Article

The Classification Performance and Mechanism of Machine Learning Algorithms in Winter Wheat Mapping Using Sentinel-2 10 m Resolution Imagery

Peng Fang [1,2], **Xiwang Zhang** [1,2,*], **Panpan Wei** [1,2], **Yuanzheng Wang** [1,2], **Huiyi Zhang** [1,2], **Feng Liu** [1,2] **and Jun Zhao** [1,2]

[1] Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River
Regions (Henan University), Ministry of Education, Kaifeng 475004, China; fangpeng@vip.henu.edu.cn (P.F.); weipan1014@henu.edu.cn (P.W.); wangyz@vip.henu.edu.cn (Y.W.); zhanghuiyi@vip.henu.edu.cn (H.Z.); liufeng@vip.henu.edu.cn (F.L.); zhaojun@vip.henu.edu.cn (J.Z.)

[2] College of Environment and Planning, Henan University, Kaifeng 475004, China

* Correspondence: zhangxiwang@vip.henu.edu.cn

check for updates

Featured Application: Machine learning algorithms are essential to crop identification and land use/cover. Our work indicates that compared to RF and CART algorithms, SVM achieves best performance in identifying winter wheat. Though SVM is sensitive to algorithm parameters, it obtains maximum value of accuracy score and minimum residuals by tuning hyperparameters. Therefore, we recommend that SVM will be a suitable and effective algorithm if researchers employ machine learning algorithms for crop identification, land use/cover and other similar research with small sample data. Meanwhile, no matter which algorithm we choose, we should focus on the importance of the performance and mechanism. The enumeration method, similar to grid search, should be used to fine-tune the hyperparameters when discussing the use of various algorithm parameters.

Abstract: Machine learning algorithms are crucial for crop identification and mapping. However, many works only focus on the identification results of these algorithms, but pay less attention to their classification performance and mechanism. In this paper, based on Google Earth Engine (GEE), Sentinel-2 10 m resolution images during a specific phenological period of winter wheat were obtained. Then, support vector machine (SVM), random forest (RF), and classification and regression tree (CART) machine learning algorithms were employed to identify and map winter wheat in a large-scale area. The hyperparameters of the three machine learning algorithms were tuned by grid search and the 5-fold cross-validation method. The classification performance of the three machine learning algorithms were compared, the results of which demonstrate that SVM achieves best performance in identifying winter wheat, and its overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), and kappa coefficient (Kappa) are 0.94, 0.95, 0.95, and 0.92, respectively. Moreover, 50 various combinations of training and validation sets were used to analyze the generalization ability of the algorithms, and the results show that the average OA of SVM, RF, and CART are 0.93, 0.92, and 0.88, respectively, thus indicating that SVM and RF are more robust than CART. To further explore the sensitivity of SVM, RF, and CART to variations of the algorithm parameters—namely, (C and gamma), (tree and split), and (maxD and minSP)—we employed the grid search method to iterate these parameters, respectively, and to analyze the effect of these parameters on the accuracy scores and classification residuals. It was found that with the change of (C and gamma) in (0.01~1000), SVM's maximum variation of accuracy score is up to 0.63, and the maximum variation of residuals is 76,215 km^2. We concluded that SVM is sensitive to the parameters (C and gamma) and presents a positive correlation. When the parameters (tree and split) change between (100~600) and (1~6), respectively, the RF's maximum variation of accuracy score is 0.08, and the maximum

variation of residuals is 1157 km^2, indicating that RF is low in sensitivity toward the parameters (tree and split). When the parameters (maxD and minSP) are between (10~60), the maximum accuracy change value is 0.06, and the maximum variation of residuals is 6943 km^2. Therefore, compared to RF, CART is sensitive to the parameters (maxD and minSP) and has poor robustness. In general, under the conditions of the hyperparameters, SVM and RF exhibit optimal classification performance, while CART has relatively inferior performance. Meanwhile, SVM, RF, and CART have different sensitivities toward the algorithm parameters; that is, SVM and CART are more sensitive to the algorithm parameters, while RF has low sensitivity toward changes in the algorithm parameters. The different parameters cause great changes in the accuracy scores and residuals, so it is necessary to determine the algorithm hyperparameters. Generally, default parameters can be used to achieve crop classification, but we recommend the enumeration method, similar to grid search, as a practical way to improve the classification performance of the algorithm if the best classification effect is expected.

1. Introduction

Wheat is one of the three major food crops across the world, providing a stable source of food and nutrition for humans [1]. In 2018, China's total wheat output was 13.14 million tons, accounting for 15.19% of the world's total output [2]. Henan province is one of the most important winter wheat production bases in China; its sown area of food crops accounted for 9.32% of the whole country in 2018, of which the winter wheat sown area was more than 23% [3]. In addition, Henan province has a large population and high levels of crop production; thus, duly and effectively mapping winter wheat across the whole province not only has great significance for local agricultural production management, but also has an impact on national wheat imports and exports, as well as the related prices.

Remote sensing technology is widely applied to many fields of agricultural production, such as crop biomass [4,5], leaf area index (LAI) [6], and yield [7], and it has been used for mapping crops [8–10]. Moreover, the way in which to utilize the remote sensing data of the growth period of crops for their quick and accurate identification has become a problem that many researchers are committed to solving. In addition to relying on the traditional spectral information, object-based image analysis (OBIA) [11], multi temporal information [12], phenology and other methods [13], more and more researchers are using machine learning algorithms for crop identification, such as support vector machine (SVM) [8], random forest (RF) [14], classification and regression tree (CART) [15], k-nearest neighbor (KNN) [16], neural networks (NN) [17], maximum likelihood (ML) [18]. Generally, these algorithms can be used for classification quickly and effectively with hyperparameters. These methods not only contribute to improving the accuracy of crop identification, but also enrich the methodologies related to the application of remote-sensing technology in agricultural fields. However, due to the complexity of remote-sensing image acquisition methods and image processing, the entire crop identification process has a long cycle and low efficiency. In addition, crop identification or land use/cover research in large-scale regions often make use of low- and medium-resolution remote-sensing images, such as Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS) [19–21]. Although they can identify crop areas and can estimate crop yield, they are susceptible to being influenced by mixed pixels, which also limits the accuracy of crop mapping [7,18,20]. Besides, Landsat, Sentinel, Spot, and other high-resolution images are usually utilized for small-scale crop classification research, while crop identification is relatively rare in large-scale regions [22–24]. Although high-resolution and multi-spectral images have higher spatial resolution and richer spectral information, traditionally, the methods of image acquisition (placing orders and downloading images) take a long time. In addition, when the desktop remote-sensing

processing software (ENVI, ERDAS, ArcGIS, etc.) is used to process images, the processing is time-consuming, which leads to the low efficiency of image processing. With Google Earth Engine (GEE) platform, the above operations can be completed with a few lines of code, which greatly improves the efficiency of data processing. These problems limit the research of high-resolution large-scale crop classification.

GEE provides a good solution to the above problems [25]. GEE is the world's most advanced cloud platform for remote sensing big data processing, which not only has efficient computing capabilities, but also assembles a large number of public geospatial data sets, including many free image resources (such as Landsat, MODIS, and Sentinel), meteorological data, land use/cover data, and population distribution data [25]. Using GEE, it is easy to conduct research pertaining to vegetation monitoring [26,27], land use/cover [28,29], water change [30], and drought analysis [31]. Moreover, lots of machine learning algorithms are built into GEE, and these have always been an important concern in crop mapping and land use/cover [8,15,29,32]. Recently, many researchers have employed machine learning algorithms to conduct research, and have achieved fruitful results. Many previous works showed that SVM is a popular non-parametric algorithm. Moreover, compared with the traditional classification algorithms maximum likehood (ML), neural networks (NN), k-nearest neighbor (KNN), SVM has better classification performance [19,33,34]. Song [35] applied SVM and artificial neural networks (ANN) to SPOT-5 image classification, and the results showed that SVM classification effect was slightly higher than ANN. Although the effects of different parameters on classification accuracy were compared in detail, the relationship between classification accuracy and final classification residual was not paid attention to. Because SVM is sensitive to the change of algorithm parameters (C and gamma), although the accuracy score is large, it may be the result of overfitting. Shao [19] compared the classification performance of SVM, NN, CART algorithms in land-classification research with MODIS data, and analyzed the accuracy change under the condition of limited training data. The results showed that compared with NN and CART algorithms, SVM has stronger generalization ability when the amount of data is small, and the highest accuracy is 83%. Similarly, although the paper notes that the parameter (C and gamma) has an impact on the classification accuracy of SVM, it does not explore how the algorithm parameters affect the accuracy change of SVM. In addition, the effect of mixed pixel produced by MODIS with medium resolution is also a factor restricting the classification accuracy of SVM. RF is also a popular classification algorithm in crop mapping and land use/cover. Many research works show that its classification accuracy is relatively stable and robust. Li [29] used RF algorithm to urban mapping with long time series Landsat images, and identified the urban development boundary. Although the number of trees can determine whether the RF is over-fitted, the split also affects the number of nodes in a single tree and the final classification result. Vuolo [36] employed RF to explore the effect of multi-temporal Sentinel-2 data on crop identification in small areas, but the research did not involve the research on RF algorithm mechanism, and the default parameters were directly applied to classification, which will have an impact on the final classification accuracy. De Alban [37] used a single classification algorithm RF to monitor land-cover changes based on GEE p, but uses fixed RF parameters for classification, and does not pay attention to the impact of RF algorithm parameters on the final classification results. Another simple and effective decision tree algorithm is CART, which has been applied to many classification studies. Johansen [26] used CART and RF algorithms to map woody plants in Queensland and Australia, and achieved good classification accuracy. However, the research did not involve the algorithm mechanism and parameter setting. The default parameters were directly used for classification, which affects the classification performance to a certain extent. Shelestov [15] employed CART, RF and other algorithms to map crops with 30 m Landsat-8 Operational Land Imager (OLI) images, and the results showed that the highest overall accuracy (OA) was 0.75. Although a variety of algorithms have been applied to crop mapping research, almost all algorithms have not involved in the discussion of classification mechanism and parameter changes, which may be an important factor causing the low classification accuracy of algorithms.

