

Technical Note

Machine Learning Comparison and Parameter Setting Methods for the Detection of Dump Sites for Construction and Demolition Waste Using the Google Earth Engine

Lei Zhou , Ting Luo *, Mingyi Du, Qiang Chen , Yang Liu, Yinuo Zhu, Congcong He, Siyu Wang and Kun Yang

School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China; zhoulei@bucea.edu.cn (L.Z.); dumingyi@bucea.edu.cn (M.D.); chenqiang@bucea.edu.cn (Q.C.); liuyang@bucea.edu.cn (Y.L.); 201704010327@stu.bucea.edu.cn (Y.Z.); 2108521519044@stu.bucea.edu.cn (C.H.); 2108521518028@stu.bucea.edu.cn (S.W.); 2108570020092@stu.bucea.edu.cn (K.Y.)

* Correspondence: 2108160319002@stu.bucea.edu.cn; Tel.: +86-1352-215-2233

Abstract: Machine learning has been successfully used for object recognition within images. Due to the complexity of the spectrum and texture of construction and demolition waste (C&DW), it is difficult to construct an automatic identification method for C&DW based on machine learning and remote sensing data sources. Machine learning includes many types of algorithms; however, different algorithms and parameters have different identification effects on C&DW. Exploring the optimal method for automatic remote sensing identification of C&DW is an important approach for the intelligent supervision of C&DW. This study investigates the megacity of Beijing, which is facing high risk of C&DW pollution. To improve the classification accuracy of C&DW, buildings, vegetation, water, and crops were selected as comparative training samples based on the Google Earth Engine (GEE), and Sentinel-2 was used as the data source. Three classification methods of typical machine learning algorithms (classification and regression trees (CART), random forest (RF), and support vector machine (SVM)) were selected to classify the C&DW from remote sensing images. Using empirical methods, the experimental trial method, and the grid search method, the optimal parameterization scheme of the three classification methods was studied to determine the optimal method of remote sensing identification of C&DW based on machine learning. Through accuracy evaluation and ground verification, the overall recognition accuracies of CART, RF, and SVM for C&DW were 73.12%, 98.05%, and 85.62%, respectively, under the optimal parameterization scheme determined in this study. Among these algorithms, RF was a better C&DW identification method than were CART and SVM when the number of decision trees was 50. This study explores the robust machine learning method for automatic remote sensing identification of C&DW and provides a scientific basis for intelligent supervision and resource utilization of C&DW.

Keywords: C&DW; Google Earth Engine; machine learning



Citation: Zhou, L.; Luo, T.; Du, M.; Chen, Q.; Liu, Y.; Zhu, Y.; He, C.; Wang, S.; Yang, K. Machine Learning Comparison and Parameter Setting Methods for the Detection of Dump Sites for Construction and Demolition Waste Using the Google Earth Engine. *Remote Sens.* **2021**, *13*, 787. <https://doi.org/10.3390/rs13040787>

Academic Editor: Antonello Bonfante

Received: 21 January 2021

Accepted: 18 February 2021

Published: 21 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Construction and demolition waste (C&DW) refers to all types of solid waste generated during construction, transformation, decoration, demolition, and laying of various buildings and structures and their auxiliary facilities, primarily including residue soil, waste concrete, broken bricks and tiles, waste asphalt, waste pipe materials, and waste wood [1]. China will inevitably produce more C&DW in the future due to its rapid economic development. Based on the available statistics, the output of solid domestic waste has reached 7 billion tons, C&DW contributes 30–40% of the total urban waste [2], and the newly produced C&DW will reach 300 million tons per year [3]. If C&DW is not treated and used appropriately, it will cause serious effects on society, the environment,

and resources [4,5]. However, random stacking and a large amount of C&DW lead to management difficulties. Therefore, quick and accurate identification of C&DW is a critical issue of C&DW management.

Many studies have investigated C&DW, most of which focused on the relationship between environmental quality and C&DW [6–8], the treatment methods of C&DW [9,10], and the location selection for C&DW [11,12]. A few studies have used GIS tools to predict existing C&DW areas [13]. However, studies of C&DW identification using satellite images are rare. Currently, the identification methods of C&DW primarily include artificial field investigation and remote sensing monitoring. Due to the broad geographical distribution and large number of C&DW stacking sites [14], manual field investigations must be performed, requiring large human and material resources, yielding a low work efficiency. Based on remote sensing data sources, researchers have studied the spatial distribution of garbage dumps and solid waste by visual interpretation [15,16]. Because C&DW is often mixed with domestic and kitchen waste, and its surfaces may be covered by weeds or green nets, the spectral and textural characteristics of C&DW are relatively complex [17], which increases the difficulty of visually interpreting C&DW [18]. Therefore, a classification method based on mathematical statistics and visual interpretation cannot meet the efficiency and accuracy requirements of C&DW classification in remote sensing images [19].

It is difficult to identify C&DW using remote sensing images simply by constructing a remote sensing monitoring index. Existing studies showed that RF can identify the trend feature of C&DW, while the extreme learning machine (ELM) can identify the amplitude characteristics of C&DW [20]. Related studies have demonstrated the identifiability of C&DW using machine learning methods, providing a new method for remote sensing identification of C&DW. Other studies have used spectral analysis and machine learning methods to identify C&DW [17], and others have used machine learning to identify and classify C&DW based on spatial and spectral features [21]. However, machine learning algorithms can have different recognition effects for the same class of objects [22]. Different algorithms contain multiple parameters, and the sensitivity of an algorithm to its model parameters is different [23]. A model's core parameters determine its recognition ability for classified objects. For example, the core parameter of classification and regression trees (CART) is the node, the core parameter of random forest (RF) is the number of trees, and the core parameters of support vector machine (SVM) include kernel types, gamma, and cost [24]. Simply relying on default parameters for classification will affect the classification performance of machine learning methods. Better recognition and classification results can be obtained by optimizing parameters of the machine learning algorithm [25]. Therefore, it is important to study the optimal parameterization scheme of machine learning to improve the intelligent C&DW identification of machine learning methods.

Machine learning algorithms have certain advantages in object recognition; however, they consume considerable data storage and computing resources, and their application on ordinary computers is limited. However, Google Earth Engine (GEE) can solve these problems. Studies have shown that GEE stores many remote sensing image datasets and can compute quickly [26]. GEE integrates rich machine learning algorithms [27], making it a powerful tool for analyzing land surface data [28]. There are an increasing number of studies of object recognition and classification using machine learning methods based on GEE such as identifying agricultural greenhouses [29], extracting urban boundaries [30], monitoring cultivated land and pasture ranges [31,32], and classifying land cover [33]. However, there are only a few studies on the identification of C&DW due to its complexity. Additionally, the acquisition and processing of many remote sensing images is a major difficulty in the study of C&DW classification. GEE has abundant high-quality data sources, and can obtain many C&DW samples, which provides the basis for long-term and large-scale remote sensing analysis of C&DW. Additionally, GEE supports parameter optimization for specific algorithms to determine the optimal machine learning method and parameter scheme for objects to be identified.

Therefore, considering the megacity of Beijing, this study explores the feasibility of identifying C&DW with complex features based on the GEE platform and machine learning algorithm. The specific problems to be solved in this study include: (1) determining the optimal parameter scheme of each algorithm by optimizing the core parameters (e.g., minimum leaf population, number of trees, kernel type, gamma, cost for CART, RF, and SVM); (2) clarifying the intelligent recognition ability of each algorithm after determining the optimal parameter scheme of each algorithm; and (3) selecting the most robust remote sensing identification method of C&DW by field survey and accuracy evaluation. This study can reduce the monitoring and control costs of C&DW and provide the basis for research on the spatial and temporal distribution of C&DW, resource utilization, and environmental pollution risk reduction in megacities.

