

Article

Vegetable Fields Mapping in Northeast China Based on Phenological Features

Jialin Hu ^{1,2}, Huimin Lu ² , Kaishan Song ¹ and Bingxue Zhu ^{1,*}

¹ Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130028, China; hujialin@iga.ac.cn (J.H.); songkaishan@iga.ac.cn (K.S.)

² School of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China; luhuimin@ccut.edu.cn

* Correspondence: zhubingxue@iga.ac.cn

Abstract: Developing vegetable agriculture is crucial for ensuring a balanced dietary structure and promoting nutritional health. However, remote sensing extraction in open-field vegetable planting areas faces several challenges, such as the mixing of target crops with natural vegetation caused by differences in climate conditions and planting practices, which hinders the development of large-scale vegetable field mapping. This paper proposes a classification method based on vegetable phenological characteristics (VPC), which takes into account the spatiotemporal heterogeneity of vegetable cultivation in Northeast China. We used a two-step strategy. First, Sentinel-2 satellite images and land use data were utilized to identify the optimal time and key indicators for vegetable detection based on the phenological differences in crop growth. Second, spectral analysis was integrated with three machine learning classifiers, which leveraged phenological and spectral features extracted from satellite images to accurately identify vegetable-growing areas. This combined approach enabled the generation of a high-precision vegetable planting map. The research findings reveal a consistent year-by-year increase in the planting area of vegetables from 2019 to 2023. The overall accuracy (OA) of the results ranges from 0.81 to 0.93, with a Kappa coefficient of 0.83. Notably, this is the first 10 m resolution regional vegetable map in China, marking a significant advancement in economic vegetable crop mapping.

Keywords: vegetable mapping; machine learning; phenological features



Academic Editor: Yanbo Huang

Received: 9 December 2024

Revised: 18 January 2025

Accepted: 23 January 2025

Published: 26 January 2025

Citation: Hu, J.; Lu, H.; Song, K.; Zhu, B. Vegetable Fields Mapping in Northeast China Based on Phenological Features. *Agronomy* **2025**, *15*, 307. <https://doi.org/10.3390/agronomy15020307>

Copyright: © 2025 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

Vegetables are an essential part of the balanced diet recommended by the World Health Organization (WHO) and the Food and Agriculture Organization of the United Nations (FAO) [1]. According to the “2023 China Vegetable Industry Development Report”, as of 2022, the national vegetable planting area reached 336 million hectares, with a total output of 791 million tons. Recent data show that greenhouse vegetables account for approximately 40% of the total vegetable planting area in China, while outdoor vegetables make up about 60%. However, compared to greenhouse vegetables, large-scale, detailed mapping studies of open-field vegetables remain limited, which in turn restricts the efficiency and accuracy of agricultural management, environmental monitoring, and agricultural decision-making to some extent.

With the growing maturity of algorithm applications and the continuous advancement of multi-source data fusion technologies, the accuracy and reliability of crop mapping have significantly improved. This progress has enabled agricultural managers to monitor crop growth distribution more effectively. However, despite significant progress in spatial

mapping of major crops such as maize, rice, and soybeans, the study of small-scale crops, particularly open-field vegetables, still faces numerous challenges. These challenges include substantial variations in climatic conditions and planting practices across different regions, the mixed distribution of target crops with natural vegetation, and the wide diversity of vegetable types, along with their short growth cycles, small planting areas, high dispersion, and strong sensitivity to environmental factors. Collectively, these issues limit the applicability of traditional feature extraction methods and machine learning models in mapping open-field vegetables.

Current research on minor crops primarily focuses on facility vegetable agriculture [2–4], orchard management [5,6], and precise monitoring of horticultural crops and outdoor vegetables. Only a small number of studies set the target objects as vegetables. For instance, using drone data and point cloud data combined with deep learning instance segmentation techniques, researchers have achieved precise monitoring of broccoli canopy structures and plant quantities within individual plots [7,8]. Similarly, unmanned aerial vehicle (UAV) RGB images and vegetation–soil height estimation methods have been employed for spatial and temporal monitoring of cabbage plant height development [9]. Hyperspectral and multispectral sensors have proven valuable in evaluating the biological characteristics, nutritional status, and physiological parameters of young cabbage seedlings, leveraging their detailed spectral band information [10]. For smaller vegetable varieties such as carrot [11], eggplant [12], and onion [13], the study areas were limited to research bases and individual fields in Australia, Turkey, and Italy, respectively. These studies primarily focused on crop productivity, growth morphology, and nutritional parameters in small areas. UAV [14] and ground sensor data were predominantly used in these studies. However, these investigations are often confined to small-scale, specific vegetable type analyses and lack large-scale, systematic mapping studies. While drones are capable of providing high-resolution images, they face significant limitations for large-scale applications, particularly in terms of scalability. Due to constraints such as limited flight time, high operational costs, and flight area restrictions (e.g., climate, terrain complexity), drones are often less efficient than satellite data over larger regions. Additionally, drones require extensive ground support and operators, which increases the potential for human error and data collection inconsistencies. In contrast, Sentinel-1/2 satellites can cover vast areas without being constrained by terrain, climate, or other factors, and offer high spatio-temporal consistency. Their data are well-suited for large-scale monitoring and long-term trend analysis, making them particularly effective for agricultural monitoring across wide areas and diverse climatic conditions.

Recent studies on large-scale crop mapping methods has primarily focused on major crops such as corn, rice, wheat, and soybeans. Using machine learning algorithms, mapping studies have been conducted to analyze coverage areas. For example, by combining Sentinel-2 (S-2) satellite data and Global Ecosystem Dynamics Investigation (GEDI) data, researchers employed the Random Forest (RF) algorithm to extract spectral features, achieving an overall accuracy of 0.91 for maize mapping from 2019 to 2022 [15]. Similarly, using the same data sources and classifiers, the mapping precision has been further improved by effectively distinguishing crop height differences, resulting in an accuracy rate of at least 0.87 [16]. Additionally, S-2 data have been used to extract rice growth characteristics across three regions: Guangdong, Chongqing, and Heilongjiang. Using the RF classifier, the overall accuracy ranged from 95.16% to 97.54%, enabling cross-regional rice mapping [17]. In feature extraction, researchers derived time-series data, texture, phenological features, and terrain characteristics from S-2 data and evaluated these features using RF, SVM, and artificial neural networks. Multiple crop maps were produced with an overall accuracy of 0.93 [18]. However, due to the wide variety of vegetable varieties, short growth cycles,

small planting areas, and high sensitivity to environmental factors, the applicability of traditional feature extraction and machine learning models is limited.

In response to the above theory, this study aimed to utilize Sentinel-2 satellite imagery to model and analyze the growth and distribution of various vegetables in farmland. Specifically, this study aimed to address the following research questions: How can the information of open-field vegetables be accurately extracted in the complex and dynamic farmland environment? How can a high-precision, large-scale distribution map of open-field vegetables be generated? And how does the distribution of open-field vegetables change over time, from year to year? To achieve these objectives, this study proposes a novel extraction strategy that combines phenological and spectral index characteristics to more accurately identify open-field vegetables. Additionally, machine learning techniques were employed to generate vegetable distribution maps with a spatial resolution of 10 m for Northeast China from 2019 to 2023. By comparing and analyzing the mapping results across different years, this study also extracted and examined the interannual changes in the distribution of open-field vegetables.

Compared to other similar studies in this field, this research is relatively novel in the following aspects: (1) A novel extraction strategy was proposed, incorporating both phenological and spectral index features. (2) A vegetable crop distribution map, with a spatial resolution of 10 m, was generated for Northeast China covering the period from 2019 to 2023 using machine learning techniques. (3) The interannual variations in the mapping results across different years were extracted and analyzed. This research provides valuable insights and foundational data for large-scale, precision-based vegetable crop management.

