




## Article

# Sentinel-2 Satellite Imagery for Urban Land Cover Classification by Optimized Random Forest Classifier

Tianxiang Zhang <sup>1,2</sup> , Jinya Su <sup>3</sup> , Zhiyong Xu <sup>1,2</sup> , Yulin Luo <sup>4</sup> and Jianguyun Li <sup>1,2,\*</sup>

<sup>1</sup> School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China; tianxiang1992@me.com (T.Z.); g20198748@xs.ustb.edu.cn (Z.X.)

<sup>2</sup> Shunde Graduate School of University of Science and Technology Beijing, Foshan 528000, China

<sup>3</sup> School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK; j.su@essex.ac.uk

<sup>4</sup> State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China; luoyl\_2020@tsinghua.org.cn

\* Correspondence: leeji@ustb.edu.cn; Tel.: +86-186-1001-8619

**Abstract:** Land cover classification is able to reflect the potential natural and social process in urban development, providing vital information to stakeholders. Recent solutions on land cover classification are generally addressed by remotely sensed imagery and supervised classification methods. However, a high-performance classifier is desirable but challenging due to the existence of model hyperparameters. Conventional approaches generally rely on manual tuning, which is time-consuming and far from satisfying. Therefore, this work aims to propose a systematic method to automatically tune the hyperparameters by Bayesian parameter optimization for the random forest classifier. The recently launched Sentinel-2A/B satellites are drawn to provide the remote sensing imageries for land cover classification case study in Beijing, China, which have the best spectral/spatial resolutions among the freely available satellites. The improved random forest with Bayesian parameter optimization is compared against the support vector machine (SVM) and random forest (RF) with default hyperparameters by discriminating five land cover classes including building, tree, road, water, and crop field. Comparative experimental results show that the optimized RF classifier outperforms the conventional SVM and the RF with default hyperparameters in terms of accuracy, precision, and recall. The effects of band/feature number and the band usefulness are also assessed. It is envisaged that the improved classifier for Sentinel-2 satellite image processing can find a wide range of applications where high-resolution satellite imagery classification is applicable.

**Keywords:** Sentinel-2 satellite; random forest; bayesian optimization; hyperparameter tuning; urban management; land cover classification



**Citation:** Zhang, T.; Su, J.; Xu, Z.; Luo, Y.; Li, J. Sentinel-2 Satellite Imagery for Urban Land Cover Classification by Optimized Random Forest Classifier. *Appl. Sci.* **2021**, *11*, 543. <https://doi.org/10.3390/app11020543>

Received: 10 December 2020

Accepted: 5 January 2021

Published: 8 January 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

Land cover classification (LCC) is able to reflect the potential natural and social process in urban development so that the vital information can be extracted to key stakeholders [1,2]. Earth observation satellite, one of the most significant platforms, is widely applied for LCC due to their customized sensors which are able to provide extensive geographical coverage while with an affordable cost for spatial and temporal land use/cover mapping [3]. In particular, LCC using remote sensing images of high spatial/spectral resolutions is playing a paramount role in urban planning, land resource management, green infrastructure monitoring, disaster management, and agricultural applications [4–7]. In China, the largest developing country, rapid urbanization has been changing its geographic characteristics, particularly for urban areas where the balance of environment and urban infrastructures is gradually being impaired. Therefore, land cover classification for urban areas is of great importance to assess its changes for its sustainable development [6].

With the advent of various earth observation satellites, image quality in terms of spatial, spectral, and temporal resolutions is constantly improving and so the LCC performance can be guaranteed. Among the freely accessible satellites, the newly launched Sentinel-2 series satellite composed of Sentinel-2A and Sentinel-2B possesses the best spatial, spectral, and temporal resolutions [8,9], which are a key part of the Global Monitoring for Environment and Security program supported by the European Space Agency (ESA). Its Multi-Spectral Instrument (MSI) features 13 bands from visible bands to short waved infrared (SWIR) bands. In addition, three different spatial resolutions (e.g., 10 m, 20 m, and 60 m) are provided for various tailored tasks [10]. A number of qualitative and quantitative studies have been done for Sentinel-2A satellite on land management, urban planning, ecosystem monitoring, and smart farming [1,3,5,7,11,12]. Sentinel-2B was launched on 7 March 2017 to complement Sentinel-2A for a better temporal resolution. Therefore, in this study, Sentinel-2A/B satellites are selected to provide the high-resolution remote sensing images for the purpose of urban land cover classification.