Previous studies have shown that for many machine algorithms, the algorithm parameters have an important impact on the classification results. Considering these problems, this paper employed SVM, RF, and CART to identify and map winter wheat with Sentinel-2 10 m resolution multi-spectral data, and these algorithms built in GEE are fast but effective for crops classification. Finally, in order to further research the sensitivity and performance of SVM, RF, and CART, we employed the grid search method to iterate (C and gamma), (tree and split), and (maxD and minSP), respectively. The effects of these parameters changes on the algorithm's accuracy scores and classification residuals were analyzed in detail.

2. Materials and Methods

2.1. Study Area

Henan province is in the central part of China (31°23′ N~36°22′ N and 110°21′ E~116°39′ E), with a total area of 167,000 km^2 (Figure 1). Most of Henan province is located in a warm temperate zone, with subtropical conditions in the south. Henan province belongs to a continental monsoon climate with a subtropical to temperate zone. The average annual temperature is between 10.5 and 16.7 °C, and the average annual precipitation is between 407.7 and 1295.8 mm. The rainy season covers June to August, and the annual average sunshine duration changes between 1285.7 and 2292.9 h. The favorable geographical location and climate of Henan province afford it advantageous conditions in agricultural development.
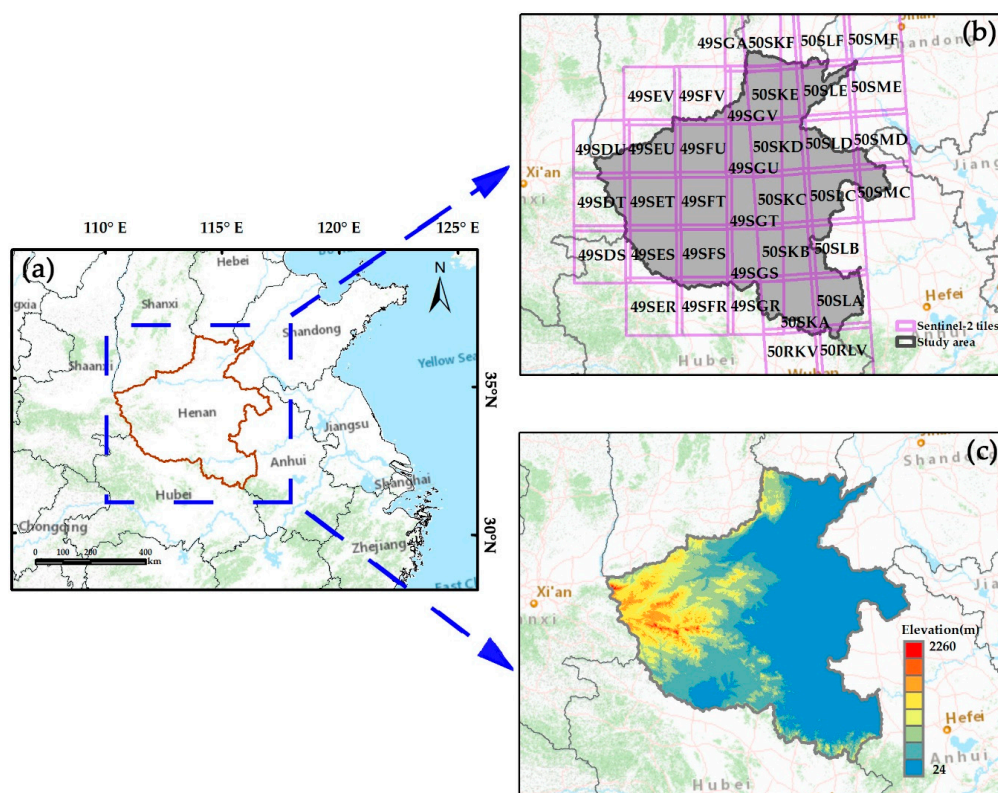

Figure 1. Location of Henan province (**a**). Sentinel 2 satellite utilizes Universal Transverse Mercator (UTM) projection and follows the US-MGRS (U.S. Military Grid Reference System) to set the satellite orbit number (**b**). The number of images covering the study area ranges from (49SD~50SM) to (50RK~50SL), with a total of 37 images. The lowest elevation in Henan province is 24 m, and the highest elevation is 2260 m (**c**).

2.2. Imagery Data and Processing

2.2.1. Sentinel-2 Data

Sentinel-2 satellites have high-resolution and multi-spectral information, and include 13 multi-spectral bands. Among them are four 10 m resolution bands, six 20 m bands, and three 60 m bands, with an orbital width of 290 km [38]. Because Sentinel-2 has multiple narrow bands in the visible and near-infrared ranges, it plays an important role in land use/cover [39], vegetation growth [40], water cover [41], and crop mapping [22].

This study utilized GEE to obtain Sentinel-2 Level 1C archived data. When selecting images, it is important to consider the specific phenological period of winter wheat, and to shorten the interval time of different orbit images as much as possible. Therefore, in order to avoid the influence of snow cover on wheat spectrum, images were selected from March to May. In this stage, the leaves and stems grow rapidly after turning green, and the chlorophyll concentration of leaves increases significantly [42]. Therefore, the spectral information of winter wheat is more prominent, which is more conducive to the identification of winter wheat (Figure 2).
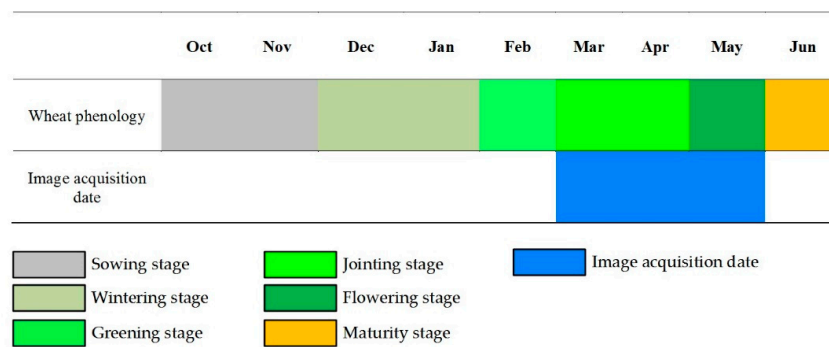

Figure 2. Wheat phenology and image acquisition date.

After selecting the image collection with the smallest cloud cover from February to May, the quality band (QA60) (Table 1) can be used for the cloud mask to remove pixels with cloud cover, and finally to composite high-quality images without clouds. Considering the spectral difference and variation of different land covers in different bands, and to more accurately distinguish the difference between winter wheat and other vegetation, urban, water and other classes, all bands were input to SVM, RF, CART classification algorithm.

Table 1. Spectral bands of the Sentinel-2 imagery.

Band	Center Wavelength (nm)	Resolution (m)	Spectral Region
B1	443	60	Coastal Aerosol
B2	490	10	Blue
B3	560	10	Green
B4	665	10	Red
B5	705	20	Red edge1
B6	740	20	Red edge2
B7	783	20	Red edge3
B8	842	10	Near infrared (NIR)
B8A	865	20	Red edge4
B9	940	60	Water vapor
B10	1375	60	Cirrus
B11	1610	20	Short -wave infrared spectral range 1 (SWIR1)
B12	2190	20	Short -wave infrared spectral range 2 (SWIR2)
QA10	—	10	—
QA20	—	20	—
QA60	—	60	Cloud mask

2.2.2. Sample Data

It is necessary to determine the number of sample points for crop mapping. We determined the total number of samples with Equation (1) [37,43], and the number of sample data involves the expected accuracy and the uncertainty of estimation.

$$n = \frac{z^2 O(1 - O)}{d^2} \tag{1}$$

where *n* is the total sample size, *z* is the percentage of the standard normal distribution, *O* is the expected overall accuracy, and *d* is the error margin. If the aim is to obtain 80% overall accuracy, *z* = 1.96 at 95% confidence interval, and the error margin is 1.5%. Therefore, the total sample is 2731. We know that in order to achieve 80% overall accuracy, at least 2731 sample points are needed to avoid potential bias [44,45]. Meanwhile, the sample points of each class should be consistent with the actual land covers. Owing to these sample data being unable to obtain a good classification accuracy, we constantly adjust the number of different classes of sample points to reduce the misclassification as much as possible, and finally we determined the number of each class in Table 2.