2. Materials

2.1. Study Area

The study area is located in the northeast region of China (40° N to 54° N, 114° E to 130° E). Average annual temperatures range north to south from -4 °C to 11 °C and average annual precipitation from 800 mm–300 mm [19]. This temperate monsoon climate is characterized by hot and rainy summers, which coincide with the rainy season, creating favorable conditions for crop growth. At the same time, the cold winters cause plants to enter a dormant period, resulting in a distinct seasonal phenological change.

The terrain is flat and extensive, with fertile black soil containing high levels of nutrients such as ammonia, phosphorus, and potassium, suitable for plant growth. Spring snowmelt provides additional water for plant growth and spring recovery. Based on a series of field surveys (23 April–24 August 2023) and visual interpretations of the latest Google Earth images, the planting types in the entire study area vary, mainly divided into bulk crops and minor crops. The growth cycle and phenological characteristics of these crops are strongly influenced by local climatic conditions. For instance, the hot and humid summer conditions promote the rapid growth and maturation of staple crops such as rice, corn, and soybeans, while the cold winter causes these crops to enter dormancy, awaiting recovery in the spring. Similarly, the growth cycle and phenological traits of smaller crops, including fruits, oilseeds, and vegetables, are significantly affected by the local climate. Vegetable-growing areas, in particular, are primarily located in farmland regions, where the growth and harvest periods of these crops are closely linked to local rainfall and temperature patterns. For example, hardy crops such as cabbage, radishes, and kale thrive in cooler climates, whereas others, like potatoes and cauliflower, are better suited to warm, moist conditions.

The spatial distribution of the research area and samples is shown in Figure 1.

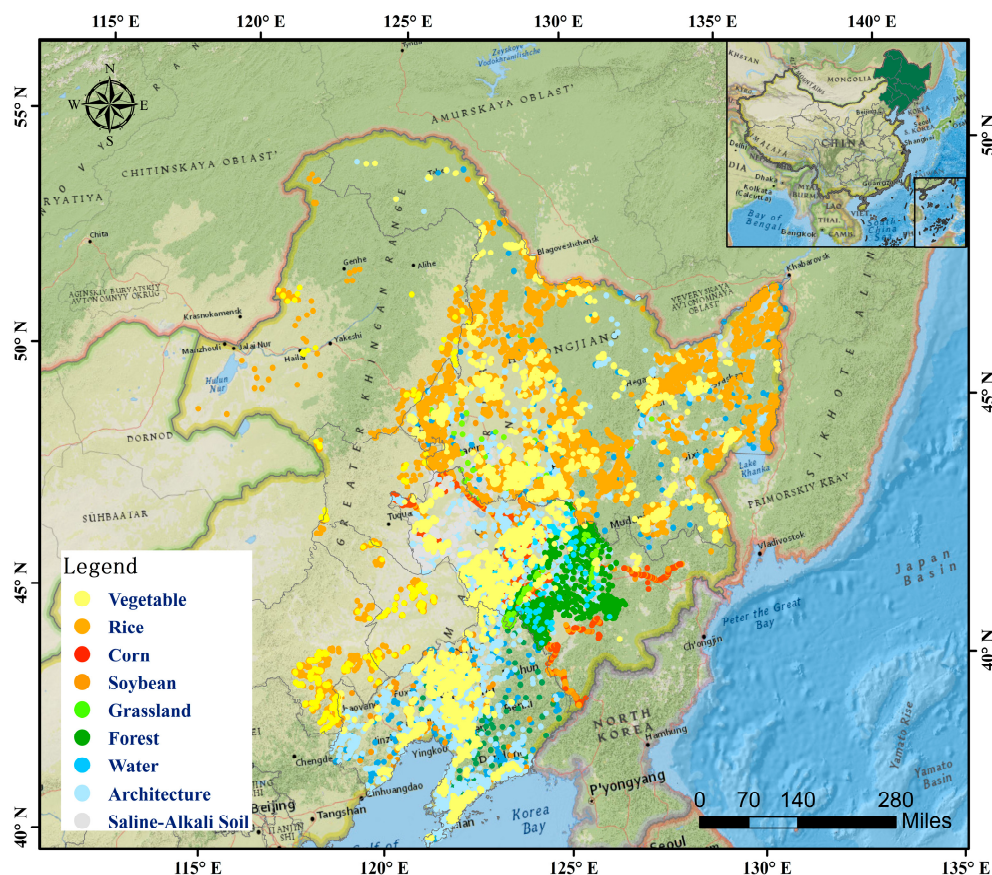




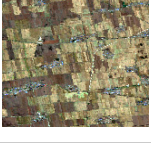



Figure 1. General information of the research area, including its geographical location and distribution of various samples.

2.2. Data and Preprocessing

2.2.1. Sentinel-2

S-2 is a high-resolution multispectral imaging satellite jointly launched by the European Space Agency (ESA) and the European Commission (EC). The images we are using are the Harmonized S-2 series. The HARMONIZED collection shifts data in newer scenes to be in the same range as in older scenes. The dataset is standardized and aligned with S-2 data through a series of preprocessing steps. These steps include atmospheric correction, geometric correction, radiometric calibration, cloud detection and removal, as well as data standardization and fusion. The assets contain 12 UTM16 spectral bands representing SR scaled by 10,000. We downloaded five different bands among them with a resolution of 10 m and an acquisition time approximately during the vegetable growing season, covering the entire northeast region of China with 129 surface reflectance images used to construct the initial time-series image collection. Examples of S-2 satellite imagery and in situ photos taken from field investigations are shown in Table 1. Among them, the vegetable area is distributed in a strip shape, mostly concentrated in farmland, usually green in color. Rice has a similar morphology to vegetable areas during the growing season. Bulk crops typically have regular texture features.

Table 1. Examples of different types of data based on phenological information.

Category	Description	Satellite Image	In Situ Photo
Vegetable Growing Area	It has a strip like distribution feature and is mostly concentrated within the farmland, presenting different colors in different seasons		
Rice	Distributed in multiple concentrations, the reflectivity is relatively similar to that of vegetable areas during the growing season		
Other Cultivated Land	Featuring regular texture features, often appearing green during the growing season		

2.2.2. Land Cover Data

ESRI 10 m Annual Land Cover is generated by the Impact Observatory. From 2017 to 2023, these maps were sourced from 10 m resolution images of the European Space Agency's Sentinel-2 spacecraft. Each map is a composite of land use and land cover (LULC) predictions for 9 classes throughout the year to generate representative snapshots for each year. The purpose is to extract the mask of cultivated land. All download, access, and processing tasks are completed on the Google Earth Engine (GEE) platform.

2.2.3. Ground Validation Samples

In 2022 and 2023, we used Global Positioning System (GPS) devices to record vegetable plot information and took field photos. Based on the field sampling points and image features, visual sampling locations were chosen on Google Earth. Table 2 presents the combined data from both visual recognition sampling points and on-site sampling information. Finally, for each province in 2023, we generated reference samples as follows: Jilin Province: 7390 samples (VP: 4010, Non-VP: 3380), Heilongjiang Province: 8002 samples (VP: 3713, Non-VP: 4289), Liaoning Province: 5230 samples (VP: 4093, Non-VP: 1137), and Neimeng Province: 972 samples (VP: 314, Non-VP: 658). The sample size of crop types in the study area is detailed in Table 2. These samples were used to train and test the results generated by the methods proposed in this study. The annotated land cover types vary across the entire study area, including vegetables, rice, corn, miscellaneous grains, grassland, woodland, water bodies, buildings, and saline–alkali land (see Figure 1). We used a random number generator to ensure the randomness of the data partitioning. By setting a random seed, we ensured that the same partitioning result was obtained each time the code was executed, which enhances repeatability. Given that the partitioning process was random and the data distribution remained consistent, the sample points were allocated to the training set and test set in an 8:2 ratio.