2. Study Areas and Datasets

2.1. Study Area

Beijing is located in the northern part of China with a central location at $116^{\circ}20'E$ and $39^{\circ}56'N$. The city has 16 districts with a total area of $16,410.54 \text{ km}^2$ (Figure 1). Based on the available statistics, the annual output of C&DW in Beijing is approximately 92,852,400 tons, while the annual transportation volume is more than 45 million tons [34]. C&DW occupies land resources, and causes air, soil, and water pollution without timely management. Most C&DW in urban areas is piled up in waste consumption farms that have been transformed from historical pit sites and kiln sites located outside the Fifth Ring Road in Beijing [35].

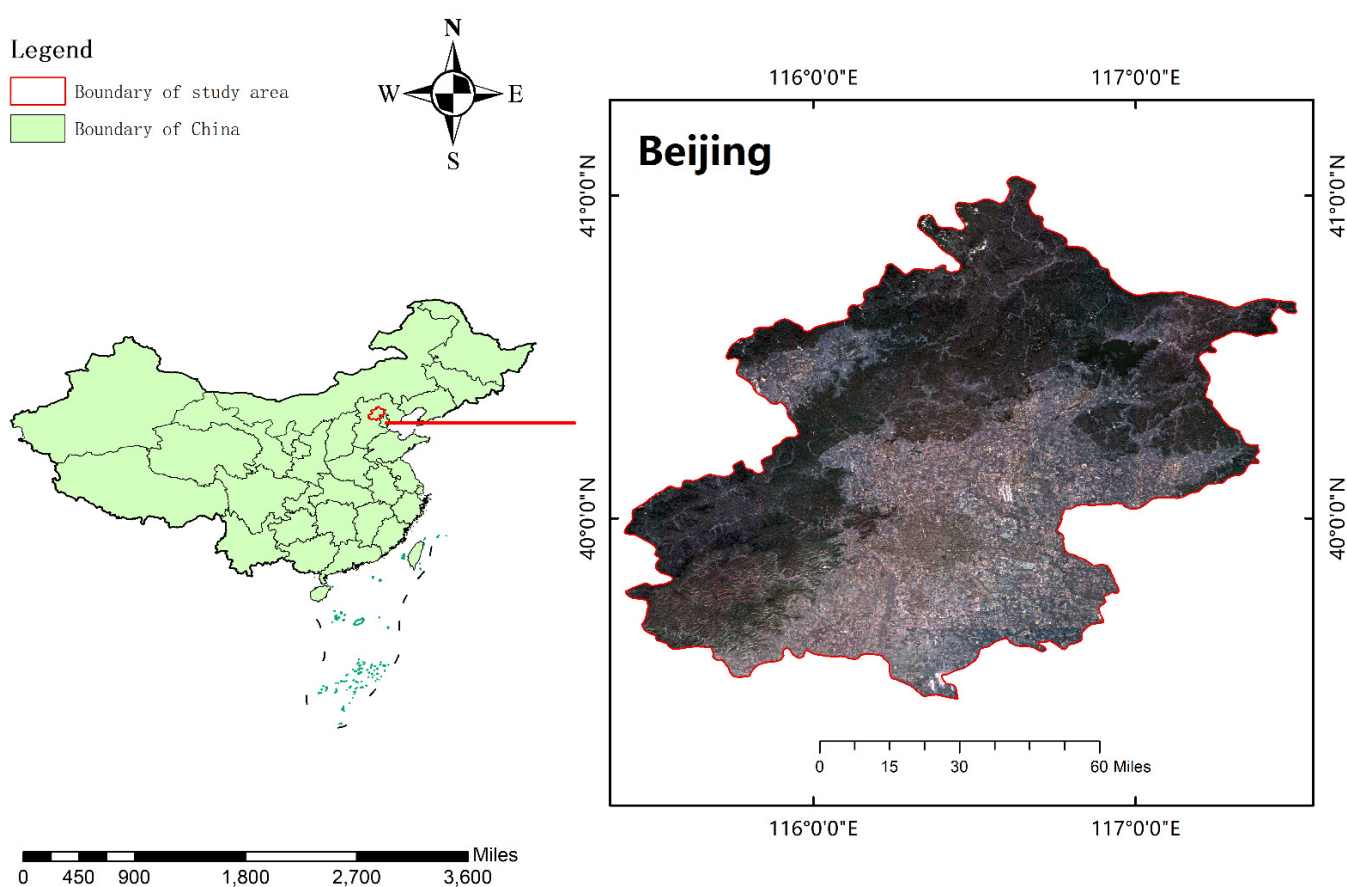


Figure 1. Location of Beijing and the Sentinel-2 remote sensing image of Beijing.

2.2. Data Source

A large amount of geospatial remote sensing data is derived from GEE. The dataset consists of remote sensing images of Earth observations, including those from Landsat,

MODIS, Sentinel-1, and Sentinel-2. The Sentinel-2 mission collects high-resolution multi-spectral imagery, which is useful for a broad range of applications, including monitoring vegetation, soil and water cover, land cover change, and humanitarian and disaster risks. According to the Earth Engine Data Catalog in GEE, this product is a Level-2A product, and the surface reflection product has been preprocessed, namely for radiation, geometry, and atmospheric correction (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR, accessed on 21 January 2021). The product details are shown in Table 1. In this study, the sentinel image data of the whole year were selected for research, and Quality Assessment 60 (QA60) wave code was used as a pre-filter to reduce turbid particles and eliminate the influence of clouds.

Table 1. The waveband parameters of Sentinel-2. The assets contain 12 UINT16 spectral bands representing surface reflectance (SR) scaled by 10,000. There are several more L2-specific bands. In addition, three QA bands are present where one (QA60) is a bitmask band with cloud mask information.

Name	Scale	Pixel Size	Description
B1	0.0001	60 m	Aerosols
B2	0.0001	10 m	Blue
B3	0.0001	10 m	Green
B4	0.0001	10 m	Red
B5	0.0001	20 m	Red Edge 1
B6	0.0001	20 m	Red Edge 2
B7	0.0001	20 m	Red Edge 3
B8	0.0001	10 m	NIR
QA10	-	10 m	Always empty
QA20	-	20 m	Always empty
QA60	-	60 m	Cloud mask

3. Methods

Remote sensing data covering Beijing are quickly acquired and processed by GEE. The C&DW sample dataset was constructed based on the spectral, textural, and topographic features of construction waste, and the CART, RF, and SVM algorithms were optimized by adjusting the parameters. The optimal identification method of construction waste was finally determined by evaluating the classification accuracy and analyzing ground verification results. A technical flow chart of the GEE process is shown in Figure 2.

3.1. Sampling

This study used Sentinel-2 images of Beijing from GEE, and all data sets were pre-processed (e.g., cloud removal). Based on the spectral and textural characteristics of the satellite data, this study focused on five categories, buildings, vegetation, water bodies, crops, and C&DW, to perform target recognition. Five types of sample points were selected in the study area based on the principle of uniform selection and prior knowledge. Visual interpretation was used to build a training dataset, which ensured the accuracy of the feature subset. In addition, different regions and shapes of C&DW samples were selected to avoid redundant and highly correlated features. The sample size of each feature type is shown in Table 2. All sample points had to be merged in the same layer to form a sample set, and the selected samples were divided into two parts (training and testing), where 70% of the samples were randomly selected for training and the other 30% of the samples were used to verify and evaluate the accuracy of the algorithm. The random column function of GEE was used to assign a random number (with the random values ranging from 0 to 1) for all sample points. Therefore, all the sample data had an extra random value. The samples with a random number ≥ 0.7 were used for validation, while the samples with a random number < 0.7 were taken as training samples.