On the other hand, it is widely acknowledged that the selection of classification method can significantly affect the land use/cover mapping performance [3]. The ever-increasing computation power and the advanced classification algorithms are making the land use/cover classification more accurate than ever before, where the commonly used algorithms may include the Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Decision Tree (DT), Artificial Neural Networks (ANN), and Random forest (RF) [5,13–15]. Machine learning based classifiers, such as SVM, ANN, and RF, are able to cope with unbalanced and noisy datasets in LCC, yielding better classification performance over traditional parametric approaches [16]. However, in these machine learning methods, model hyperparameters should be appropriately set in order to get satisfying classification results. In conventional approaches, hyperparameters are usually set empirically or tuned manually. As a result, these manually set hyperparameters are insufficient to obtain accurate and reliable land cover classification performance and therefore alternative approaches should be sought. Consequently, it is desirable to develop an automated and systematic approach to determine the model hyperparameters before a reliable and accurate classifier being realized.

Bayesian optimization is a promising method for parameter tuning/optimization; however, until now, very few studies have been available to apply it for model hyperparameter tuning, especially for land cover classification with satellite images. Therefore, Bayesian optimization is adopted to tune the hyperparameters of the widely used RF classifier for land cover classification. To summarize, the aim of this study is to optimize RF classifiers and compare against other machine learning methods for urban land cover classification (five classes including building, tree, road, water, and crop field) by using Sentinel-2A/B remote sensing satellite images, where the study area is located in Beijing, China. The optimized RF classifier by Bayesian optimization is compared against the SVM and the random forest with default hyperparameters. In addition, both Red–Green–Blue (RGB) features and full band features are selected for training and testing in different methods so that their effects on classification performance can be assessed. It is expected that a better classification result can be achieved by the optimized RF over the conventional SVM and RF with default hyperparameters.

To be more exact, the main contributions of this study are summarized:

- (1) State-of-the-art earth observation satellite Sentinel-2A/B with the best spatial/spectral/temporal resolution among freely available satellites are evaluated for urban land cover classification;
- (2) Bayesian optimization is drawn to automatically tune the hyperparameters of random forest classifiers for satellite remote sensing image classification.
- (3) Both RGB band and full multispectral bands available on Sentinel-2A/B of an urban scenario with five classes are adopted to evaluate the classification performance of the optimized RF against the SVM and the RF with default hyperparameters.

The remainder of this paper is organized as follows: Section 2 introduces some related work; Section 3 introduces related materials in this case study; Section 4 proposes the methodology of the optimized random forest classifier; Section 5 demonstrates the comparative results by using various methods; Finally, discussion and conclusions along with future work are drawn in Sections 6 and 7, respectively.

## 2. Related Work

Land cover classification is usually formulated as a pixel-wise classification task in the remote sensing community, where the pixels that belong to the same classes are labeled accordingly [17]. With the development of remote sensing technology, the commonly used classifiers can be divided into two branches including the machine learning based classifiers and the deep learning based classifiers. Both of the aforementioned classifiers will be introduced with their advantages and shortcomings in the following sections.

### 2.1. Machine Learning Classifier

Machine learning based classifiers such as Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Decision Tree (DT), and Random forest (RF) are widely used in remote sensing image classification. Zhang proposed to combine the SVM classifier and a mutual information ranking method to obtain more efficient band information, which achieves the state-of-the-art performance in the land cover classification problem [5]. DT classification algorithms have significant potential for land cover mapping problems since they are flexible and robust against the nonlinear and noisy relations among input features and the corresponding class labels [18]. The KNN classifier is widely used because of its implementation simplicity but will perform poorly when training samples distribute unevenly or the sample number of each class is very different [19]. With consideration of the applications of the classifiers in city scenes, SVM and RF classifiers are proved to outperform the traditional classifiers [3]. However, the classification performance of the aforementioned classifiers including RF classifier is highly related to the hyperparameters involved in the model, which normally rely on experience or trial and error tuning. Therefore, in this paper, we take the RF classifier as the baseline and evaluate the influence of the hyperparameters. In this paper, we propose to adopt the Bayesian optimization to automatically optimize the hyperparameters of RF classifier for the city scenario.