Table 2. Quantity of different crop types.

Class	Crop Type	JiLin	HeiLongJiang	LiaoNing	NeiMeng	Count
1	Vegetable	4010	3713	4093	314	12,130
2	Rice	236	-	417	-	653
3	Corn	1537	-	-	-	1537
4	Soybean	185	-	-	-	185
5	Grassland	135	41	-	-	176
6	Saline–Alkali Soil	163	21	-	-	184
7	Others	1124	4227	720	658	6729
8	Total	7390	8002	5230	972	21,594

Other crop types in the table mainly include mixed samples of rice, wetland, grassland and other crops.

3. Methods

We designed a classification method based on phenological characteristics. Spectral features used for classification were extracted according to the differences in time-series spectral indices of different crop sample points. Effective classification of crops and vegetables was carried out within the cropped farmland areas. Using this method, we generated annual maps of vegetable planting areas in the northeast region from 2019 to 2023. Figure 2 presents the overall framework of this study.

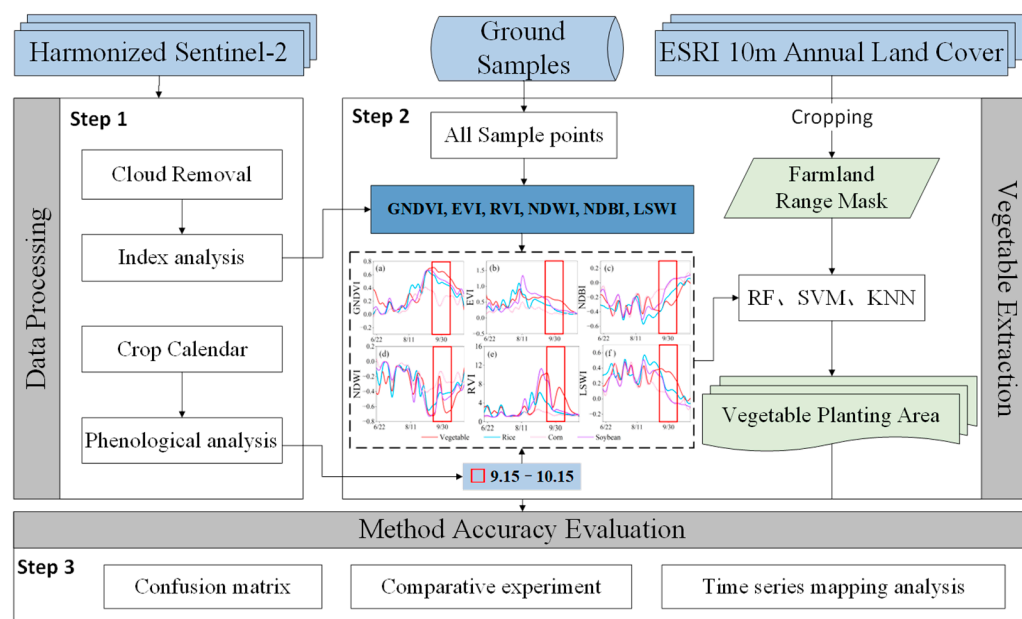


Figure 2. Overall framework for the Vegetable Phenological Feature-based Classification method (VPC) developed in this paper.

3.1. Remote Sensing Imagery and Indices

By combining the unique spectral characteristics of various land cover types, feature indices were selected for differentiation. Vegetation indices were used to generate index curves with time-series characteristics, further revealing the differences between vegetables and other land cover types, providing effective variables for classification. The indices we used included the Green Normalized Difference Vegetation Index (GNDVI), Enhanced Vegetation Index (EVI), Ratio Vegetation Index (RVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-up Index (NDBI). The index list is shown in Table 3.

GNDVI quantifies vegetation greenness by leveraging the difference between the red and near-infrared spectral bands. As plants grow and mature, their chlorophyll con-

centration increases, leading to higher GNDVI values [20]. Therefore, the GNDVI value is directly proportional to the growth status of the vegetable. NDWI is primarily used to identify water bodies and water-rich areas. For vegetable growing areas, NDWI can reflect the degree of irrigation and soil moisture. During the flourishing period of paddy field crops, irrigation is frequent, and soil moisture is high, which causes the NDWI value to differ from that of vegetables. This distinction helps to differentiate between areas of vegetable cultivation and paddy fields based on their water content and irrigation patterns. It provides auxiliary information for distinguishing these features [21] and differentiates urban built-up areas from natural surfaces by comparing the reflection differences between near-infrared and shortwave infrared bands. In areas near vegetable growing zones, the presence of artificial structures such as buildings or roads leads to relatively high NDBI values for these impervious surfaces [22]. By comparing the GNDVI with NDBI, it becomes possible to further distinguish vegetable growing areas from the surrounding urban environment, enhancing the accuracy of identifying agricultural zones amidst urban landscapes. Primarily used to distinguish urban built-up areas from natural surfaces, NDBI can identify natural surfaces after field harvesting, effectively separating vegetable planting areas from their surrounding backgrounds. The RVI index correlates strongly with leaf area index (LAI), leaf dry biomass (DM), and chlorophyll content, making it a reliable indicator for monitoring vegetation growth and health. RVI helps exclude crops with relatively low chlorophyll content or distinct growth cycles, such as certain early or late maturing varieties of vegetables [23]. EVI accurately captures vegetation greenness, offering a significant advantage in eliminating weeds or low-growing crops that share similar growth environments with vegetables but have lower biomass. Additionally, EVI performs well in vegetation monitoring under varying lighting conditions [24,25]. LSWI serves as a reference for irrigation management, reflecting the water status of crops. It is particularly useful for crops with pronounced seasonal characteristics, such as rice. By distinguishing mature rice plants, LSWI maximally highlights the differences between vegetable fields and rice fields [26].

Table 3. List of indices.

Index	Equation	References
GNDVI	$\frac{\rho_{nir} - \rho_{green}}{\rho_{nir} + \rho_{green}}$	[20]
NDWI	$\frac{\rho_{green} - \rho_{swir}}{\rho_{green} + \rho_{swir}}$	[27]
NDBI	$\frac{\rho_{swir} - \rho_{nir}}{\rho_{swir} + \rho_{nir}}$	[28]
EVI	$2.5 \times \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + 6\rho_{red} - 7.5\rho_{blue} + 1}$	[29]
RVI	$\frac{\rho_{nir}}{\rho_{red}}$	[30]
LSWI	$\frac{\rho_{nir} - \rho_{swir}}{\rho_{nir} + \rho_{swir}}$	[31]

Where ρ_{red} , ρ_{green} , ρ_{blue} , ρ_{nir} , and ρ_{swir} represent the green (560 nm), red (665 nm), blue (490 nm), near-infrared (842 nm), and shortwave infrared (1610 nm) bands of the S-2 image, respectively.

3.2. Phenological Features Analysis

We selected the maturity period of vegetables with significant differences in vegetation indices as the optimal identification period for crop classification. Due to geographical differences in climate, there are different phenological periods for vegetable crop growth in the different provinces. Vegetables exhibit two growth patterns, unimodal and bimodal, on individual vegetation index time series. Although vegetables with bimodal growth curves show significant differences from bulk crops during late June to mid-August, vegetables

with unimodal curves overlap with them. Therefore, we chose vegetables with bimodal curves as effective sample indicators for distinguishing crop categories.