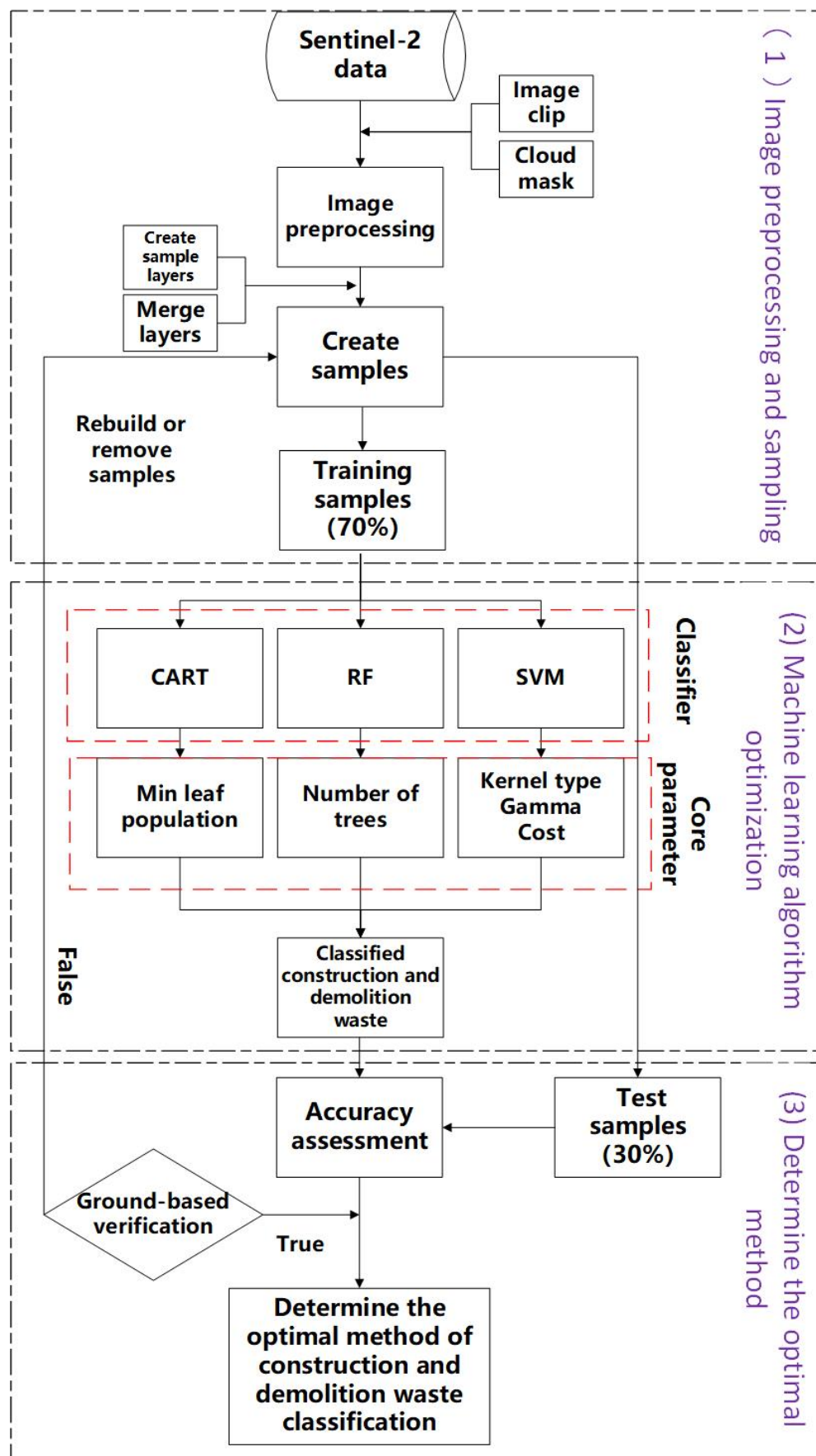


Figure 2. Methodological steps conducted for this study.

Table 2. Sample data selected in this study. A total number of 520 samples were generated as the training dataset for the machine learning methods. C&DW: construction and demolition waste.

Code	Type of Objects	Feature Extraction	Sample Size
0	Building	Land for construction, such as for houses and roads	120
1	Vegetation	Urban greening and other evergreen vegetation	100
2	Water	Water bodies such as rivers, lakes, and swamps	100
3	C&DW	Bare or strained-covered construction and demolition waste	100
4	Crops	Non-green crops	100
Total			520

3.2. Machine Learning

3.2.1. CART and Parametric Optimization Scheme

The CART algorithm is a supervised machine learning algorithm [36] that uses training samples to identify and construct trees to solve the problem of “same object with different spectrum, foreign object with same spectrum” in remote sensing recognition and classification [37]. Therefore, CART is widely used for remote sensing classification.

The CART algorithm divides n -dimensional space into non-overlapping rectangles by recursion. Firstly, an independent variable x_i is selected, and then a value u_i of x_i is selected. The n -dimensional space is divided into two parts. Certain points satisfy $x_i \leq u_i$, and the others satisfy $x_i > u_i$. For a discontinuous variable, there are only two values for the attribute value: equal or not equal. During recursive processing, these two parts based on the first step to reselect an attribute continue to partition until the entire n -dimensional space is divided. Attributes with minimum *GINI* coefficient values are used as partition indexes. For a dataset D , the *GINI* coefficient is defined as follows:

$$GINI * (D) = \sum_{i=1}^k p_i * (1 - p_i) = 1 - \sum_{i=1}^k p_i^2 \quad (1)$$

where k is the number of categories of samples and $p(i)$ represents the probability that a sample is classified into category i . The smaller the *GINI* value, the higher is the “purity” of the sample, and the better is the division effect [38].

The decision tree is composed of multilevel and multi-leaf nodes. Therefore, the decision tree can be pruned by controlling the parameters or thresholds of the new branch (Figure 3). Max nodes refers to the maximum number of leaves per tree, and min leaf population is the minimum number of nodes that are created only for the training set. To construct a suitable tree, sufficient nodes and branches must be created. The max node value is unlimited if it is not specified in GEE. The empirical method is used to select the optimal Min leaf population.

3.2.2. RF and Parametric Optimization Scheme

RF is an integrated learning algorithm that can integrate many decision trees and then form a forest (Figure 4). The algorithm combines random features or a combination of random features to generate a tree. The bagging method is used to generate training samples, and each selected feature is randomly drawn by replacing N (the size of the original training set) samples. Then, the prediction of multiple decision trees is combined, and the final prediction result is obtained by voting [39]. The final classification decision is as follows:

$$H(x) = \operatorname{argmax}_Y \sum_{i=1}^k I(h_i(x) = Y) \quad (2)$$

where $H(x)$ is the combination model, h_i is a single decision tree's classification model, Y is the output variable (or target variable), and $I(\cdot)$ is the indicator function. The formula shows that the RF uses the majority of voting decisions to determine the final classification.

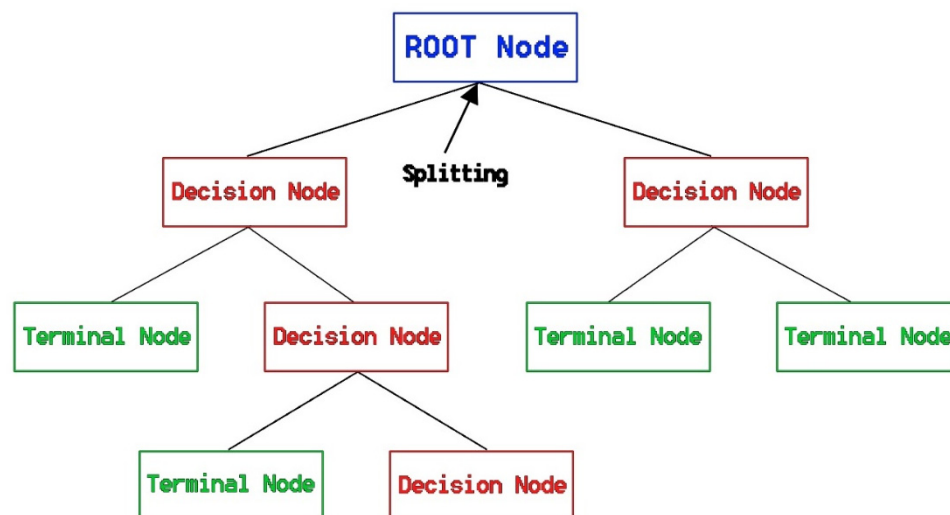


Figure 3. The decision tree can be pruned by nodes in classification and regression trees (CART).