Table 2. Sample data selected in this study.

Land Cover	Abbreviated	Description	Samples
Wheat	WHE	Wheat pixels during the observation period	910
Vegetation	VEG	Other crops, urban greening, and other evergreen vegetation	900
Urban	URB	Land for construction, such as for houses and roads	480
Water	WAT	Water bodies such as rivers, lakes, and swamps	210
Others	OTH	Bare, idle, or deserted land during the observation period	290
Total			2790

The wheat, vegetation, urban, water, and others classes (Table 2) are labeled in GEE, and all sample points are selected by visual interpretation. Since different bands of images have various spectral information, a combination of different bands can better highlight feature information [46]. The optimal sample points are selected after comparing the Sentinel-2 false color (B8-B4-B3\B9-B4-B3) and true color (B4-B3-B2) images. Finally, all of the selected sample data are compared to the high-resolution image in Google Earth (GE). If the selected sample label is not matched with the actual feature in GE, the sample is deleted to ensure the accuracy and reliability of the selected sample. With machine learning algorithms, the sample data is usually split into training data sets and validation data sets. This facilitates the usage of the available sample data as efficiently as possible in the absence of independent validation data. In addition, the model is trained by training data, and validated by unseen validation data, which can test the classification performance of algorithms in the face of unseen data. In order to split data into two sets, the *randomColumn* function of GEE was used to generate a random number in all sample points, with the random values ranging from (0~1). Therefore, all of the sample data generate an extra random value. If sample values were ≥0.7, they were taken as the validation set, and the sample with values <0.7 as the training set. Finally, the training set was used for training, and the validation set was utilized to validated the performance of algorithms.

2.3. Overall Workflow

The overall workflow is shown in Figure 3, which includes the following four parts: Image preprocessing, image classification, classification results, and sensibility analysis. Image preprocessing includes filtering Sentinel-2 data in a specific phenological period based on the GEE, and obtaining

high-quality cloudless images in the study area through steps such as minimum cloud cover, cloud mask, splicing, and clipping. For image classification, three machine learning algorithms, namely, SVM, RF, and CART, were employed. After dividing all sample data into training and validation sets, the training set served as the input data for these algorithms, and the validation set was used to validate the classification accuracy. In the classification results section, the three classification results were compared to the officially published wheat area data in order to analyze the classification differences. The OA, user's accuracy (UA), producer's accuracy (PA), and kappa coefficient (Kappa) were calculated by error matrixes, and these classification assessment indicators were compared to the three machine-learning algorithms. Meanwhile, to explore the generalization ability and the mechanisms of the three algorithms, different training and validation data sets were created and inputted into the three algorithms for classification. Finally, those algorithms with robust classification effects and generalization ability were chosen. In the sensitivity analysis section, various algorithm parameters were iterated to explore their effect on the accuracy score and classification residuals.
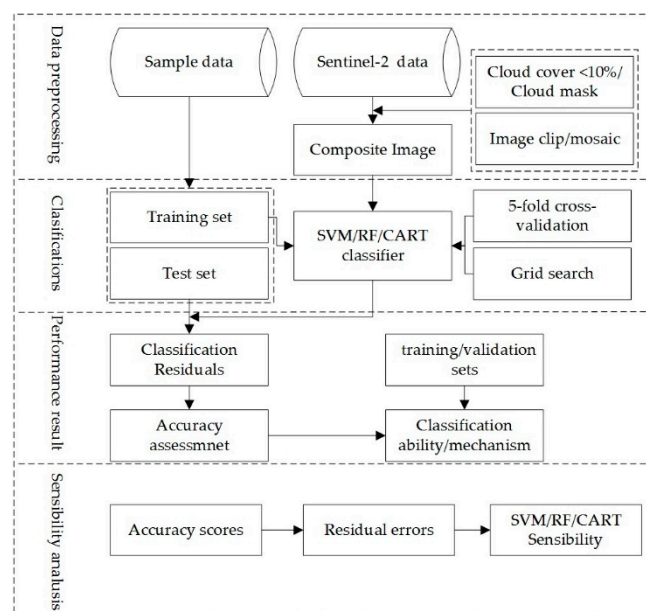

Figure 3. Overall workflow of this study.

2.4. Machine-Learning Algorithms

Based on the GEE platform, Sentinel-2 hyper-spectral images were utilized for processing. A large number of high-quality samples were selected by visual interpretation, and the three popular machine-learning algorithms of SVM, CART, and RF were employed to identify and map winter wheat in a large-scale region.

$$y_i(wx_i + b) \geq 0 \tag{2}$$

where x_i is the point lying on the boundary of hyperplane, w is the weight vector and b is called the bias.

SVM is a classification algorithm based on statistical learning theory, which was first proposed by Vapnik [47]. The basic theory is to find the optimal hyperplane of the feature space in order to determine the maximum interval between different classes (Equation (2)) [47]. The hyperplane of SVM is defined by Equation (2) [47], and these points on the boundary of the hyperplane are called support vectors, which directly affect the final classification performance. Owing to this feature, many previous works have proved that SVM performs well in small data and multi-feature data sets [33,48].

At present, it is very common to employ SVM to conduct land cover research, crop identification, and classification [49–52].

$$G = \sum\sum_{j\neq i}(f(C_i,T)/|T|)\big(f\big(C_j,T\big)/|T|\big) \tag{3}$$

where T is a given training set, $f(C_i,T)$ is the probability that the selected case belongs class C_i.

CART is a tree-based machine learning algorithm, which Breiman discussed in detail [53]. It is a non-parametric computationally intensive algorithm that can be used for both classification and regression [54]. The basic mechanism is based on a split criterion to perform binary recursive classification on data sets, which can process continuous or discrete attributes as targets or predictors [53]. The Gini index defined by Equation (3) is used to select the feature at each internal node of the decision tree [55]. CART is a popular machine-learning algorithm, which is also widely used in disaster monitoring [56], vegetation growth [26], and crop classification [15].

The RF algorithm was first proposed by Breiman [57]. Its essence is that a classifier is ensembled by many decision tree classifiers $\{h(x,\Theta_k), k = 1, \ldots,\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and x is an input pattern [57,58]. Each tree votes on the results, and the classification result with the most votes is chosen as the final classification result. Moreover, the Gini index is chosen for RF, owing to its simplicity [55]. Because RF assembles many trees as the basis for classification, the more independent the trees are, the smaller the generalization error is [57]. Owing to the fact that RF has high classification efficiency, even though the number of samples is large, increasingly researchers employ RF to carry out related research in agricultural fields [9,50].