Based on year-round S-2 images and corresponding ground reference data in the four provinces of the northeast region, land cover types include three main vegetables (Chinese cabbage, potatoes, and cabbage, collectively referred to as “vegetables” hereafter). The phenological calendars of these three main vegetables and other crops are shown in Figure 3. Specifically, the potato planting stage typically occurs in late April, followed by germination, growth, filling, and maturation stages. Chinese cabbage is sown in August, followed by germination, seedling stage, heading stage, and balling stage [32]. Cabbage generally goes through the bolting stage, flowering stage, and fruiting stage [33], similar to the growth cycle of Chinese cabbage. From the graph, it can be seen that there is some overlap in the phenological characteristics of the three main crops, mainly concentrated in the maturity period of potatoes, the heading period of Chinese cabbage, and the fruiting period of cabbage. Therefore, based on the Sentinel-2 (S-2) images and the corresponding ground reference data from the four provinces in Northeast China throughout the year, we selected the maturity stages of the vegetables, which show significant differences in vegetation indices, as the optimal classification period for crop classification. The window period was limited to mid-June to late October, when these differences in vegetation indices were most pronounced.

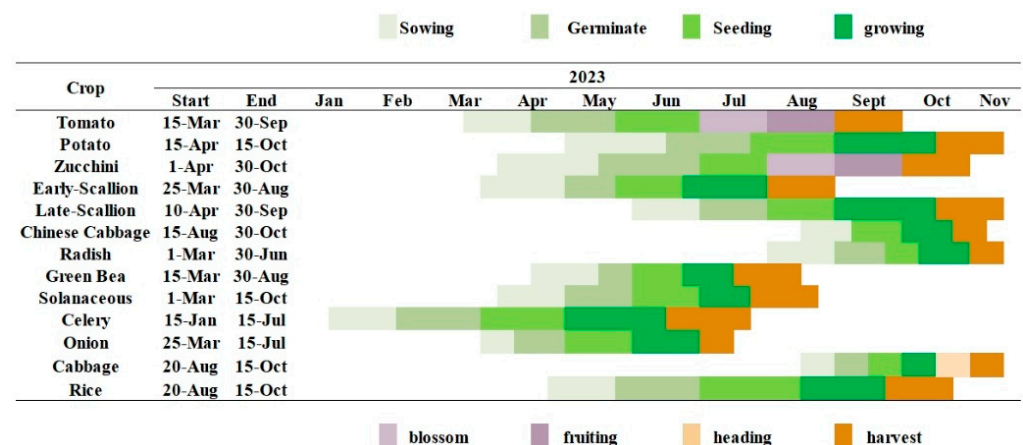


Figure 3. Phenological calendars for vegetables. Based on year-round field surveys and data collection results, we established the phenological calendars of the main crop types in the study area and analyzed the phenological information of different crops.

During this time interval, the contrast between the target and background in the image is higher, which can effectively filter out easily confused crops through vegetation indices and reflect the information differences between vegetable areas and other land cover types to the maximum extent possible, accurately reflecting the growth range of vegetables.

Subsequently, we used the GEE platform to obtain data at a 5 day frequency throughout the year. Based on the performance characteristics of different land cover types, vegetable crops exhibited two distinct growth patterns in the vegetation index time series: single peak and double peak. Upon observation, from mid-June to late October, vegetable crops with bimodal growth curves showed clear differences from bulk crops, while vegetable crops with unimodal growth curves partially overlapped with bulk crops. Therefore, vegetable crops with bimodal curves were selected as valid sample indicators for distinguishing crop categories. Finally, this study defined the time window from the end of September to early October each year as the key period for crop classification. This period coincides with the harvest of bulk crops and the peak growth of vegetable crops. During this time, the contrast between the target crops and the background is high, allow-

ing easily confused crops to be effectively filtered using the vegetation index. This time window maximizes the differentiation between vegetable areas and other land cover types, providing an accurate reflection of the vegetable growing range.

3.3. Machine Learning Methods

Random Forest (RF) [34], Support Vector Machine (SVM) [35], and k-Nearest Neighbors (KNN) [36] are three popular machine learning algorithms used for remote sensing classification tasks. First, as an ensemble learning method, RF improves model accuracy and generalization by constructing a “forest” of decision trees trained on random subsets of data and features, reducing the risk of overfitting. It excels in processing complex remote sensing data, especially for vegetable planting areas with diverse vegetation and complex growth environments, enabling more accurate classification. We initialized a random forest classifier with 25 trees due to its strong processing capabilities and adaptability to high-dimensional data.

Second, SVM aims to find an optimal hyperplane that separates different categories while maximizing the margin between them. It performs well in both binary and multi-class tasks, especially when distinguishing vegetable planting areas with clear vegetation differences. SVM also generalizes well to unseen data and handles nonlinear classification problems, which is crucial for vegetable mapping with complex growth patterns and terrain. We chose SVM for its ability to handle nonlinear data and define clear classification boundaries.

Finally, KNN is a simple and effective classification method, suitable for smaller datasets. In vegetable mapping, KNN classifies based on the feature space distance of remote sensing data without complex model training. However, for large-scale datasets, KNN’s computational complexity can be high. Despite this, we included KNN as a comparison algorithm to evaluate its performance in specific cases. In this study, we set the K value to 1, considering only the nearest neighbor’s category for classification to simplify calculations and compare algorithm performance.

3.4. Accuracy Assessments

The classification results from our VPC method showed strong agreement with ground validation samples. To evaluate the accuracy of the vegetable map we generated, we compared ground sampling data with existing results. We calculated the OA [37,38] and KAPPA [39] of the map area using statistical data from 2019–2023. The formulae are as follows:

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

$$\text{KAPPA} = \frac{p_0 - p_e}{1 - p_e} \quad (2)$$

where TP is the number of correctly identified positive samples, FP is the number of negative samples with false positives, TN is the number of correctly identified negative samples, FN is the number of missed positive samples, p_0 is the sum of the number of correctly classified samples for each class divided by the total number of samples, and p_e is an accidental consistency error, which refers to the expected consistency rate.

4. Results Analysis

4.1. Index Effectiveness Analysis

We analyzed the characteristic changes of rice, maize, soybean, and vegetable sample points during the 2023 growing season by combining six spectral indices: GNDVI, EVI, RVI, NDWI, NDBI, and LSWI (Figure 4). During the peak growing season (late September to early October), vegetables exhibit their highest GNDVI values, significantly differing

from those of corn crops. At this time, corn enters the maturing stage, with leaves turning yellow, a reduction in leaf area, and decreased ground coverage. In contrast, vegetables are at their growth peak, with the highest biomass content and maximum leaf area, resulting in markedly higher GNDVI values compared to other land cover types.