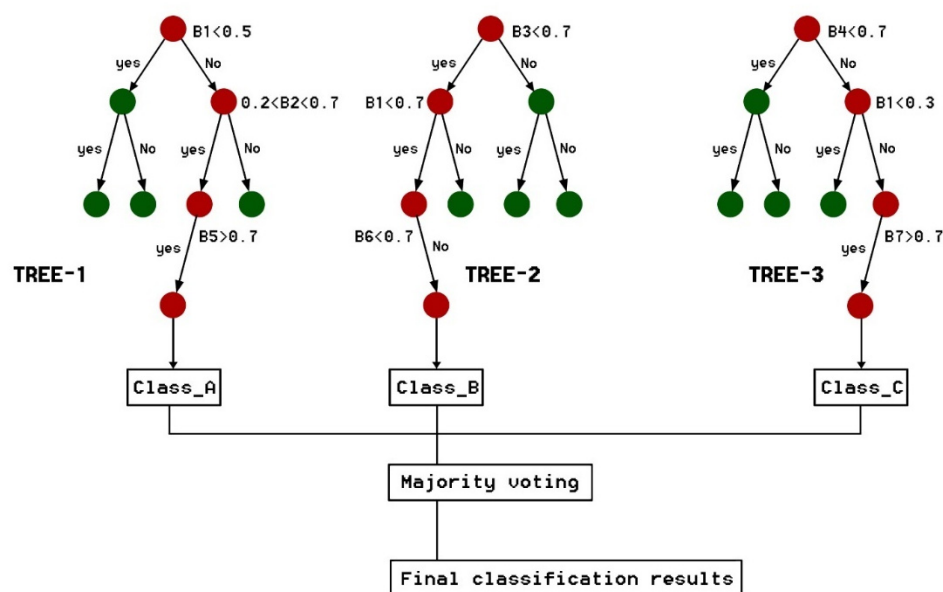


Figure 4. Example of the trees ensemble in the random forest (RF).

RF has a good tolerance for outliers and noise and is not easy to over fit (i.e., is stable) [40]. The adjustable parameter of the RF algorithm is the number of trees, and the number of trees is selected empirically.

3.2.3. SVM and Parametric Optimization Scheme

SVM is a supervised machine learning algorithm that can manage sample scarcity, is robust, and typically yields good results during classification and regression. SVM uses hyperplanes to divide support vectors to classify log data points clearly [41] with the goal of finding the two types of independent support vectors with the largest margin (maximum distance) (Figure 5).

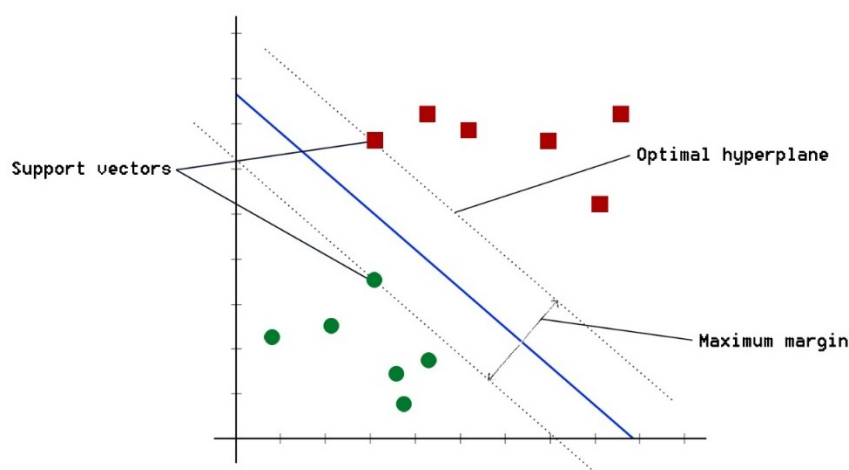


Figure 5. Optimal hyperplane identification in support vector machine (SVM).

The SVM algorithm is based on the kernel method, and the selection of the kernel type has a strong effect on the classification results [42]. Currently, there are three types of kernels commonly used:

- Polynomial kernel:

$$K(x, x') = (\gamma(x \cdot x') + c)^q \quad (3)$$

where q is the polynomial order and c and q are artificially defined parameters.

- Radial basis function (RBF) kernel:

$$K(x, x') = \exp\left(-\gamma \frac{\|x - x'\|^2}{\sigma^2}\right) \quad (4)$$

where γ is greater than 0 and defined manually.

- SIGmoID kernel.

$$K(x, x') = \tanh(v(x \cdot x') + c) \quad (5)$$

where c is the epsilon, a manually defined parameter.

In GEE, the SVM algorithm is described by the kernel type, gamma, and cost parameters [24]. The kernel type can be polynomial, SIGmoID, or RBF. Gamma represents a parameter of the function after selecting the RBF function and implicitly determines the distribution of the data mapped to the new eigenspace. The larger the gamma is, the smaller is the support vector, and vice versa. The number of support vectors affects the speed of training and prediction. Cost represents the regularization parameter C of the error term.

Currently, SVM parameter selection methods primarily include empirical method, experimental trial method, gradient descent method, cross validation method, Bayesian method, etc. [43,44]. Among the three kernel functions, the RBF kernel function is relatively stable, while the polynomial kernel function and sigmoid kernel function have relatively low stability [43]. Therefore, the RBF kernel type was used in this study. The values of gamma and cost were selected by a combination of the experimental trial method and the grid search method.

3.3. Verification Methods

The confusion matrix is the core method of accuracy evaluation, which can describe the classification accuracy and show the confusion between categories. The basic statistics for the confusion matrix include the overall accuracy (OA), consumer accuracy (CA), producer accuracy (PA), and Kappa coefficient. In this study, 156 sample points were taken to make the confusion matrix.

To verify the classification results, the ground verification method was used to visually interpret the optimized classification results of each algorithm. Different classification algorithms for C&DW based on satellite images were compared and analyzed. Based on accuracy assessment and ground verification, the optimal scheme of CART, RF, and SVM for C&DW identification was determined.

4. Results and Analysis

4.1. Accuracy Assessment

The detail codes are accessible from Code links (see Supplementary Materials). From the algorithm optimization study, the statistics of each algorithm's precision index are as follows:

For optimizing the CART algorithm, the min leaf population was between 0 and 10. As shown in Figure 6a, the OA, CA, PA and kappa coefficient remained unchanged as the number of nodes increased, with values of 73.13%, 65.22%, 50%, and 65.22%, respectively. The CART algorithm yielded relatively low classification accuracy for C&DW, and the change in parameters had little influence on the classification. When optimizing the RF algorithm, as shown in Figure 6b, the OA, CA, PA and kappa coefficient all increased rapidly as the number of trees increased from 1 to 20. The OA and PA tended to decrease and then increase slowly when the number of trees was approximately 40. Classification precisions reached their highest values and remained stable when the number of trees was 50. The highest values of OA, CA, PA, and the kappa coefficients for RF were 98.05%, 100%, 96.67%, and 98.38%, respectively.