3. Results

3.1. Classification Results and Accuracy Assessment

3.1.1. Classification Results

This study selected high-quality images of the specific phenological period of winter wheat growth. Furthermore, the winter wheat, vegetation, urban, water, and others' class were labeled and mapped in GEE (Figure 4). The detail codes are accessible from the Code link (see Supplementary Materials).
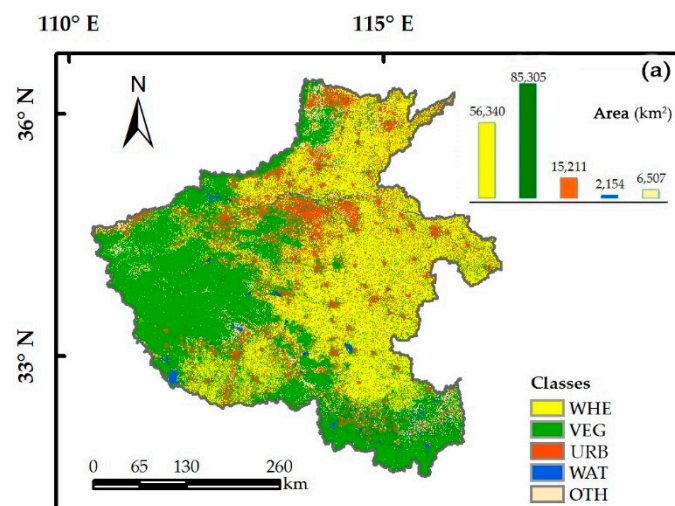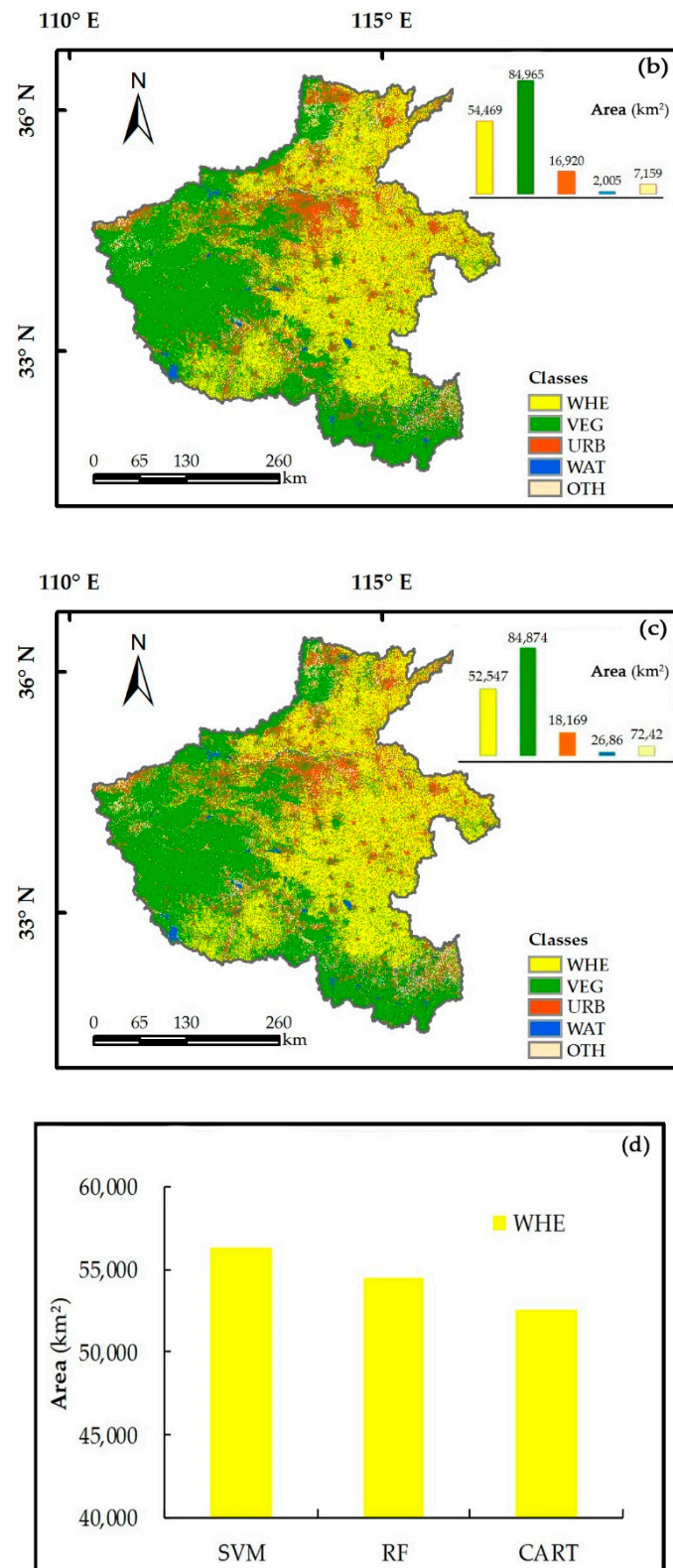

Figure 4. *Cont.*

Figure 4. The classification results of the study area in 2019. (**a**–**c**) The classification results of support vector machine (SVM), random forest (RF), and classification and regression tree (CART), respectively. (**d**) Winter wheat areas of the three machine learning algorithms. WHE = Wheat, VEG = Vegetation, URB = Urban, WAT = Water, OTH = Other.

To compare the differences in areas of each class identified by SVM, RF, and CART, these areas were calculated (Figure 4). Meanwhile, in order to validate the accuracy of the three algorithms in identifying winter wheat pixels, the statistics of the official yearbook of winter wheat area were compared.

The area of winter wheat identified by the three algorithms was 56,340 km^2 for SVM (34% of the study area), 54,469 km^2 for RF (33% of the study area), and 52,547 km^2 for CART (31% of the study area). Meanwhile, the official statistics of Henan province in 2019 were 57,066 km^2 [59]. By calculating winter wheat area residuals of the three algorithms, the following pattern was observed: CART (4519 km^2 residuals) > RF (2597 km^2 residuals) > SVM (726 km^2 residuals). The most accurate identification of the winter wheat was given by SVM, which has the lowest residual error of 726 km^2, followed by RF, for which the residual error compared to the actual area is 2597 km^2. The largest residual error was for CART with 3862.40 km^2. Moreover, we found that CART was able to identify the most urban and others' class pixels, while SVM and RF identified relatively few.

3.1.2. The Classification Difference of Regions

In order to more specifically show the classification differences of SVM, RF, and CART in different regions and classes (i.e., Wheat (WHE)- Other (OTH), WHE-Urban (URB), WHE- Vegetation (VEG), and WHE-Water (WAT)), four 5 × 5 km sub-regions (i.e., a, b, c, and d) were randomly selected from the classification results of the three algorithms to discuss the misclassification of winter wheat pixels (Figure 5).

By carefully comparing the mapping results of SVM, RF, and CART in the four sub-regions, it is obvious that the three algorithms are affected by the other classes when identifying winter wheat, resulting in misclassification. In region a, the spectrum of vegetation is more similar to winter wheat than the other classes, resulting in misclassification. Compared to the RF and CART algorithms, SVM can effectively distinguish vegetation and winter wheat pixels. In region b, we analyzed the differences between winter wheat and urban pixels. It can be seen from the figure that the three algorithms can distinguish the urban pixels well, but there are differences when classifying roads with vegetation growth. SVM and RF can effectively classify wheat and urban pixels, while CART classifies such pixels as winter wheat pixels. It is difficult for CART to distinguish winter wheat pixels from winter wheat pixels when classifying mixed pixels. In region c, we analyzed the misclassification of winter wheat and water pixels. Obviously, due to the strong absorption of solar radiation in the visible and near-infrared bands, the spectrum of water is significantly different from that of winter wheat. Therefore, the three algorithms can distinguish the boundary between water bodies and winter wheat well. For region d, it is obvious that SVM is more accurate than RF and CART in classifying wheat class and other class.

Through the analysis of the misclassification of the four different sub-regions, we determined that the SVM and RF algorithms can distinguish winter wheat pixels from the other classes. However, RF is not as effective as SVM in identifying vegetation and winter wheat pixels. According to Figure 5, it can be found that except for water bodies, vegetation and urban pixels could not be accurately classified. Compared to SVM and RF, the classification ability of CART is relatively inferior. This indicates that SVM and RF have strong performance in classifying winter wheat from the other classes in regions with complex classes.
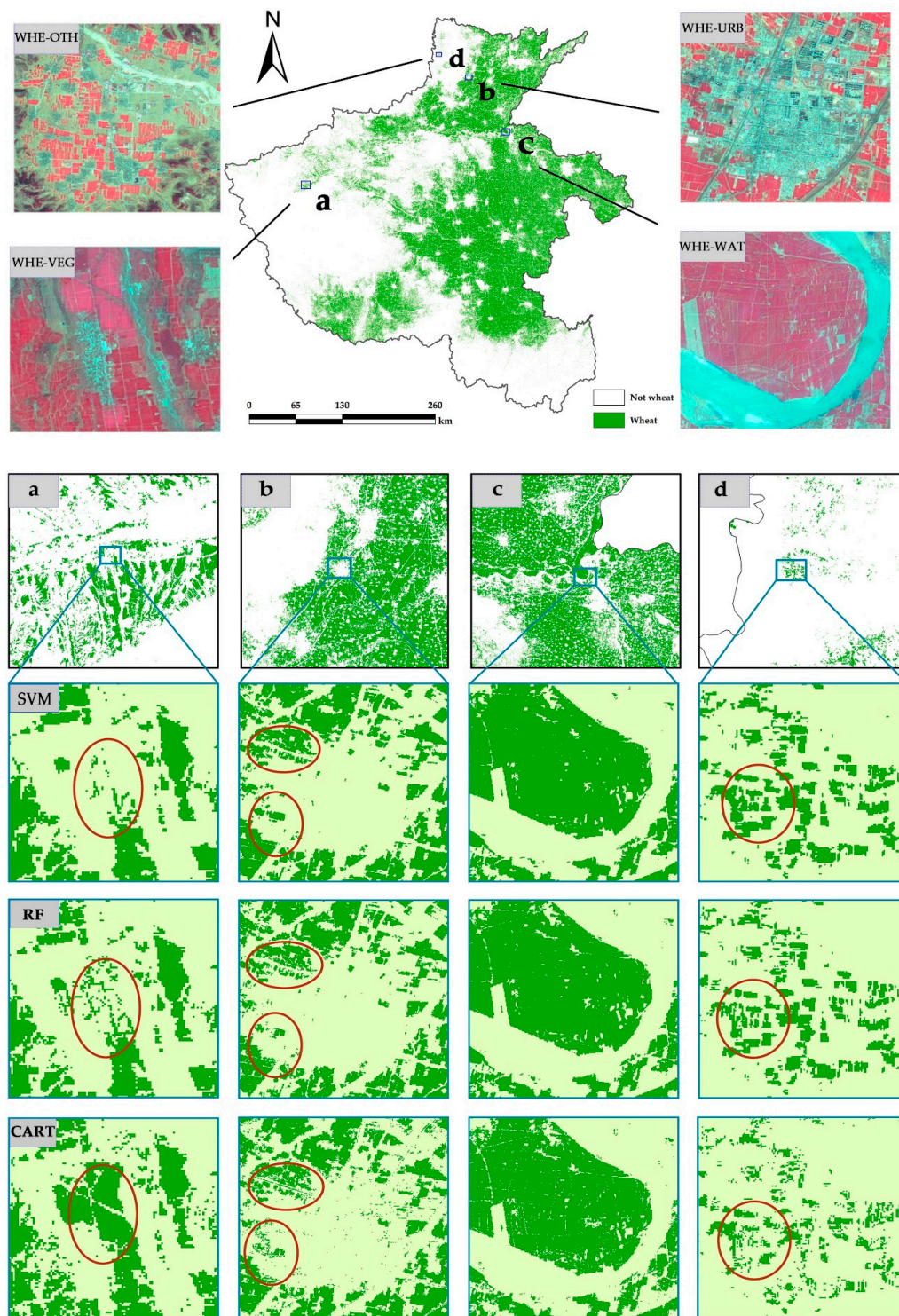
Figure 5. Winter wheat mapping in different regions using the three machine-learning algorithms. (**a–d**) The different regions of this study area.