### 2.2. Deep Learning Classifier

Artificial neural network (ANN) kicks off the prelude to deep learning, which can simulate the human brain to make the decision [13]. Now, the deep learning based methods mostly take convolutional neural networks (CNN) as the backbone of the algorithms. The CNN architecture can automatically learn the image features via lots of parameters (usually billions), which are trained with a large volume of training data. The CNN classification performance is usually higher than machine learning based classifiers with sufficient computation resource and samples [20]. However, deep learning classifiers highly rely on personal experience and a huge amount of training samples. By considering the limitation of the dataset, in this paper, the machine learning method is considered as a classification approach.

## 3. Materials

This section introduces the related materials involved in the land cover classification problem by using Sentinel-2A/B satellites and machine learning based classifiers. Both satellite imagery and experimental field information are detailed in this section.

### 3.1. Sentinel-2 Satellite Imagery

The earth observations satellite Sentinel-2 series are able to provide remote sensing imageries of high spatial, spectral and temporal resolutions due to its customized Multi-spectral Instrument (MSI) sensor. The spatial resolutions of Band1, Band9, Band10 (60 m),

Band5, Band6, Band7, Band8A, Band11, Band12 (20 m) and Band2, Band3, Band4, Band8 (10 m) can meet various requirements in atmospheric and geophysical parameters correction, vegetation detecting, and land classification [8,10,21]. Moreover, the 13 bands provided by Sentinel-2 series cover visible bands, near Infrared (NIR) bands, and short waved Infrared information (SWIR) bands.

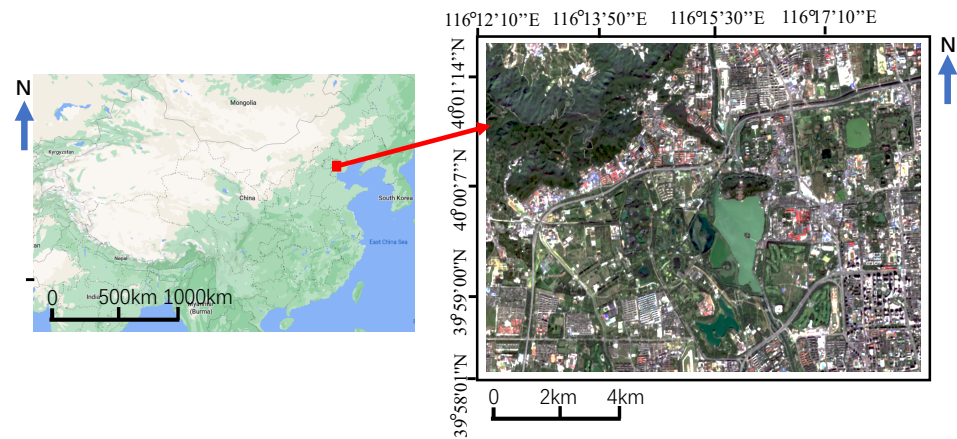
In particular, compared against the popular Landsat 8 and other freely accessible mainstream satellites [5,22,23], Sentinel-2 satellites are able to provide more details in NIR and SWIR bands, which can promote the land cover classification performance in urban monitoring, forest monitoring, and smart farming, among many others [4,24]. Moreover, Sentinel-2 series satellites also improve the temporal resolution, where a 5-day revisit time is available with the introduction of Sentinel2-B. The Sentinel-2 satellite information in terms of band characteristics, wavelength, and spatial resolution are summarized in Table 1, where the band wavelength information is at the central wavelength. It is also noted that Band 10 is particularly for cirrus; therefore, this band is omitted in the land cover classification problem in this study.

**Table 1.** Band information of Sentinel-2A/B.

Band No.	Characteristic	Wavelength ( $\mu\text{m}$ )	Resolution (m)
1	Coastal Aerosol	0.443	60
2	Blue	0.490	10
3	Green	0.560	10
4	Red	0.665	10
5	Near Infrared	0.705	20
6	Near Infrared	0.740	20
7	Near Infrared	0.783	20
8	Near Infrared	0.842	10
8A	Near Infrared	0.865	20
9	Water Vapour	0.945	60
10	Cirrus	1.375	60
11	Shortwave Infrared	1.610	20
12	Shortwave Infrared	2.190	20