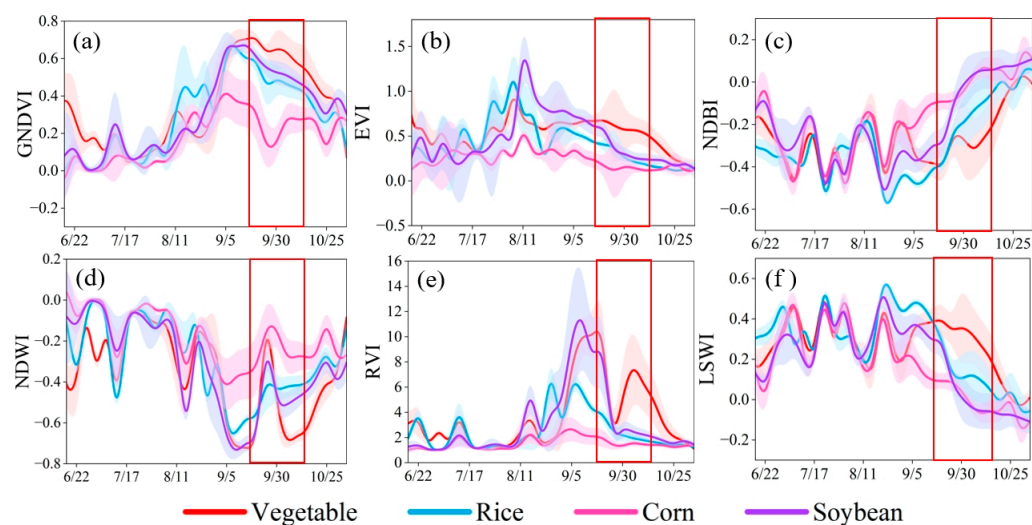


Figure 4. Spectral index characteristics of five-day time series datasets (GNDVI, EVI, RVI, NDWI, NDBI, LSWI); subfigures (a–f) show the differences in spectral index characteristics between vegetables and other samples. The error band represents the standard deviation of the sample.

Similar to GNDVI, vegetables generally exhibit higher EVI and RVI values due to their high chlorophyll content and rapid growth, which effectively distinguishes green vegetation from non-vegetation areas. As part of natural vegetation, vegetables typically have low NDBI values, which can help differentiate vegetable planting areas from periods when bulk crops are in the bare soil stage. During the middle and later stages of the study period, significant differences in NDWI values were observed between vegetables and other crops. And wetlands, rice fields, and grasslands may be misclassified as vegetable planting areas due to their similar spectral characteristics. To address this, NDWI and LSWI should be used to further eliminate interference from areas such as rice fields. By intersecting the classification results of existing farmland mapping with the exclusion of rice reflectance, rice field targets can be filtered out, enabling the exclusion of irrigated areas and paddy fields, allowing for the accurate extraction of vegetable planting areas.

4.2. Classification Result Analysis

The confusion matrix presented in Table 4 details the classification results of vegetable planting areas across various provinces from 2019 to 2023. Each pair of rows represents the classification results for a year and crops in different provinces. The four indicators for different years and provinces, listed in order from left to right and from top to bottom, are as follows: (1) the area that is actually a vegetable and correctly predicted by the model as a vegetable, (2) the area that is actually a vegetable but incorrectly classified by the model as another crop, (3) the area that is actually another crop but incorrectly classified by the model as a vegetable, and (4) the area that is actually another crop and correctly predicted by the model as another crop. It is important to note that the quantity of different crops shows fluctuations over the years. In 2021, there was a significant decline in classification performance, which may be attributed to specific environmental factors or data quality issues during that year. In Liaoning Province, over time, the degree of confusion between vegetables and other crops has gradually increased. This confusion may result from changes in crop planting structures, variations in crop growth cycles, or the model's limitations in

adapting to new data. Furthermore, there are significant differences in the distribution of vegetables and other crops across different cities and years. This variation is evident not only in the diversity of crop species but also in the fluctuations of planting areas. In the Inner Mongolia region, although the sample size is relatively small, leading to a lower absolute number of classifications, the model still exhibits good overall performance. This suggests that our model can maintain a certain level of accuracy and stability, even with limited sample sizes. In contrast, Jilin Province exhibits the best classification performance among all provinces. This may be closely linked to its favorable local climate conditions, fertile soil, high-quality image data, and a sufficient sample size. These advantageous factors in Jilin collectively provide a solid foundation for the training and validation of the model, ensuring optimal classification performance.

Table 4. Confusion matrix of crop type maps from 2019 to 2023.

Year	Province Crop	JiLin		LiaoNing		HeiLongJiang		NeiMeng	
		Vegetable Samples	Other Samples	Vegetable Samples	Other Samples	Vegetable Samples	Other Samples	Vegetable Samples	Other Samples
2019	Vegetable Samples	692	85	826	19	510	168	48	6
	Other Samples	55	369	26	163	53	792	1	83
2020	Vegetable Samples	709	80	830	26	471	173	46	6
	Other Samples	56	233	22	150	51	735	4	104
2021	Vegetable Samples	696	64	766	24	468	144	49	5
	Other Samples	58	320	24	179	83	763	6	102
2022	Vegetable Samples	725	69	800	20	510	142	54	5
	Other Samples	59	359	19	165	53	740	6	91
2023	Vegetable Samples	762	90	756	21	590	101	52	6
	Other Samples	43	351	19	147	35	811	8	92

4.3. Comparison Accuracy of Different Models

We introduced three different classifiers and conducted comparative tests. The results indicated that the RF classifier demonstrated optimal performance in distinguishing these easily confused crops. This approach significantly enhanced classification accuracy, enabling us to more precisely identify low-growing crops and vegetable planting areas. Using the classification results from 2023, which are the most representative, our study demonstrated that the RF model outperformed other methods in terms of OA and KAPPA evaluation metrics. Specifically, RF achieved improvements of approximately 4% to 9% compared to the K-Nearest Neighbors (KNN) and SVM algorithms (Figure 5). The OA of vegetable area classification exceeded 85%, with the KAPPA coefficient consistently above 0.75, indicating high reliability in mapping vegetable planting areas.

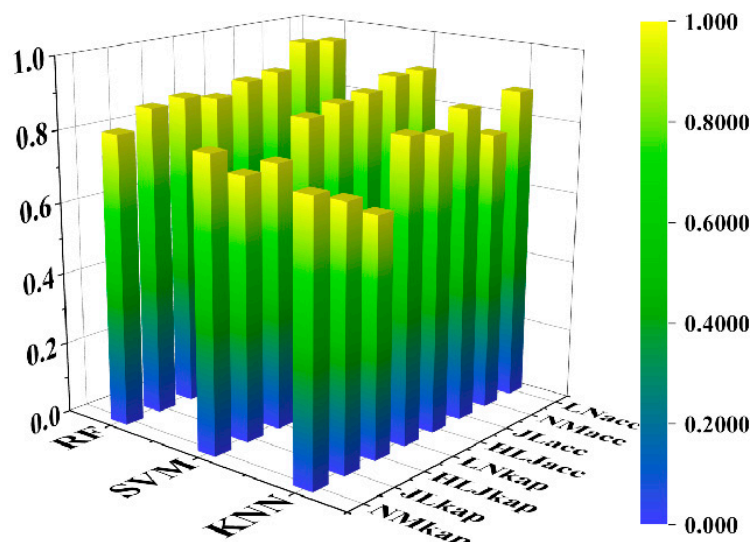


Figure 5. The classification accuracy of different classifiers in the four provinces in 2023.

4.4. Time Series Mapping and Analysis

In this study, we generated a detailed spatial distribution map and an intuitive area distribution chart covering vegetable planting areas in Northeast China from 2019 to 2023. As illustrated in Figure 6, vegetable cultivation exhibited a distinct clustering pattern across the region, primarily concentrated in five key areas: Qiqihar City in Heilongjiang Province, Baicheng City and Siping City in Jilin Province, and Shenyang City and Dalian City in Liaoning Province. These regions, with their unique natural conditions, rich agricultural resources, and long history of cultivation, have become the core areas for vegetable farming in Northeast China. Collectively, these areas account for over 45% of the total vegetable cultivation area in the region. To further visualize the proportion of vegetable cultivation relative to arable land and total land area, we created bar charts and line charts. As shown in Figure 6f, the bar chart clearly illustrates the vegetable planting area across different years. From 2019 to 2023, vegetable fields expanded by 7000 square kilometers, underscoring the significant role and development trend of vegetable cultivation in agricultural production. Meanwhile, the line chart depicts the change in the ratio of vegetable planting area to total cultivated land over time, showing a year-on-year increase of two percentage points. This trend intuitively reflects the growth and scaling of the vegetable planting industry in Northeast China. We observed a year-on-year increase in vegetable planting areas, aligning with the expansion of arable land in the region in recent years.