For SVM optimization, the optimal gamma parameter was determined first, followed by the optimal cost parameter. The empirical method was used to set the cost value to 10. Then, the experimental trial method and grid search method were used to obtain and compare the statistical values of each precision index in the cases of the gamma and cost parameters. As shown in Figure 6c, the OA, CA, PA and kappa coefficient rose markedly and then increased slowly when gamma changed between 0.2 and 0.6. The OA, CA, PA, and kappa coefficient reached a small wave peak when gamma was 5. Additionally, the classification accuracy was good. As shown in Figure 6d, the classification precision tended to be stable when $\gamma \geq 10$. The highest levels of the OA, CA, PA, and the kappa coefficients were 84.68%, 81.82%, 60%, and 75.6%, respectively, when gamma was 16. The cost parameter was studied and compared when gamma was 5 and 16 for experimental rigor. As shown in Figure 6e, OA and CA increased with increasing cost when gamma was 5, while PA and the kappa coefficient fluctuated marginally from 1 to 20. The classification precision reached its highest value and remained stable when the cost was 40. The highest values of OA, CA, PA, and the kappa coefficients were 84.38%, 85.71%, 60%, and 76.37%, respectively. As shown in Figure 6f, the OA, CA, and kappa coefficient increased with increasing cost when gamma was 5, except for cost near 14. The PA increased first and then decreased, and finally tended to be stable with increasing cost. The highest values of OA, CA, PA, and the kappa coefficient were 85.62%, 85.71%, 60%, and 78.78%, respectively, when cost was 34. For the SVM algorithm, the combination of gamma = 16 and cost = 34 yielded better accuracy than when gamma = 5 and cost = 40.

This study selected certain experimental results of C&DW identification, and drew a distribution comparison map of C&DW in Beijing, as shown in Figure 7. The detected C&DW distribution results were less affected by CART parameters, followed by RF parameters, and then the SVM parameters.

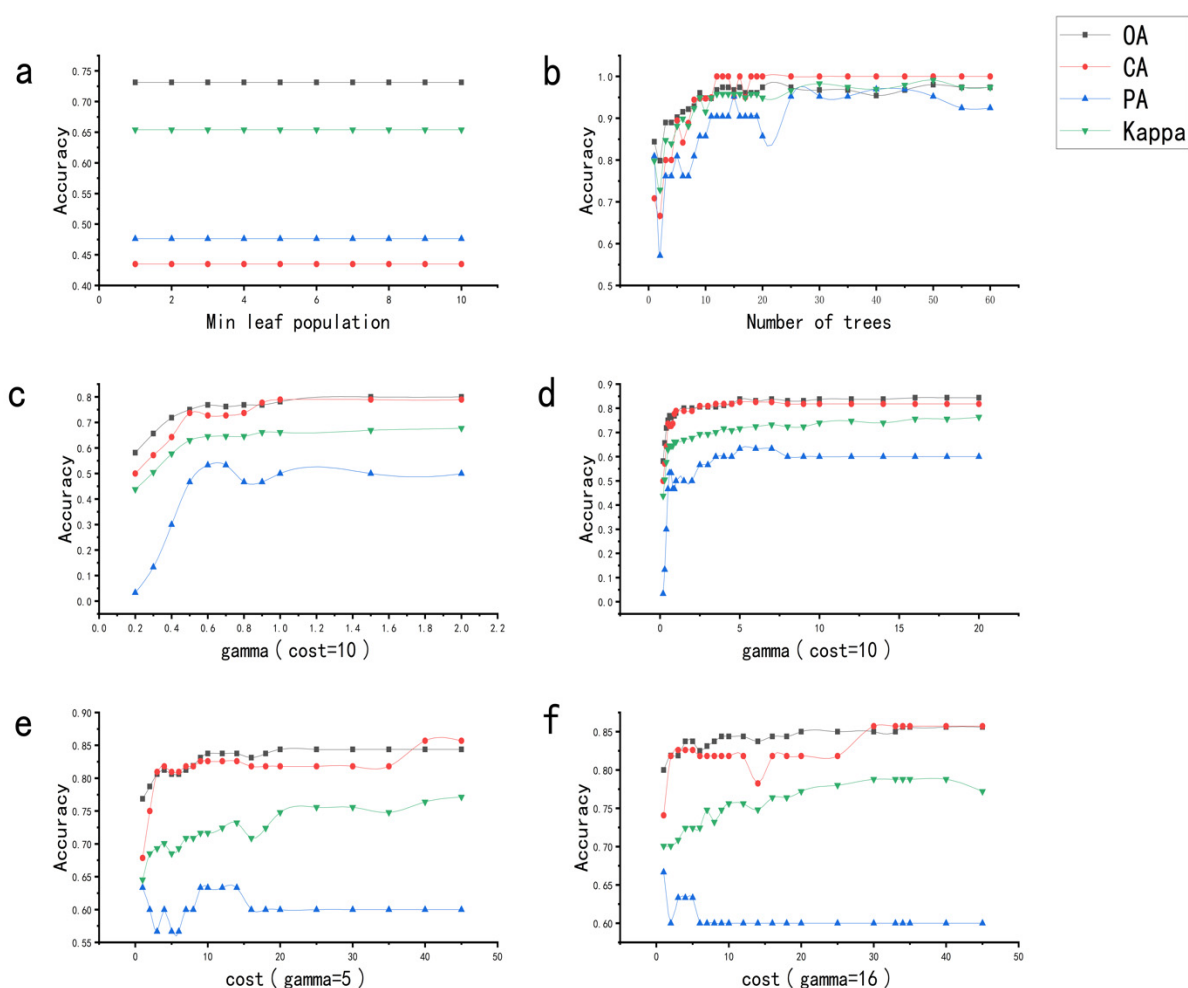


Figure 6. Accuracy index statistics of each algorithm. (a) The result of CART; (b) the result of RF; (c) the result of SVM when cost is 10 and gamma changes between 0.2 and 0.6; (d) the result of SVM when cost is 10 and gamma changes between 0.2 and 20; (e) the result of SVM when gamma is 5; (f) the result of SVM when gamma is 16.

Statistical analyses of each accuracy index described an optimal parameterization scheme of CART, RF, SVM and the corresponding precision. The results of this process are shown in Table 3.

4.2. Ground Verification

This study verified the classification results using ground truth data to compare the classification effect and the reliability of various algorithms for C&DW detection. In certain areas of Beijing, the real state and primary characteristics in C&DW were examined using remote sensing images at construction waste dumps on 15 December 2020, including formal consumption sites and informal construction waste dump areas. As shown in Figure 8, C&DW was classified and absorbed into a regular consumption area. The C&DW was piled in an orderly manner based on brick size and covered with green grids to prevent air pollution. However, certain informal C&DW dumps have existed for a long time. After many years of weathering, weeds grow on the surface of the C&DW accumulation, making it difficult to identify that area as a C&DW dump.