3.1.3. Accuracy Assessment

In Section 3.1, based on the statistics of the winter wheat area under different machine-learning algorithms, we determined that the residual error of winter wheat area identified by the SVM algorithm is the smallest. In Tables 3–5, the error matrix of the three algorithms was used, and the OA, UA, PA, and Kappa indicators were calculated to compare the classification accuracy of the three algorithms.

Table 3. The error matrix of SVM. WHE = Wheat, VEG = Vegetation, URB = Urban, WAT = Water, OTH = Other, OA = Overall accuracy, UA = User's accuracy, PA = Producer's accuracy,.

Classification	Reference						
	WHE	**VEG**	**URB**	**WAT**	**OTH**	**Total**	**UA**
WHE	254	13	0	0	0	267	0.95
VEG	10	245	4	1	2	271	
URB	1	3	143	0	6	153	
WAT	1	7	1	58	0	67	
OTH	0	0	3	0	84	87	
Total	266	268	151	59	92	836	
PA	0.95						
OA	0.94	Kappa	0.92				

Table 4. The error matrix of RF.

Classification	Reference						
	WHE	**VEG**	**URB**	**WAT**	**OTH**	**Total**	**UA**
WHE	250	16	0	0	0	266	0.94
VEG	13	253	5	1	1	273	
URB	2	10	127	0	4	143	
WAT	2	4	0	52	0	58	
OTH	0	1	0	0	90	91	
Total	267	284	132	53	95	831	
PA	0.94						
OA	0.93	Kappa	0.90				

Table 5. The error matrix of CART.

Classification	Reference						
	WHE	**VEG**	**URB**	**WAT**	**OTH**	**Total**	**UA**
WHE	249	24	0	0	0	269	0.93
VEG	19	233	11	6	4	276	
URB	1	8	127	0	10	146	
WAT	0	5	0	55	1	61	
OTH	0	6	8	0	66	81	
Total	273	273	146	61	80	833	
PA	0.91						
OA	0.87	Kappa	0.83				

By observing the above tables, it can be seen that the OA, UA, and PA of the three classifiers are all above 80%, and the OA values of SVM and RF are above 90%. Therefore, these machine learning algorithms all achieve good classification accuracy. However, SVM and RF are more robust than CART.

The value of OA is the ratio of the number of correctly classified samples to the total number, which can directly reflect the classification accuracy. The OA values of SVM, CART, and RF are 0.94, 0.93, and 0.87, respectively. This indicates that the number of correctly classified pixels of SVM is the largest, and OA is slightly higher than other two algorithms. Since the main purpose of this paper is to validate the accuracy of the three machine learning algorithms in identifying winter wheat, the UA and PA of winter wheat were calculated. For UA, the values of SVM, RF, and CART are 0.95, 0.94, and 0.93, while for PA, they are 0.95, 0.94, and 0.91, respectively. Thus, the classification accuracy of SVM is slightly higher than that of CART and RF. It can be found the values of the UA and PA of SVM have the largest values, which indicates that SVM can correctly classify winter wheat pixels with the largest proportion and has better classification performance compared to CART and RF. Kappa is a statistical parameter used to test the consistency of classifications, and it is able to take into account the

differences caused by different proportions of classes. According to the above tables, the Kappa values of SVM, CART, and RF are 0.92, 0.90, and 0.83, respectively, indicating that the Kappa of SVM is also slightly higher than that of the other two algorithms.

The above analysis concludes that the values of the OA, UA, PA, and Kappa of SVM and RF are higher than that of CART. In brief, based on the existing sample data, SVM and RF have stronger generalization ability, better robustness, and the best performance in this classification.

3.2. The Performance and Mechanism of Support Vector Machine (SVM), Random Forest (RF), and Classification and Regression Tree (CART)

In order to further explore the classification performance of SVM, RF, and CART using the existing sample data, this study created a training set (70%) and a validation set (30%). Fifty different random seeds were generated, which were combined into 50 groups of different training sets and validation sets. Furthermore, changes in the various accuracy indicators under different sample data for each algorithm were analyzed (Figure 6).
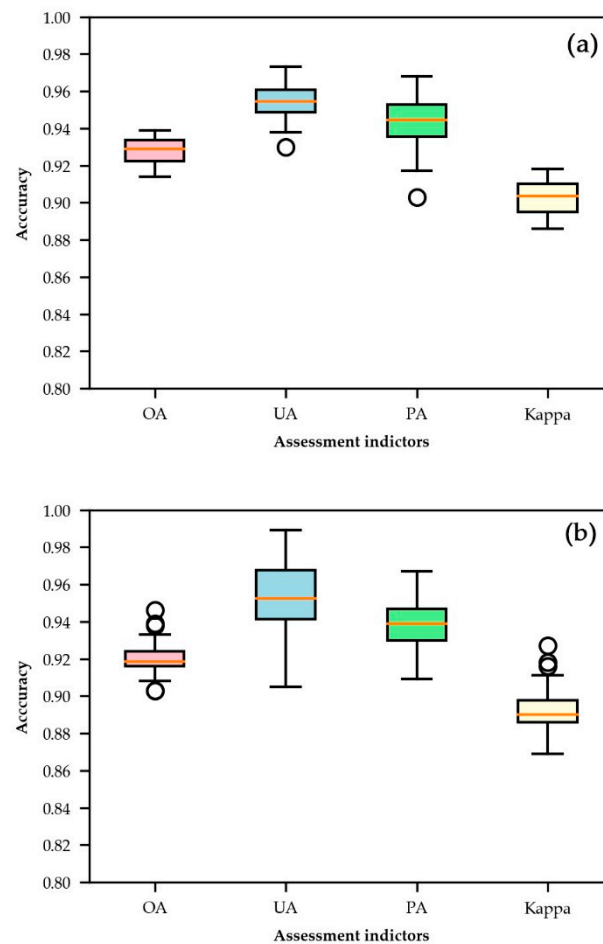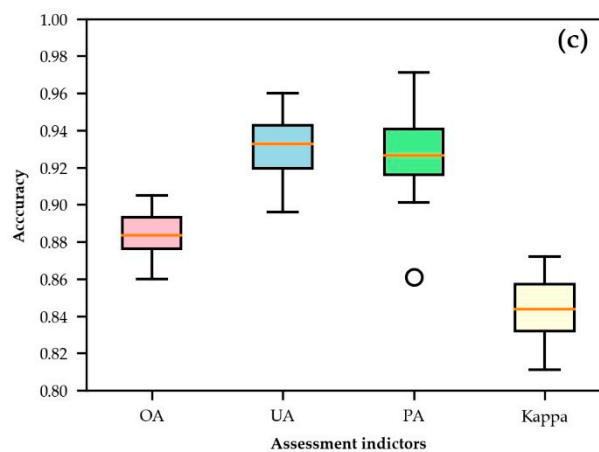

Figure 6. *Cont.*

Figure 6. Changes in various assessment indicators with SVM, RF, and CART. (**a**) The result of SVM;
(**b**) the result of RF; (**c**) the result of CART. See Table S1 and Table S2 for detail.

In Figure 6, although the values of the assessment indicators calculated by the three algorithms are varied, all values of the three algorithms are large (all values are above 0.8). It can be concluded that the three machine learning algorithms have better performance in winter wheat mapping. However, compared to CART, the four assessment indicators of SVM and RF are higher. Specifically, the average OA values of SVM, RF, and CART are 0.93, 0.92, and 0.88, respectively. Meanwhile, the average UA values are 0.95, 0.95, and 0.93, and the average PA values are 0.94, 0.94, and 0.93, respectively. Finally, the average Kappa values are 0.90, 0.89, and 0.84, respectively. Obviously, through inputting the 50 different training and validation set combinations into the three algorithms, it is shown that when training the different data sets, both the SVM and RF algorithms achieve higher classification accuracy, and the interquartile range values of the four classification indicators are more concentrated. From a statistical point of view [60], the accuracy values obtained by SVM and RF have stronger generalization ability. In comparison, the interquartile range of CART is larger and the box body is longer, indicating that with the same sample data, the variation of CART is relatively large and the algorithm is not robust.