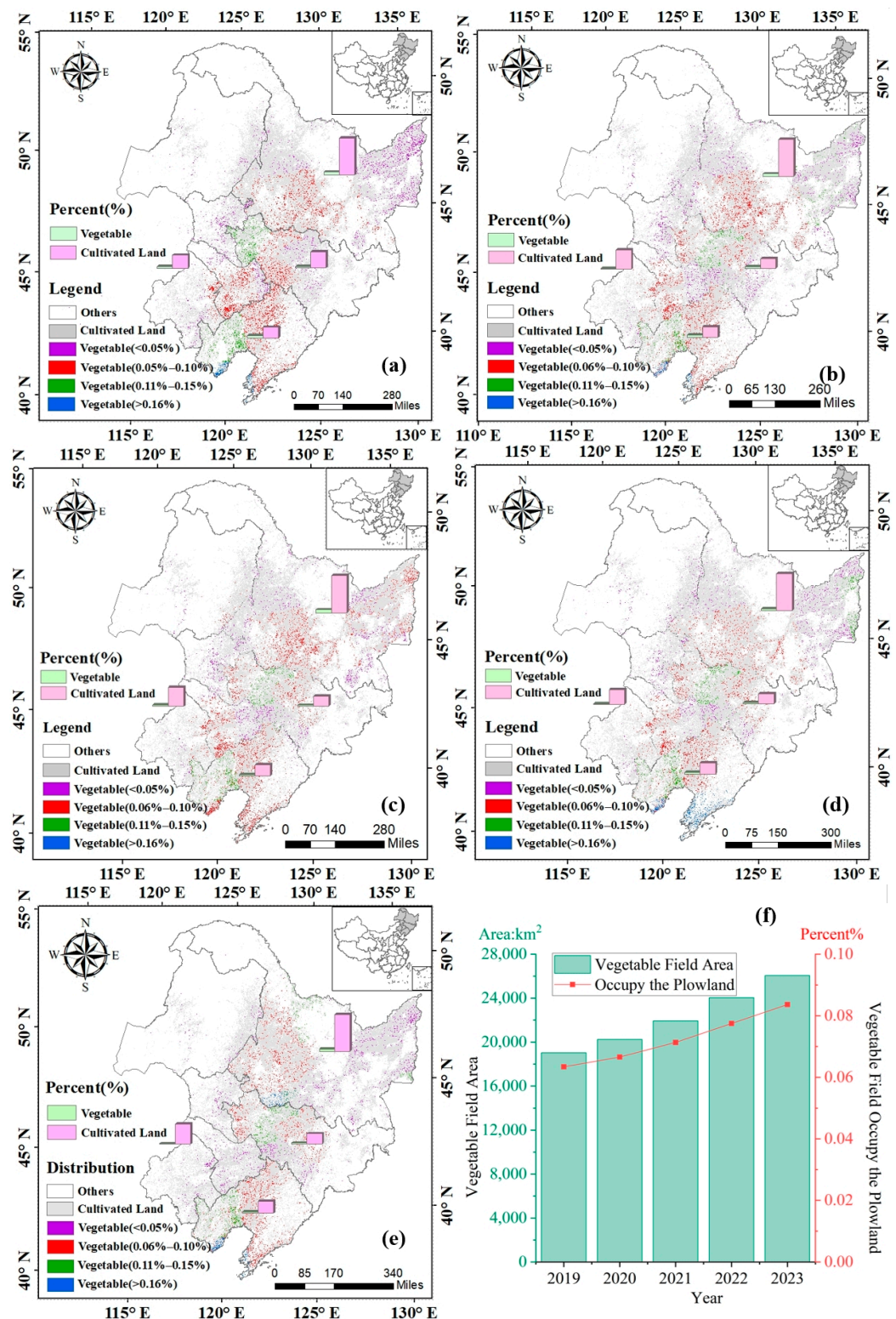


Figure 6. The vegetable distribution maps of Northeast China in 2019 (a), 2020 (b), 2021 (c), 2022 (d), and 2023 (e). Proportion of vegetable area in 2019 to 2023 (f). Based on the proportion of vegetables to cultivated land, we divide vegetables into four distribution intensities, represented by yellow, red, green, and blue from low to high.

5. Discussion

Figure 7 presents a visual comparison of vegetable images using various classification methods and classifiers across four representative regions in 2023. In sparse vegetable

planting areas in Figure 7(a1–a3), other classifiers showed considerable omissions and misclassifications. In contrast, our Random Forest classifier performed better in these situations. In regions where rice and vegetables are interplanted in Figure 7(b1–b3), the similar peak biomass periods of both crops resulted in a relatively small or even absent pixel area for vegetables. Nevertheless, the Random Forest classifier successfully eliminated rice crops, minimizing classification confusion. In large vegetable planting areas that intersect with complex urban environments in Figure 7(c1–c3), our algorithm effectively reduced missed classifications and ensured overall accuracy. Furthermore, in densely planted vegetable areas in Figure 7(d1–d3), our hierarchical classification method, combined with the RF classifier, accurately removed bulk crops used for spacing in vegetable fields, further enhancing classification precision.