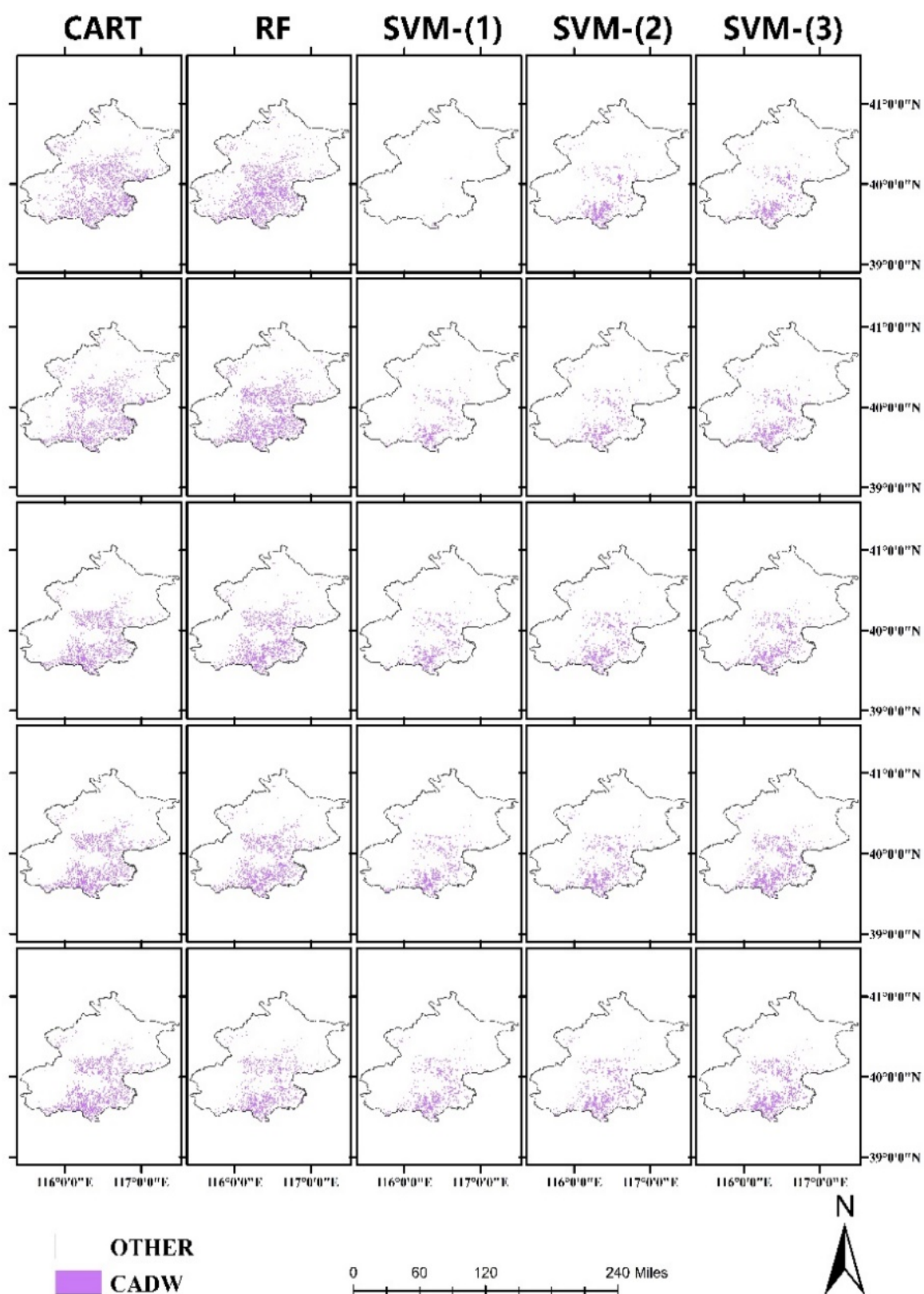


Figure 7. Distribution of C&DW by CART, RF, and SVM algorithms in Beijing. SVM-(1) cost = 10; SVM-(2) gamma = 5; SVM-(3) gamma = 16.

Table 3. Statistics of optimal parameters and accuracy of various algorithms. OA: overall accuracy; CA: consumer accuracy; PA: producer accuracy; RBF: radial basis function.

	Parameters	OA	CA	PA	Kappa
CART	Min leaf population: 6	73.12%	65.22%	50%	65.22%
RF	Number of trees: 50	98.05%	100%	96.67%	98.38%
	Kernel type: RBF				
SVM	Gamma: 16	85.62%	85.71%	60%	78.78%
	Cost: 34				



Figure 8. Field investigation of C&DW in Beijing on 15 December 2020.

C&DW is scattered and irregular in shape. To analyze the prediction results of each algorithm in more detail, certain C&DW areas with concentrated accumulation and relatively regular shapes were selected for visual interpretation and comparison. Based on the distribution map of C&DW under the optimal parameterization scheme, the details of C&DW identification in different areas were analyzed, as shown from Figure 9. Typically, each algorithm can identify C&DW; however, the recognition results were marginally different. In regions A and D, compared to SVM and RF, the classification ability of CART was inferior. In regions B and C, CART could not completely identify C&DW, and SVM incorrectly classified ground objects that were not C&DW as C&DW. Additionally, RF was shown to be more accurate in classifying C&DW. The analysis of four different regions shows that RF yielded the best recognition results compared to that of CART and SVM. However, SVM had good recognition results along edges of C&DW, such as in regions B and C.

The numbers of pixels identified as C&DW in the four areas are shown in Table 4. The number of C&DW areas identified by SVM was typically higher than those identified by CART and RF. Combined with the visual results shown in Figure 9, the SVM algorithm may have experienced overfitting.

Based on the accuracy assessment indices and ground verification of C&DW, the RF algorithm was shown to yield better recognition results for C&DW detection among the tested algorithms. Identification was best when RF had 50 decision trees.

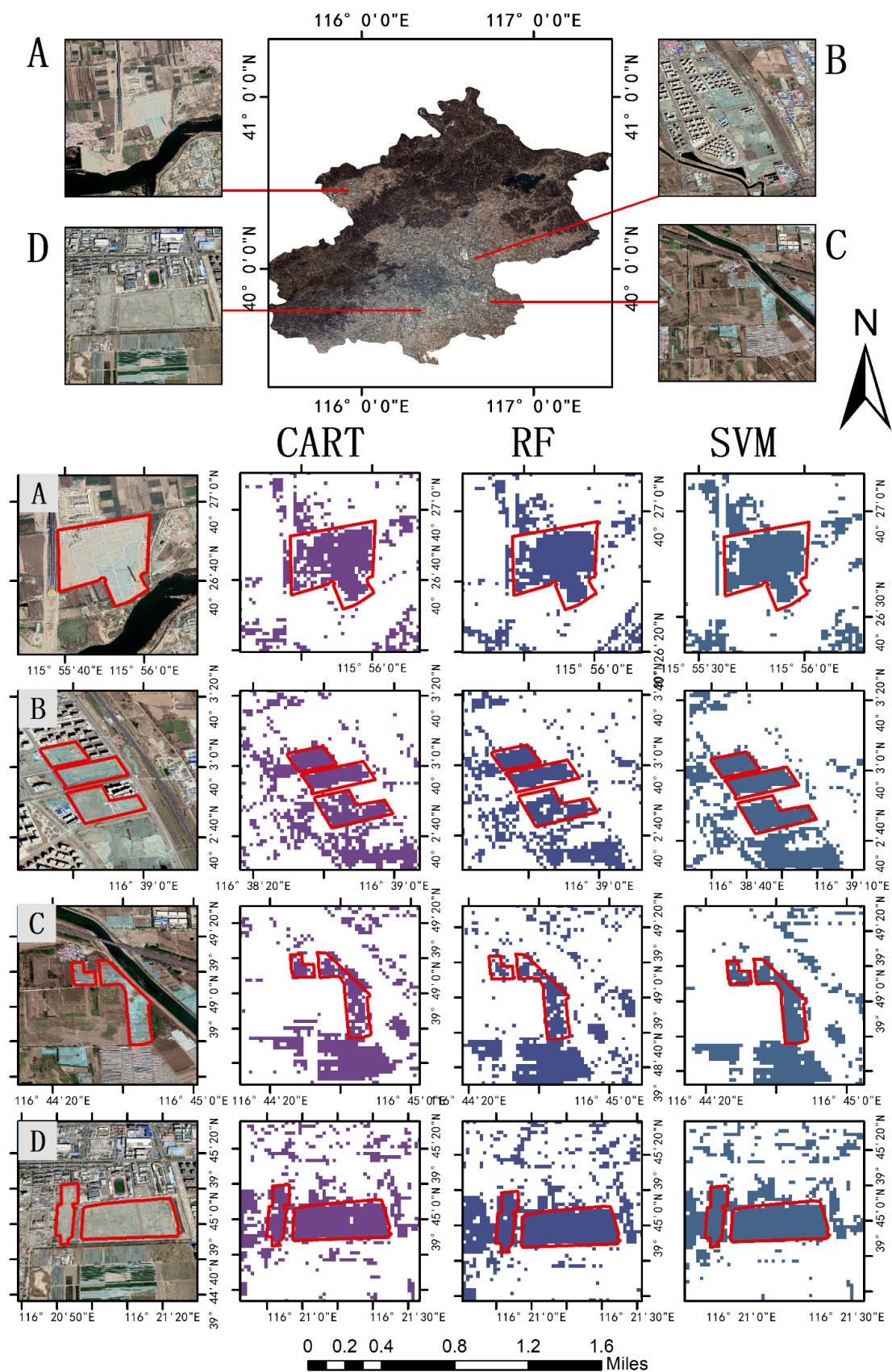


Figure 9. Algorithm to identify the details of construction waste. (A–D) Different regions of this study area. The areas surrounded by red are C&DW areas with concentrated accumulation and relatively regular shapes.