During classification, the optimal hyperplane of SVM serves as a condition to classify different data labels. The way in which the optimal hyperplane and kernel function are determined is an important factor for the classification results. When processing binary classification data, linear classification can be utilized to obtain better classification accuracy. When the feature dimension is multi-dimensional, non-linear classification is required. The cost parameter C should be considered in both linear and non-linear classifications. The value of the parameter C directly affects the maximization of the decision boundary of SVM, thereby affecting the number of misclassified pixels [19,48]. In addition, the choice of kernel function also affects the classification accuracy of SVM. SVM provides linear kernel function, polynomial kernel function, radial kernel function, etc. For this classification, we chose the classification radial basis function (RBF) $k_{RBF}(x, x') = \exp(-gamma\|x - x'\|_2^2)$. This kernel function has been used in many classification research works and has achieved good classification results [19]. The choice of kernel function affects the classification performance of SVM, and the value of the kernel function parameter gamma has an impact on the classification accuracy. If gamma is small, the decision boundary is close to linear, which affects the non-linear classification data; however, when gamma is large, it causes overfitting [48]. In order to obtain the hyperparameters (C and gamma), this research utilized 5-fold cross-validation and the grid search method to tune the parameters, based on the Python scikit-learn library. Finally, we obtained the hyperparameters C:10 and gamma:10.

Unlike the SVM algorithm mechanism, the CART and RF algorithm are both tree-based algorithms. CART is a decision-tree algorithm that can be used for classification and regression. During classification, CART recursively divides the feature space based on the purity of the nodes until the number of samples is less than a predetermined threshold, or the number of the samples in the nodes is homogeneous [53,58].

Since there are many classification features, the whole classification may produce a large number of child nodes. The information contained in these child nodes may be noisy and may thus affect the classification results. Therefore, the classification tree is pruned to improve the classification accuracy. In the growth of the classification tree, the tree growth depth (maxD) and the minimum samples of split (minSP) affect the classification results [53,61]. Generally, default parameters can be selected to achieve classification results. In this study, we obtained the hyperparameters maxD: 20 and minSP: 10.

For RF, the algorithm assembles many decision trees. In the classification process, each tree votes on the classification results, and then the classifier selects the result with the most votes as the final classification result [50,57] In addition, RF employs bootstrap for randomly sampling. Each node randomly selects a subset of features to split. By following these steps, randomness can be achieved to the greatest extent. In addition, the number of RF-assembled trees (tree) and the largest split node feature (split) will impact the classification results [9,22]. Consequently, the hyperparameters for RF are tree: 300 and split: 3.

3.3. The Sensitivity Analysis of SVM, RF, CART

The above analysis demonstrates that the algorithm parameters are crucial for the classification performance of the three machine-learning algorithms. In order to further research the sensitivity of SVM, RF, and CART, this study employed the grid search method to iterate (C and gamma), (tree and split), and (maxD and minSP), respectively, and analyzed the effects of these parameter changes on the algorithm accuracy scores and classification residuals. We set the parameters {C: (102~103), gamma: (102~103}, {tree: 100~600, split: 1~6}, {maxD: 10~60, minSP: 10~60} [19,22,48]. SVM, RF, and CART were iterated with the corresponding algorithm parameters to obtain the accuracy scores under the different parameters (Figure 7a,c,e) and classification residual (Figure 7b,d,f).

By analyzing the sensitivity of SVM toward (C and gamma) (Figure 7a,b), it can be seen that as the values of (C and gamma) gradually increase, the accuracy score increases, and the residuals decrease. When the values of (C and gamma) are less than (1 and 1), the accuracy scores are small. In particular, when the (C and gamma) values are (0.01 and 0.01), (0.01 and 0.1), (0.01 and 1000), and (0.1 and 0.01), the accuracy scores all are 0.323, which are the smallest scores. In Figure 7b, as the (C and gamma) increase gradually, the residual values decrease accordingly. When the values of (C and gamma) are (0.01 and 0.01), (0.01 and 0.1), (0.01 and 1000), and (0.1 and 0.01), the maximum residuals are 76,700 km^2. When the values of (C and gamma) are (1000 and 1000), the residuals are 1540 km^2. When parameter C is small, the penalty for error is relatively large [19,48], which results in the phenomenon of underfitting, leading to large classification residuals. With the increase in gamma values, although the decision boundary becomes smoother and individual features can be better fitted, overfitting may also occur [48]. When (C and gamma) are (1 and 100), the accuracy score is 0.95. In general, when the values of (C and gamma) are small, the error penalty for the classified data is larger, and the accuracy scores are generally smaller. As the values of (C and gamma) increase, so does the accuracy score. When the values of (C and gamma) exceed (1 and 1), the accuracy score gradually tends to be stable. In general, SVM is sensitive to parameters (C and gamma).

In Figure 7c,d, RF is not sensitive to the parameters (tree and split); the smallest accuracy score of (tree and split) is 0.932, while the largest score is 0.940, with a small variation. In Figure 7d, the maximum value is 2859 km^2 and the minimum value is 1702 km^2, and the variation range of the residuals is relatively small. RF is an assemble algorithm based on a decision tree, which integrates multiple weak classifiers into a strong classifier. Multiple weak classifiers vote on the final classification result to obtain the optimal result [9,50]. Therefore, the classification results of RF are generally stable when sample data are sufficient. In this study, under the control of parameters (tree and split), the number of trees have a small impact on RF, while splits have a large impact. The split determines the number of features required to split; when the value of the split is small, each tree grows deeper and overfitting occurs easily. Therefore, controlling the split value is conducive to improving the classification performance of RF.
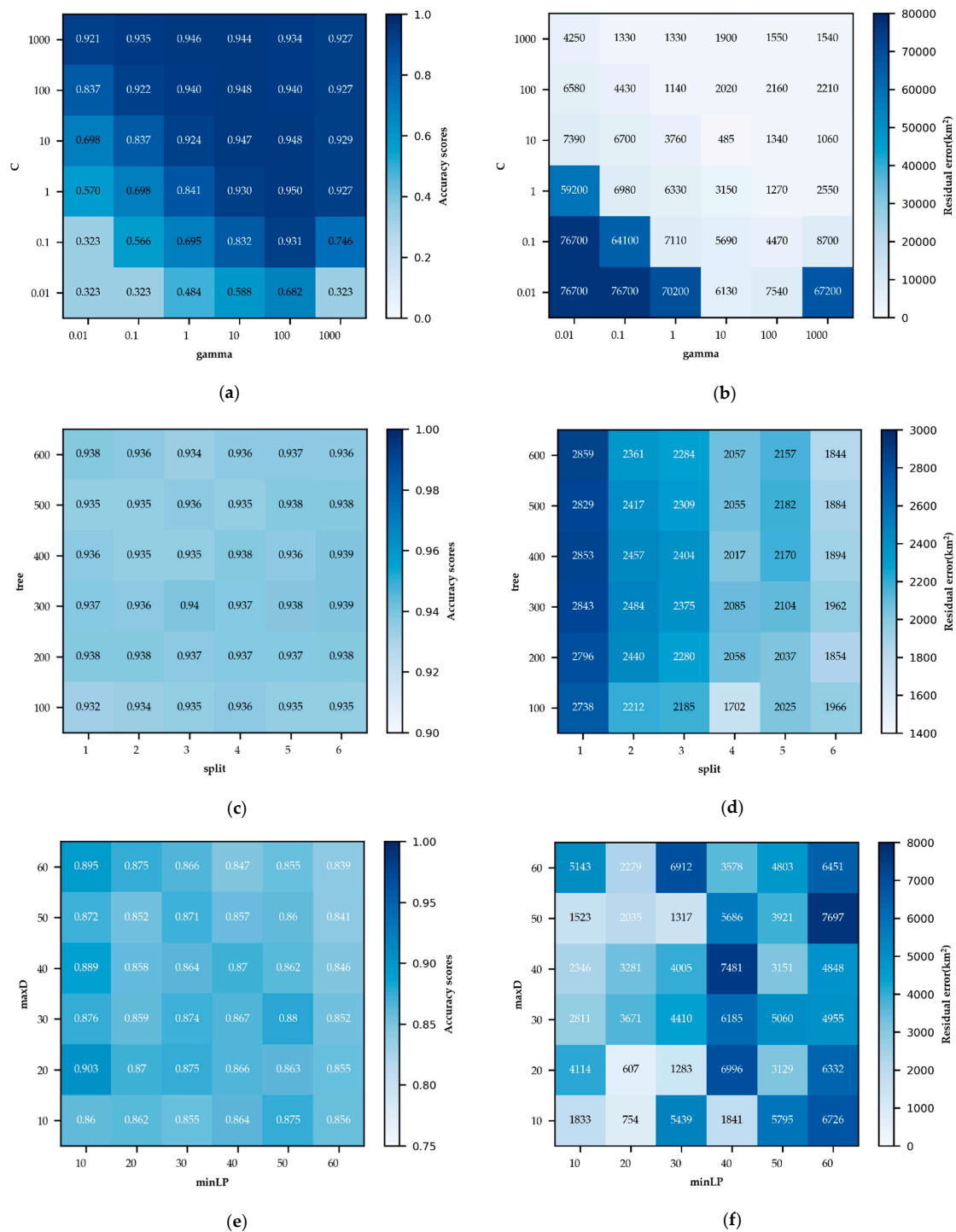
Figure 7. The accuracy scores and residual errors of SVM, RF, and CART with different algorithm parameters. (**a**,**c**,**e**) The accuracy score chart; (**b**,**d**,**f**) the residual error chart.