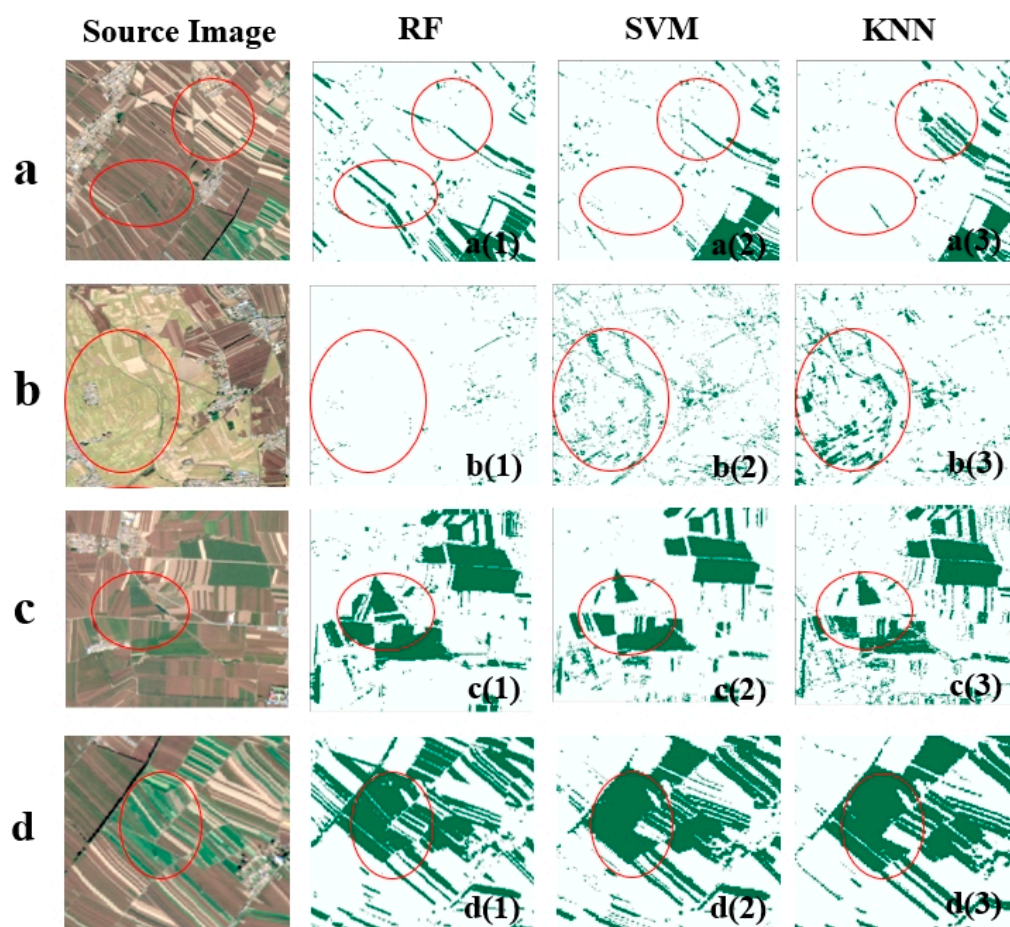


Figure 7. Visual comparison of different classification methods and classifiers for vegetable maps in 2023: The lush growth period of vegetables is composed of multiple bands in S-2, including NIR (band 8), RED (band 4), GREEN (band 3) and BLUE (band 2), forming the original image: (a) are sparse vegetable planting areas; (b) is a vegetable area mixed with rice and with a relatively small pixel area; (c) is a vegetable area that intersects with complex urban features and has a large planting area; (d) is a dense vegetable planting area. (1–3) classification mapping, from left to right: RF, SVM, and KNN.

Based on the observations shown in Figure 6, we found that the vegetable planting area is increasing year by year, a trend closely linked to the recent expansion of cultivated land in Northeast China. By 2023, the total vegetable cultivation area in the region reached 26,062 square kilometers, marking significant progress in agricultural development. Given the current growth trajectory, the vegetable planting area is expected to continue expanding in the coming years. In this context, the cartographic method proposed in this study

will play an increasingly important role in continuous change monitoring. By regularly monitoring remote sensing data and integrating machine learning algorithms, changes in vegetable planting areas and structures can be tracked in real time.

The VPC method has demonstrated good performance with the current dataset, and we believe it is applicable to regions with similar planting habits in Central China. However, further validation is needed in Southern China. Additionally, the acquisition and processing of remote sensing data can be affected by various factors, such as extreme weather events, cloud cover, and other conditions, which lead to changes in vegetation spectral characteristics. This can complicate the relationship between spectral indices and vegetable growth states, making them unstable. Due to differences in planting structures, crop types, and growth cycles, land cover types are relatively complex, potentially leading to issues like mixed pixels and loss of detailed information.

To address these limitations, we propose several avenues for future research. First, expanding the study area and sample size, and focusing on more accurate time windows for single or multiple crops based on local planting habits and climate conditions, would help verify the applicability and accuracy of the VPC method in diverse regions. Additionally, multi-temporal data fusion, meteorological data correction, or incorporating other data sources (e.g., UAV remote sensing or commercial satellite data), along with ground observation data and advanced image processing techniques (e.g., super-resolution reconstruction), could enhance classification accuracy and resolution. For different climate regions, more granular regional models and high-resolution climate data could improve model accuracy. Considering the potential impacts of climate change, adopting multi-scenario forecasting methods to simulate agricultural production under various climatic conditions would help increase the robustness and adaptability of the model.

6. Conclusions

This study, based on the phenological variation patterns of vegetables and other land cover types, identified suitable spectral indices for vegetable mapping. By combining machine learning classifiers, we developed the VPC method. This method accurately distinguishes vegetable areas from other land cover types. RF has the best effect, with an overall accuracy of 0.87 and a Kappa coefficient of 0.93. By analyzing the results of fine mapping of vegetables in Northeast China from 2019 to 2022, we observed a year-on-year increase in the vegetable planting area. Notably, Jilin Province, a major agricultural hub, saw a significant increase of 171 square kilometers in its vegetable planting area. In contrast, Inner Mongolia (Eastern Four Leagues) experienced a decline in vegetable cultivation, with a decrease of 90 square kilometers. Looking ahead, we will continue to explore and optimize VPC approaches, incorporating localized climate data (such as temperature, precipitation, humidity, and wind speed), planting structure models (including crop species, planting density, and growth cycles in different regions), and enhancing ground data collection (covering crop growth status, chlorophyll content, soil moisture, etc.) to better address these challenges.

Author Contributions: J.H., responsible for data collection and analysis, resource provision, drafting the initial version of the paper, and revising the paper. H.L., in charge of experimental methods, resource provision, and drafting the initial version of the paper. K.S., responsible for conceiving the paper, resource provision, revising the paper, obtaining project funding, and supervising the implementation of the research. B.Z., responsible for data collection and analysis, resource provision, experimental methods, and drafting the initial version of the paper. All authors have agreed to the final version of the manuscript and are willing to take responsibility for the accuracy and authenticity of the entire research work to ensure that any issues related to the accuracy or integrity of any part of

the manuscript are appropriately investigated and resolved. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Jilin Province Natural Science Foundation (YDZJ202301-ZYTS239) and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA28050400).

Data Availability Statement: Data will be available upon request. If you need the data, please contact us.

Acknowledgments: Thanks to the National Earth System Science Data Center, China (www.geodata.cn, accessed on 1 January 2024), which has provided a solid data foundation for this study.