Table 4. Number of pixels identified as C&DW by each algorithm in the four study areas.

	CART	RF	SVM
A	326	336	339
B	256	247	283
C	118	114	145
D	316	331	332

5. Discussion and Conclusions

In recent years, due to the serious threat of construction and demolition waste (C&DW) to society, the economy, and the environment, C&DW management has received increasing attention. Intelligent identification of C&DW is an important method of waste supervision and resource utilization. This study attempted to determine a method suitable for C&DW identification and classification by optimizing parameters based on Google Earth Engine (GEE) and machine learning algorithms. The results of this study provide a method for C&DW remote sensing identification. The types of ground objects were divided into buildings, vegetation, water, C&DW, and crops. The parameter optimization studies of machine learning found that the overall classification accuracy of each algorithm was optimal when 6 nodes were used with the classification and regression trees (CART) algorithm, 50 trees are used with the random forest (RF) algorithm, and $\gamma = 16$ and $\text{cost} = 34$ with the support vector machine (SVM) algorithm. Ground verification results of C&DW distribution points in Beijing show that the results of this study are reliable. The results showed that CART, RF, and SVM have different recognition abilities for C&DW in Beijing. Compared to CART and SVM, RF performed better in terms of the overall accuracy (OA) and identification ability of C&DW. In some other studies, four methods (characteristic reflectivity and extreme learning machine; first-order derivative of characteristic reflectivity and extreme learning machine; grey level co-occurrence matrix and extreme learning machine; and convolutional neural network) were proposed for the automatic identification of C&DW. The given correct rate was around 80% in these studies [20]. Our results were consistent with previous studies. They further prove the effectiveness of the machine learning method in the intelligent identification of C&DW based on remote sensing images.

This study confirmed the feasibility of intelligent identification of C&DW based on GEE and machine learning algorithms. A parameter optimization scheme of the machine learning algorithm was proposed to improve the remote sensing recognition ability for C&DW. Based on a field-based accuracy assessment, the best machine learning method and its parameterization scheme for remote sensing identification of C&DW were determined. This study could provide the basis for research on the spatial and temporal distribution of C&DW, resource utilization, and environmental pollution risk reduction in megacities. The results of this study are helpful for the treatment and management of C&DW and associated cost reductions, which are beneficial for saving land resources and promoting energy conservation and emission reduction [45]. Additionally, the distribution map of C&DW in Beijing provides a scientific basis for intelligent supervision and resource utilization of C&DW.

However, the proposed method still has certain limitations, which should be further explored in future research. For example, during identification and classification, failure to consider the elevation of C&DW may lead to the classification of construction waste as buildings, and C&DW covered with green grids may be misclassified as vegetation or crops. In addition to the influence of algorithm parameters, the source of the sample data, the number of samples, and the partition ratio of the training/validation set will affect the final classification performance [46]. What is more, deep learning methods also have great potential in the intelligent identification of C&DW, and we will undertake more studies on deep learning identification methods of C&DW identification based on TensorFlow. In addition, we will make further in-depth comparisons with other methods to find the optimal method of intelligent identification of C&DW. The accuracy validation

of C&DW remote sensing identification results is also an important aspect of our future research. Based on the existing data, this study discussed the influence of algorithm parameter combinations on the performance of C&DW identification and classification, and preliminarily determined an optimal scheme. This paper focused on the new method of intelligent identification of C&DW based on machine learning and remote sensing data. The uncertain factors should be investigated in future research.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2072-4292/13/4/787/s1>, Code links of the detection of dump sites for construction and demolition waste are uploaded with supplementary file.

Author Contributions: Conceptualization, L.Z. and T.L.; Data curation, L.Z. and T.L.; Formal analysis, L.Z. and T.L.; Funding acquisition, L.Z., M.D., Q.C., and Y.L.; Investigation, L.Z. and T.L.; Methodology, T.L.; Project administration, L.Z. and M.D.; Resources, T.L. and L.Z.; Software, T.L.; Supervision, L.Z., Q.C., Y.L., and M.D.; Validation, T.L.; Visualization, T.L.; Writing—original draft, T.L. and L.Z.; Writing—review and editing, L.Z., T.L., M.D., Q.C., Y.L., Y.Z., C.H., S.W., and K.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Key Project # 2018YFC0706003), the National Natural Science Foundation (NSFC) (Key Project # 42077439, 41930650), and the Beijing Advanced Innovation Center for Future Urban Design (Key Project # UDC2018030611).

Acknowledgments: The authors would like to thank the reviewers of the manuscript for their helpful comments.

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

References

1. Zhou, W.; Chen, J.; Lu, H. Status And Countermeasures Of Domestic Construction Waste Resources. *Archit. Technol.* **2009**, *40*, 741–744.
2. Huang, B.; Wang, X.; Kua, H.; Geng, Y.; Bleischwitz, R.; Ren, J. Construction and demolition waste management in China through the 3R principle. *Resour. Conserv. Recycl.* **2018**, *129*, 36–44. [\[CrossRef\]](#)
3. Zhou, Z.H.; Su, Z. Current Situation and Measures of Construction Waste Treatment. *Sci. Technol. Innov.* **2019**, *07*, 114–115.
4. Yazdani, M.; Kabirifar, K.; Frimpong, B.E.; Shariati, M.; Mirmozaffari, M.; Boskabadi, A. Improving construction and demolition waste collection service in an urban area using a simheuristic approach: A case study in Sydney, Australia. *J. Clean. Prod.* **2021**, *280*, 124138. [\[CrossRef\]](#)
5. Ma, M.; Tamm, V.W.Y.; Le, K.N. Challenges in current construction and demolition waste recycling: A China study. *Waste Manag.* **2020**, *118*, 610–625. [\[CrossRef\]](#)
6. Oh, D.; Noguchi, T.; Kitagaki, R.; Choi, H. Proposal of demolished concrete recycling system based on performance evaluation of inorganic building materials manufactured from waste concrete powder. *J. Renew. Sustain. Energy Rev.* **2021**, *135*, 110147. [\[CrossRef\]](#)
7. Cristelo, N.; Vieira, C.S.; Lopes, M.L. Geotechnical and Geoenvironmental Assessment of Recycled Construction and Demolition Waste for Road Embankments. *Procedia Eng.* **2016**, *143*, 51–58. [\[CrossRef\]](#)
8. Yu, D.; Duan, H.; Song, Q.; Li, X.; Zhang, H.; Zhang, H.; Liu, Y.; Shen, W.; Wang, J. Characterizing the environmental impact of metals in construction and demolition waste. *Environ. Sci. Pollut. Res. Int.* **2018**, *25*, 13823–13832.
9. Khajuria, A.; Matsui, T.; Machimura, T.; Morioka, T. Decoupling and Environmental Kuznets Curve for municipal solid waste generation: Evidence from India. *J. Int. J. Environ. Sci.* **2012**, *2*, 1670–1674.
10. Wang, H. Technology and Demonstration Project of Construction Waste in-situ Treatment in the Northern Part of Haidian District, Beijing. *Build. Energy Effic.* **2016**, *44*, 84–88.
11. Biluca, J.; Aguiar, C.R.; Trojan, F. Sorting of suitable areas for disposal of construction and demolition waste using GIS and ELECTRE TRI. *Waste Manag.* **2020**, *114*, 307–320. [\[CrossRef\]](#)
12. Lin, Z.; Xie, Q.; Feng, Y.; Zhang, P.; Yao, P. Towards a robust facility location model for construction and demolition waste transfer stations under uncertain environment: The case of Chongqing. *Waste Manag.* **2020**, *105*, 73–83. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Nissim, S.; Portnov, A.B. Identifying areas under potential risk of illegal construction and demolition waste dumping using GIS tools. *Waste Manag.* **2018**, *75*, 22–29.
14. Zhang, F.; Ju, Y.; Dong, P.; Gonzalez, E.D.S. A fuzzy evaluation and selection of construction and demolition waste utilization modes in Xi'an, China. *Waste Manag. Res.* **2020**, *38*, 792–801. [\[CrossRef\]](#)
15. Yalan, L.; Yuhuan, R.; Chengjie, W.; Gaihua, W.; Huizhen, Z.; Yaobin, C. Study on monitoring of informal open-air solid waste dumps based on Beijing-1 images. *J. Remote Sens.* **2009**, *13*, 320–326.