Compared to RF, CART is more sensitive to changes in the algorithm parameters (maxD and minSP). In Figure 7e, the accuracy score decreases gradually as (maxD and minSP) increase slightly. The maximum accuracy score is 0.903 and the minimum is 0.841. Correspondingly, in Figure 7f, we can see that as (maxD and minSP) increase, the residuals also increase. This indicates that changes in the algorithm parameters (maxD and minSP) directly affect the classification performance of CART. MaxD is the growth depth control parameter of the tree; when the value of maxD is large, it is easy to ignore the common characteristics of the data and to result in overfitting, although it trains the details of the

data better. The minSP is the minimum number of samples required for splitting the nodes; the smaller the minSP values, the lower the number of split nodes. In Figure 7f, when maxD < 60 and minSP < 30, the classification performance of CART is relatively good.

According to the above analysis, we conclude that SVM, RF, and CART have different sensitivities toward the algorithm parameters. SVM is more sensitive to parameters (C and gamma). Generally, there is a positive correlation between the accuracy score and the values of (C and gamma). RF has a lower sensitivity toward value variation, and the split effect is relatively stable. Under the control of the parameters (tree and split), the split value has a greater impact on classification performance. Finally, CART has strong sensitivity toward (maxD and minSP). We found that accuracy values are generally better when (maxD and minSP) are small. However, the classification performance of the CART algorithm is not stable, and its robustness is lower than that of RF.

4. Discussion

Machine learning algorithms are widely used in agricultural fields such as land use/cover, crop identification and mapping, etc. However, the effect of parameters setting on their classification performance has not received enough attention. Many works have shown that when SVM is used for classification, the selection of kernel function and the value of parameters will have an important impact on the classification performance [33,48,62,63]. The small parameter C tends to emphasize hyperplane margin while ignoring outliers, while the larger parameter C tends to overfitting easily [49]. Parameter gamma control kernel width, and large values of gamma also lead to overfitting. For tree-based machine learning algorithms, RF and CART, the algorithm parameters also have an important impact on the classification performance. RF is classification algorithms that assembles many decision trees. Each tree votes the classification results and selects the result with the most votes as the final classification result. Therefore, RF shows more stable classification results and stronger robustness in many classification studies [64,65]. Generally, the number of trees affects the accuracy of RF classification. When the tree value is large, the classification results tend to overfitting. In the study, other RF parameters (spit) also affect the classification performance, but it is usually ignored in research. CART is a single decision tree algorithm, which divides the data features into binary recursion, and finally obtains the classification results. Similarly, the setting of algorithm parameters (maxD, minSP) is easy to ignore.

How to determine the appropriate C and gamma of SVM is not a problem that cannot be ignored, and also an important basis for correct research conclusions. This paper comprehensively discussed the classification mechanism of SVM, RF and CART, and explored the performance of SVM, RF, and CART to variations of the algorithm parameters. We found that the classification accuracy of SVM is better than that of RF and CART. This is because SVM has a unique classification advantage in the face of multiple features, and previous studies have also confirmed this point. However, SVM is sensitive to the change of algorithm parameters. Different C and gamma values affect the accuracy scores and residuals of the algorithm. Generally, If the C and gamma are small, the model tends to be underfitting, while if it is large, it tends to overfitting. In addition, for RF and CART algorithm, the classification performance of RF is better than CART, because a multi-trees ensemble algorithm has stronger robustness than a single decision tree. This point is also consistent with previous studies. However, RF is less sensitive to parameter changes. Under the joint control of tree and split, the effect of split's values on RF is greater than tree's values. This is because the final classification effect of RF depends on each tree, while the growth of a single tree is affected by the values of split. For CART algorithm, we found that it is also sensitive to the change of algorithm parameters. Moreover, under the control of parameters (minD, minSP), the values of minSP have a greater impact on the performance of CART. In conclusion, this study found that the classification performance of the three algorithms were affected by the algorithm parameters, although different algorithms have different effects. This further shows that it is necessary to pay attention to the effect of algorithm parameters when using a machine-learning algorithm to perform classification research. Generally, the default

parameter value does not help to improve the accuracy of algorithms, so the optimal result can be obtained by tuning parameters properly.

There are still some uncertainties in this study, which can be further explored in the future. Cross validation is widely used in machine-learning algorithms, and its advantage is to evaluate the classification performance of prediction models for unknown data. However, when using this method, we should also pay attention to its limitations. Juan D, etc., [66,67] comprehensively discussed the k-fold cross validation method. When the data set is large, a small k value (5 or less) will reduce the variance of the prediction model, and reduce the calculation time of model, so as to improve the classification efficiency. Moreover, a small value of k trends to increase the bias of the predictor. In this study, 5-fold cross validation was employed to achieve good results, but the effect of the number of k on the performance of the model was not discussed in depth. Therefore, it can be further discussed in future research. In addition, many works have shown that, except for the influence of algorithm parameters, the source of sample data, the number of samples, the division proportion of training/validation set, etc., will affect the final classification performance. Based on the existing data, this paper comprehensively discusses the influence of algorithm parameters on the classification performance. However, many uncertain factors mentioned above should be further studied in the future.

5. Conclusions

Crop mapping is crucial for agricultural production management and food security, and machine-learning algorithms such as SVM, RF, and CART provide important support for this purpose. Based on the GEE platform, we utilized Sentinel-2 10 m resolution multi-spectral images, and combined them with the SVM, RF, and CART algorithms to identify and map winter wheat in large-scale areas. By analyzing the classification results of the three machine-learning algorithms, the value of the OA for SVM is 0.95, and it achieves the best classification performance. In addition, the sensitivity of SVM, RF, and CART were discussed in detail. This research demonstrates that, in general, SVM is more sensitive to parameters (C and gamma); the classification accuracy is optimal when these hyperparameters are obtained. RF is less sensitive to (tree and split), and it is more robust. Compared to RF, CART is more sensitive to parameters (maxD and minSP), with unstable classification performance and inferior robustness. In this research, considering the algorithms' complexity, SVM, RF, and CART, three simple but efficient algorithms were employed for winter wheat identification and mapping. In future research, more machine and deep learning algorithms should be applied to crop identification. However, no matter which algorithms we choose, we should pay more attention to the algorithms' internal mechanism, which is more conducive to improving the accuracy of classifications. Meanwhile, the enumeration method, similar to grid search, should be employed to fine-tune the hyperparameters when discussing the use of various algorithm parameters.

References

1. Ren, S.; Qin, Q.; Ren, H. Contrasting wheat phenological responses to climate change in global scale. *Sci. Total Environ.* **2019**, *665*, 620–631. [CrossRef] [PubMed]

2. FAOSTAT. Crops. Available online: http://www.fao.org/faostat/en/#data/QC (accessed on 7 March 2020).

3. HPBS. Henan Statistical Yearbook 2019. Available online: http://www.ha.stats.gov.cn/hntj/lib/tjnj/2018/indexch.htm (accessed on 2 March 2020).

4. Zheng, G.; Chen, J.M.; Tian, Q.J.; Ju, W.M.; Xia, X.Q. Combining remote sensing imagery and forest age inventory for biomass mapping. *J. Environ. Manag.* **2007**, *85*, 616–623. [CrossRef] [PubMed]

5. Lu, D. The potential and challenge of remote sensing-based biomass estimation. *Int. J. Remote Sens.* **2006**, *27*, 1297–1328. [CrossRef]

6. Yang, G.; Liu, J.; Zhao, C.; Li, Z.; Huang, Y.; Yu, H.; Xu, B.; Yang, X.; Zhu, D.; Zhang, X.; et al. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci.* **2017**, *8*, 1111. [CrossRef]

7. Franch, B.; Vermote, E.F.; Skakun, S.; Roger, J.C.; Becker-Reshef, I.; Murphy, E.; Justice, C. Remote sensing based yield monitoring: Application to winter wheat in United States and Ukraine. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *76*, 112–127. [CrossRef]

8. Aneece, I.; Thenkabail, P. Accuracies achieved in classifying five leading world crop types and their growth stages using optimal earth observing-1 hyperion hyperspectral narrowbands on google earth engine. *Remote Sens.* **2018**, *10*, 2027. [CrossRef]

9. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [CrossRef]

10. Zhang, X.; Qiu, F.; Qin, F. Identification and mapping of winter wheat by integrating temporal change information and Kullback–Leibler divergence. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *76*, 26–39. [CrossRef]

11. Aguilar, M.A.; Vallario, A.; Aguilar, F.J.; Lorca, A.G.; Parente, C. Object-based greenhouse horticultural crop identification from multi-temporal satellite imagery: A case study in Almeria, Spain. *Remote Sens.* **2015**, *7*, 7378–7401. [CrossRef]