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

References

1. Yang, Q.; Chen, J.; Dai, J.; He, Y.; Wei, K.; Gong, M.; Chen, Q.; Sheng, H.; Su, L.; Liu, L. Total coliforms, microbial diversity and multiple characteristics of Salmonella in soil-irrigation water-fresh vegetable system in Shaanxi, China. *Sci. Total Environ.* **2024**, *924*, 171657. [[CrossRef](#)] [[PubMed](#)]
2. Hong, R.; Xiao, B.; Yan, H.; Liu, J.; Liu, P.; Song, Z. Multitemporal greenhouse mapping for high-resolution remote sensing imagery based on an improved YOLOX. *Comput. Electron. Agric.* **2023**, *206*, 107689. [[CrossRef](#)]
3. Chen, D.; Ma, A.; Zheng, Z.; Zhong, Y. Large-scale agricultural greenhouse extraction for remote sensing imagery based on layout attention network: A case study of China. *ISPRS J. Photogramm. Remote Sens.* **2023**, *200*, 73–88. [[CrossRef](#)]
4. Yang, D.; Chen, J.; Zhou, Y.; Chen, X.; Chen, X.; Cao, X. Mapping plastic greenhouse with medium spatial resolution satellite data: Development of a new spectral index. *ISPRS J. Photogramm. Remote Sens.* **2017**, *128*, 47–60. [[CrossRef](#)]
5. Ozdarici-Ok, A.; Ok, A.O. Using remote sensing to identify individual tree species in orchards: A review. *Sci. Hortic.* **2023**, *321*, 112333. [[CrossRef](#)]
6. Zhang, T.; Hu, D.; Wu, C.; Liu, Y.; Yang, J.; Tang, K. Large-scale apple orchard mapping from multi-source data using the semantic segmentation model with image- to- image translation and transfer learning. *Comput. Electron. Agric.* **2023**, *213*, 108204. [[CrossRef](#)]
7. Lee, C.-J.; Yang, M.-D.; Tseng, H.-H.; Hsu, Y.-C.; Sung, Y.; Chen, W.-L. Single-plant broccoli growth monitoring using deep learning with UAV imagery. *Comput. Electron. Agric.* **2023**, *207*, 107739. [[CrossRef](#)]
8. Zhou, C.; Ye, H.; Sun, D.; Yue, J.; Yang, G.; Hu, J. An automated, high-performance approach for detecting and characterizing broccoli based on UAV remote-sensing and transformers: A case study from Haining, China. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103055. [[CrossRef](#)]
9. Jamil, N.; Kootstra, G.; van Apeldoorn, D.F.; Van Henten, E.J.; Kooistra, L. UAV time-series imagery show diversity treatment effects on cabbage growth. *Smart Agric. Technol.* **2024**, *8*, 100443. [[CrossRef](#)]
10. Martins, G.D.; Sousa Santos, L.C.; dos Santos Carmo, G.J.; da Silva Neto, O.F.; Castoldi, R.; Machado, A.I.M.R.; de Oliveira Charlo, H.C. Multispectral images for estimating morphophysiological and nutritional parameters in cabbage seedlings. *Smart Agric. Technol.* **2023**, *4*, 100211. [[CrossRef](#)]
11. Suarez, L.A.; Robson, A.; McPhee, J.; O'Halloran, J.; van Sprang, C. Accuracy of carrot yield forecasting using proximal hyperspectral and satellite multispectral data. *Precis. Agric.* **2020**, *21*, 1304–1326. [[CrossRef](#)]
12. Taşan, S.; Cemek, B.; Taşan, M.; Cantürk, A. Estimation of eggplant yield with machine learning methods using spectral vegetation indices. *Comput. Electron. Agric.* **2022**, *202*, 107367. [[CrossRef](#)]
13. Messina, G.; Pratico, S.; Badagliacca, G.; Di Fazio, S.; Monti, M.; Modica, G. Monitoring onion crop “cipolla rossa di Tropea Calabria IGP” growth and yield response to varying nitrogen fertilizer application rates using UAV imagery. *Drones* **2021**, *5*, 61. [[CrossRef](#)]
14. Canicatti, M.; Vallone, M. Drones in vegetable crops: A systematic literature review. *Smart Agric. Technol.* **2024**, *7*, 100396. [[CrossRef](#)]
15. Li, Z.; Xuan, F.; Dong, Y.; Huang, X.; Liu, H.; Zeng, Y.; Su, W.; Huang, J.; Li, X. Performance of GEDI data combined with Sentinel-2 images for automatic labelling of wall-to-wall corn mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *127*, 103643. [[CrossRef](#)]
16. Di Tommaso, S.; Wang, S.; Vajipey, V.; Gorelick, N.; Strey, R.; Lobell, D.B. Annual Field-Scale Maps of Tall and Short Crops at the Global Scale Using GEDI and Sentinel-2. *Remote Sens.* **2023**, *15*, 4123. [[CrossRef](#)]
17. Sun, L.; Lou, Y.; Shi, Q.; Zhang, L. Spatial domain transfer: Cross-regional paddy rice mapping with a few samples based on Sentinel-1 and Sentinel-2 data on GEE. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *128*, 103762. [[CrossRef](#)]
18. Luo, K.; Lu, L.; Xie, Y.; Chen, F.; Yin, F.; Li, Q. Crop type mapping in the central part of the North China Plain using Sentinel-2 time series and machine learning. *Comput. Electron. Agric.* **2023**, *205*, 107577. [[CrossRef](#)]

19. Niu, H.; Sun, Y.; Wang, J.; Marquer, L.; Vessies, J.; Sack, D.; Chambers, F.M.; Jie, D. Drivers of land cover and plant compositional changes in Northeast China since the mid-Holocene: Climate versus human activities. *J. Archaeol. Sci.* **2024**, *163*, 105938. [[CrossRef](#)]
20. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
21. Cho, M.A.; van Aardt, J.; Main, R.; Majeke, B. Evaluating variations of physiology-based hyperspectral features along a soil water gradient in a Eucalyptus grandis plantation. *Int. J. Remote Sens.* **2010**, *31*, 3143–3159. [[CrossRef](#)]
22. 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](#)]
23. Yanlin, T.; Renchao, W.; Xiuzhen, W.; Yunmei, L.I. Study on the Leaf Area Index and Biochemical Contents of Rice Blade by the Spectral Method. *J. South China Agric. Univ.* **2003**, *24*, 4–7.
24. Vijith, H.; Dodge-Wan, D. Applicability of MODIS land cover and Enhanced Vegetation Index (EVI) for the assessment of spatial and temporal changes in strength of vegetation in tropical rainforest region of Borneo. *Remote Sens. Appl. Soc. Environ.* **2020**, *18*, 100311. [[CrossRef](#)]
25. Gurung, R.B.; Breidt, F.J.; Dutin, A.; Ogle, S.M. Predicting Enhanced Vegetation Index (EVI) curves for ecosystem modeling applications. *Remote Sens. Environ.* **2009**, *113*, 2186–2193. [[CrossRef](#)]
26. Boles, S.H.; Xiao, X.; Liu, J.; Zhang, Q.; Munkhtuya, S.; Chen, S.; Ojima, D. Land cover characterization of Temperate East Asia using multi-temporal VEGETATION sensor data. *Remote Sens. Environ.* **2004**, *90*, 477–489. [[CrossRef](#)]
27. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [[CrossRef](#)]
28. 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](#)]
29. Huete, A.R.; Liu, H.Q.; Batchily, K.; van Leeuwen, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sens. Environ.* **1997**, *59*, 440–451. [[CrossRef](#)]
30. Clevers, J.G.P.W.; Gitelson, A.A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 344–351. [[CrossRef](#)]
31. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. [[CrossRef](#)]
32. Lacz, E.; Apahidean, A.; Luca, E.; Dumitrascedilla, A.; Boancă, P. Headed Chinese cabbage growth and yield influenced by different manure types in organic farming system. *Hort. Sci.* **2016**, *43*, 42–49. [[CrossRef](#)]
33. Paranhos, L.G.; Barrett, C.E.; Zotarelli, L.; Darnell, R.; Migliaccio, K.; Borisova, T. Planting date and in-row plant spacing effects on growth and yield of cabbage under plastic mulch. *Sci. Hortic.* **2016**, *202*, 49–56. [[CrossRef](#)]
34. Saffari, A.; Leistner, C.; Santner, J.; Godec, M.; Bischof, H. On-line Random Forests. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Kyoto, Japan, 27 September–4 October 2009; pp. 1393–1400.
35. Cristianini, N.; Ricci, E. Support Vector Machines. In *Encyclopedia of Algorithms*; Kao, M.-Y., Ed.; Springer: Boston, MA, USA, 2008; pp. 928–932.
36. Steinbach, M.; Tan, P.-N. kNN: K-nearest neighbors. In *The Top Ten Algorithms in Data Mining*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2009; pp. 165–176.
37. You, N.; Dong, J.; Huang, J.; Du, G.; Zhang, G.; He, Y.; Yang, T.; Di, Y.; Xiao, X. The 10-m crop type maps in Northeast China during 2017–2019. *Sci. Data* **2021**, *8*, 41. [[CrossRef](#)] [[PubMed](#)]
38. Mei, Q.; Zhang, Z.; Han, J.; Song, J.; Dong, J.; Wu, H.; Xu, J.; Tao, F. ChinaSoyArea10m: A dataset of soybean-planting areas with a spatial resolution of 10 m across China from 2017 to 2021. *Earth Syst. Sci. Data* **2024**, *16*, 3213–3231. [[CrossRef](#)]
39. Chen, H.; Li, H.; Liu, Z.; Zhang, C.; Zhang, S.; Atkinson, P.M. A novel Greenness and Water Content Composite Index (GWCCI) for soybean mapping from single remotely sensed multispectral images. *Remote Sens. Environ.* **2023**, *295*, 113679. [[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.