16. Wu, W.W.; Liu, J. The Application of Remote Sensing Technology on the Distribution Investigation of the Solid Waste in Beijing. *Environ. Sanit. Eng.* **2000**, *8*, 76–78.
17. Kuritcyn, P.; Anding, K.; Linß, E.; Latyev, S.M. Increasing the Safety in Recycling of Construction and Demolition Waste by Using Supervised Machine Learning. *J. Phys. Conf. Ser.* **2015**, *588*, 012035. [[CrossRef](#)]
18. Ku, Y.; Yang, J.; Fang, H. Researchers at Huaqiao University Release New Data on Robotics (Deep Learning of Grasping Detection for a Robot Used In Sorting Construction and Demolition Waste). *J. Robot. Mach. Learn.* **2020**, *23*, 84–95.
19. Gu, X.; Gao, X.; Ma, H.; Shi, F.; Liu, X.; Cao, X. Comparison of Machine Learning Methods for Land Use/Land Cover Classification in the Complicated Terrain Regions. *Remote Sens. Technol. Appl.* **2019**, *34*, 57–67.
20. Xiao, W.; Yang, J.; Fang, H.; Zhuang, J.; Ku, Y. A robust classification algorithm for separation of construction waste using NIR hyperspectral system. *Waste Manag.* **2019**, *90*, 1–9. [[CrossRef](#)]
21. Xiao, W.; Yang, J.; Fang, H.; Zhuang, J.; Ku, Y. Classifying construction and demolition waste by combining spatial and spectral features. *Proc. Inst. Civil Eng.—Waste Resour. Manag.* **2020**, *173*, 79–90. [[CrossRef](#)]
22. Ge, G.; Shi, Z.; Zhu, Y.; Yang, X.; Hao, Y. Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms. *Glob. Ecol. Conserv.* **2020**, *22*, e00971. [[CrossRef](#)]
23. Fang, P.; Zhang, X.; Wei, P.; Wang, Y.; Zhang, H.; Liu, F.; Zhao, J. The Classification Performance and Mechanism of Machine Learning Algorithms in Winter Wheat Mapping Using Sentinel-2 10 m Resolution Imagery. *Appl. Sci.* **2020**, *10*, 5075. [[CrossRef](#)]
24. Tharwat, A. Parameter investigation of support vector machine classifier with kernel functions. *Knowl. Inform. Syst.* **2019**, *61*, 1269–1302. [[CrossRef](#)]
25. Shaharum, N.S.N.; Shafri, H.Z.M.; Ghani, W.A.W.A.K.; Samsatli, S.; Al-Habshi, M.M.A.; Yusuf, B. Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms. *Remote Sens. Appl. Soc. Environ.* **2020**, *17*, 100287. [[CrossRef](#)]
26. 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](#)]
27. Sidhu, N.; Pebesma, E.; Câmara, G. Using Google Earth Engine to detect land cover change: Singapore as a use case. *Eur. J. Remote Sens.* **2018**, *51*, 486–500. [[CrossRef](#)]
28. Thieme, A.; Yadav, S.; Oddo, P.C.; Fitz, J.M.; McCartney, S.; King, L.; Keppler, J.; McCarty, G.W.; Hively, W.D. Using NASA Earth observations and Google Earth Engine to map winter cover crop conservation performance in the Chesapeake Bay watershed. *Remote Sens. Environ.* **2020**, *248*, 111943. [[CrossRef](#)]
29. Dehai, Z.; Yiming, L.; Quanlong, F.; Cong, O.; Hao, G.; Jiantao, L. Spatial-temporal Dynamic Changes of Agricultural Greenhouses in Shandong Province in Recent 30 Years Based on Google Earth Engine. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 168–175.
30. Goldblatt, R.; You, W.; Hanson, G.; Khandelwal, A. 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](#)]
31. Gumma, M.K.; Thenkabail, P.S.; Teluguntla, P.G.; Oliphant, A.; Xiong, J.; Giri, C.; Pyla, V. Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud. *Giscience Remote Sens.* **2020**, *57*, 302–322. [[CrossRef](#)]
32. Zhou, B.; Okin, G.S.; Zhang, J. Leveraging Google Earth Engine (GEE) and machine learning algorithms to incorporate in situ measurement from different times for rangelands monitoring. *Remote Sens. Environ.* **2020**, *236*, 111521. [[CrossRef](#)]
33. Hongwei, Z.; Bingfang, W.; Shuai, W.; Walter, M.; Fuyou, T.; Eric, M.Z.; Nitesh, P.; Mavengahama, S. A Synthesizing Land-cover Classification Method Based on Google Earth Engine: A Case Study in Nzhelele and Levhuvu Catchments, South Africa. *Chin. Geogr. Sci.* **2020**, *30*, 397–409.
34. Pang, S.; Bi, J.; Luo, Z. *Annual Report on Public Service of Beijing (2017–2018)*; Social Sciences Academic Press(CHINA): Beijing, China, 2018.
35. Xu, J.; Miao, W.; Mao, J.; Lu, J.; Xia, L. Research and Practice on Resource Utilization Technology of Construction and Demolition Debris in Beijing under Venous Industry Model. *Environ. Sanit. Eng.* **2020**, *28*, 22–25.
36. Han, J.; Mao, K.; Xu, T.; Guo, J.; Zuo, Z.; Gao, C. A soil moisture estimation framework based on the CART algorithm and its application in China. *J. Hydrol.* **2018**, *563*, 65–75. [[CrossRef](#)]
37. Qi, L.; Yue, C. Remote Sensing Image Classification Based on CART Decision Tree Method. *For. Inventory Plan.* **2011**, *36*, 62–66.
38. El-Alem, A.; Chokmani, K.; Laurion, I.; El-Adlouni, S. An Adaptive Model to Monitor Chlorophyll-a in Inland Waters in Southern Quebec Using Downscaled MODIS Imagery. *Remote Sens.* **2014**, *6*, 6446–6471. [[CrossRef](#)]
39. Du, L.; Feng, Y.; Lu, W.; Kong, L.; Yang, Z. Evolutionary game analysis of stakeholders' decision-making behaviours in construction and demolition waste management. *Environ. Impact Assess. Rev.* **2020**, *84*, 106408. [[CrossRef](#)]
40. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *J. Photogramm. Remote Sens.* **2011**, *67*, 93–104. [[CrossRef](#)]
41. Lu, L.; Tao, Y.; Di, L. Object-Based Plastic-Mulched Landcover Extraction Using Integrated Sentinel-1 and Sentinel-2 Data. *Remote Sens.* **2018**, *10*, 1820. [[CrossRef](#)]
42. Wang, K.; Cheng, L.; Yong, B. Spectral-Similarity-Based Kernel of SVM for Hyperspectral Image Classification. *Remote Sens.* **2020**, *12*, 2154. [[CrossRef](#)]

-
43. Guohe, F. Parameter optimizing for Support Vector Machines classification. *Computer Engineering and Applications. Comput. Eng. Appl.* **2011**, *47*, 123–124+128.
 44. Chen, W.; Li, X.; Wang, L. Fine Land Cover Classification in an Open Pit Mining Area Using Optimized Support Vector Machine and WorldView-3 Imagery. *Remote Sens.* **2019**, *12*, 82. [[CrossRef](#)]
 45. Zoghi, M.; Kim, S. Dynamic Modeling for Life Cycle Cost Analysis of BIM-Based Construction Waste Management. *Sustainability* **2020**, *12*, 2483. [[CrossRef](#)]
 46. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [[CrossRef](#)]