12. Vega, F.A.; Ramírez, F.C.; Saiz, M.P.; Rosúa, F.O. Multi-temporal imaging using an unmanned aerial vehicle for monitoring a sunflower crop. *Biosyst. Eng.* **2015**, *132*, 19–27. [CrossRef]

13. Pan, Z.; Huang, J.; Zhou, Q.; Wang, L.; Cheng, Y.; Zhang, H.; Blackburn, G.A.; Yan, J.; Liu, J. Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 188–197. [CrossRef]

14. Tatsumi, K.; Yamashiki, Y.; Torres, M.A.C.; Taipe, C.L.R. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Comput. Electron. Agric.* **2015**, *115*, 171–179. [CrossRef]

15. Shelestov, A.; Lavreniuk, M.; Kussul, N.; Novikov, A.; Skakun, S. Exploring Google earth engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. *Front. Earth Sci.* **2017**, *5*, 17. [CrossRef]

16. Wei, C.; Huang, J.; Mansaray, L.R.; Li, Z.; Liu, W.; Han, J. Estimation and mapping of winter oilseed rape LAI from high spatial resolution satellite data based on a hybrid method. *Remote Sens.* **2017**, *9*, 488. [CrossRef]

17. Chen, C.; McNairn, H. A neural network integrated approach for rice crop monitoring. *Int. J. Remote Sens.* **2006**, *27*, 1367–1393. [CrossRef]

18. Arvor, D.; Jonathan, M.; Meirelles, M.S.; Dubreuil, V.; Durieux, L. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *Int. J. Remote Sens.* **2011**, *32*, 7847–7871. [CrossRef]

19. Shao, Y.; Lunetta, R.S. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 78–87. [CrossRef]

20. Galford, G.L.; Mustard, J.F.; Melillo, J.; Gendrin, A.; Cerri, C.C.; Cerri, C.E. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sens.* **2008**, *112*, 576–587. [CrossRef]

21. Kastens, J.H.; Kastens, T.L.; Kastens, D.L.A.; Price, K.P.; Martinko, E.A.; Lee, R.Y. Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sens. Environ.* **2005**, *99*, 341–356. [CrossRef]

22. Thanh Noi, P.; Kappas, M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* **2018**, *18*, 18. [CrossRef]

23. Conrad, C.; Fritsch, S.; Zeidler, J.; Rücker, G.; Dech, S. Per-field irrigated crop classification in arid Central Asia using SPOT and ASTER data. *Remote Sens.* **2010**, *2*, 1035–1056. [CrossRef]

24. Zhang, X.; Liu, J.; Qin, Z.; Qin, F. Winter wheat identification by integrating spectral and temporal information derived from multi-resolution remote sensing data. *J. Integr. Agric.* **2019**, *18*, 2628–2643. [CrossRef]

25. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]

26. Johansen, K.; Phinn, S.; Taylor, M. Mapping woody vegetation clearing in Queensland, Australia from Landsat imagery using the Google Earth Engine. *Remote Sens. Appl. Soc. Environ.* **2015**, *1*, 36–49. [CrossRef]

27. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M. High-Resolution Global Maps of 21st-century Forest Cover Change. *Science* **2013**, *342*, 850–853. [CrossRef] [PubMed]

28. Ran, G.; Wei, Y.; Gordon, H.; Amit, K. Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine. *Remote Sens.* **2016**, *8*, 634. [CrossRef]

29. Li, X.; Gong, P.; Liang, L. A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data. *Remote Sens. Environ.* **2015**, *166*, 78–90. [CrossRef]

30. Busker, T.; Roo, A.d.; Gelati, E.; Schwatke, C.; Cottam, A. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrol. Earth Syst. Ences* **2019**, *23*, 669–690. [CrossRef]

31. Nazmus, S.; Iliana, M.; John, B. Leveraging the Google Earth Engine for Drought Assessment Using Global Soil Moisture Data. *Remote Sens.* **2018**, *10*, 1265. [CrossRef]

32. Naghibi, S.A.; Pourghasemi, H.R.; Dixon, B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* **2016**, *188*, 44. [CrossRef]

33. Huang, C.; Davis, L.S.; Townshend, J.R.G. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [CrossRef]

34. PAL, M.; Mather, P.M. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* **2005**, *26*, 1007–1011. [CrossRef]

35. Song, X.; Duan, Z.; Jiang, X. Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image. *Int. J. Remote Sens.* **2012**, *33*, 3301–3320. [CrossRef]

36. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130. [CrossRef]

37. De Alban, J.D.T.; Connette, G.M.; Oswald, P.; Webb, E.L. Combined Landsat and L-band SAR data improves land cover classification and change detection in dynamic tropical landscapes. *Remote Sens.* **2018**, *10*, 306. [CrossRef]

38. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]

39. Forkuor, G.; Dimobe, K.; Serme, I.; Tondoh, J.E. Landsat-8 vs. Sentinel-2: Examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GISci. Remote Sens.* **2018**, *55*, 331–354. [CrossRef]

40. Frampton, W.J.; Dash, J.; Watmough, G.; Milton, E.J. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J. Photogramm.* **2013**, *82*, 83–92. [CrossRef]

41. Du, Y.; Zhang, Y.; Ling, F.; Wang, Q.; Li, W.; Li, X. Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sens.* **2016**, *8*, 354. [CrossRef]

42. Zhang, Y.; Qin, Q.; Ren, H.; Sun, Y.; Li, M.; Zhang, T.; Ren, S. Optimal Hyperspectral Characteristics Determination for Winter Wheat Yield Prediction. *Remote Sens.* **2018**, *10*, 2015. [CrossRef]

43. Cochranwrited, W. *Sampling Techniques*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1977.

44. Zhen, Z.; Quackenbush, L.J.; Stehman, S.V.; Zhang, L. Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. *J. Remote Sens.* **2013**, *34*, 6914–6930. [CrossRef]

45. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 270–279. [CrossRef]

46. Dwivedi, R.; Rao, B. The selection of the best possible Landsat TM band combination for delineating salt-affected soils. *Int. J. Remote Sens.* **1992**, *13*, 2051–2058. [CrossRef]

47. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [CrossRef] [PubMed]

48. Ben-Hur, A.; Weston, J. A user's guide to support vector machines. In *Data Mining Techniques for the Life Sciences*; Carugo, O., Ed.; Humana Press: New York, NY, USA, 2010; pp. 223–239. ISBN 978-1-60327-241-4.

49. Zheng, B.; Myint, S.W.; Thenkabail, P.S.; Aggarwal, R.M. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 103–112. [CrossRef]

50. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [CrossRef]

51. Li, X.; Lyu, X.; Tong, Y.; Li, S.; Liu, D. An object-based river extraction method via optimized transductive support vector machine for multi-spectral remote-sensing images. *IEEE Access* **2019**, *7*, 46165–46175. [CrossRef]

52. Rußwurm, M.; Korner, M. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.

53. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; Wadsworth & Brooks/Cole Advanced Books & Software: Monterey, CA, USA, 1984; ISBN 978-0-412-04841-8.

54. Strobl, C.; Malley, J.; Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* **2009**, *14*, 323. [CrossRef]

55. Robnik-Šikonja, M. Improving Random Forests. *Lect. Notes Comput. Sci.* **2004**, *3201*, 359–370. [CrossRef]

56. Pham, B.T.; Prakash, I.; Bui, D.T. Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* **2018**, *303*, 256–270. [CrossRef]

57. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

58. Steinberg, D. CART: Classification and regression trees. In *The Top Ten Algorithms in Data Mining*; Chapman and Hall/CRC: New York, NY, USA, 2009; pp. 193–216.

59. HPBS. Statistical Bulletin on National Economic and Social Development of Henan Province in 2019. Available online: http://www.ha.stats.gov.cn/hntj/lib/tjnj/2018/indexch.htm (accessed on 14 March 2020).

60. Williamson, D.F.; Parker, R.A.; Kendrick, J.S. The box plot: A simple visual method to interpret data. *Ann. Intern. Med.* **1989**, *110*, 916–921. [CrossRef] [PubMed]

61. Loh, W.Y. Classification and regression trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 14–23, 1942–4787. [CrossRef]

62. Knorn, J.; Rabe, A.; Radeloff, V.C.; Kuemmerle, T.; Kozak, J.; Hostert, P. Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sens. Environ.* **2009**, *113*, 957–964. [CrossRef]

63. Peng, G.; Jie, W.; Le, Y.; Zhao, Y.; Zhao, Y. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654.

64. Teluguntla, P.; Thenkabail, P.S.; Oliphant, A.; Xiong, J.; Gumma, M.K.; Congalton, R.G.; Yadav, K.; Huete, A. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 325–340. [CrossRef]

65. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

66. Rodriguez, J.D.; Perez, A.; Lozano, J.A. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 569–575. [CrossRef]

67. Bengio, Y.; Grandvalet, Y. No unbiased estimator of the variance of k-fold cross-validation. *J. Mach. Learn. Res.* **2004**, *5*, 1089–1105.