### 3.2. Study Area

In this study, to evaluate the classification capabilities for different machine learning based classifiers, an image of  $636 \times 954$  pixels for an urban area in Beijing, China (see Figure 1) is selected. A summary of the geographic location, number of spectral bands, imagery pixels, and cloud cover is displayed in Table 2. In particular, all satellite images of Sentinel series could be freely downloaded from Sentinel Open Hub (<https://scihub.copernicus.eu/>). The officially customized software Sentinel Application Platform (SNAP) is utilized to import all the sensor information and export tailored data for follow-up analysis in comparison to other geo-software such as ENVI [25,26]. The selected field is a typical area composed of five main classes: buildings (such as universities, factories and companies), trees, roads, water, and crop fields.



**Figure 1.** The study area in Beijing, China and its zoom-in satellite image (false-color RGB).

**Table 2.** A summary of the study area and satellite image.

Location	No. of Bands	Image Size	Cloud Cover (%)
40°01'23" N, 116°12'10" E	12	6360 m × 9540 m	1.67
39°58'01" N, 116°18'52" E			

#### 4. Methodology

This section introduces the overall methodology including problems formulation, the developed framework, and algorithm of RF with Bayesian optimization.

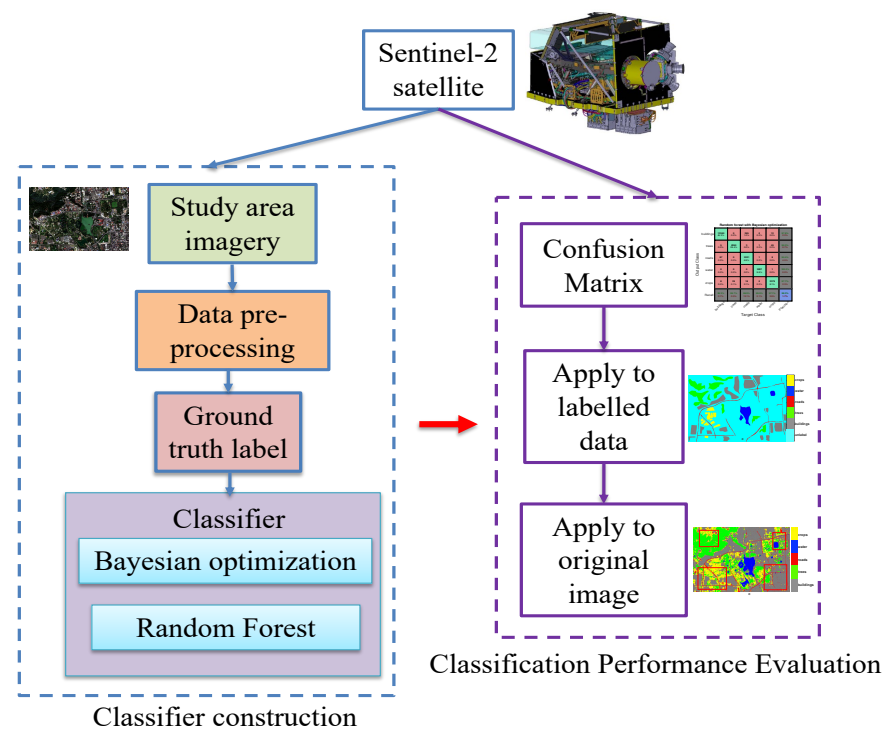
##### 4.1. Problems Formulation

The land cover classification problem in this study can be formulated as a supervised classification problem where bands (or typical indices) are selected as features into a supervised classifier for training. In this study, the Sentinel-2A/B image pixels are represented by  $\mathcal{D} = \{1, \dots, n\}$  where  $n$  means the total number of individual pixels in the original satellite map. Here, the pixel matrix of this image with  $f$  being the number of features (bands or indices) is defined as  $x = (x_1; \dots; x_n) \in R^{n \times f}$ . Let  $\mathcal{L} = \{1, \dots, k\}$  and  $C = (c_1, \dots, c_n)$  be a set of class labels and classification map corresponding to the label, respectively, where  $k$  denotes the number of class. Therefore, the training dataset  $\mathcal{T}$  can be generated by the number of features  $f$  and the corresponding labels  $C$  in the form of  $\mathcal{T} = \{(x_1, c_1), \dots, (x_\tau, c_\tau)\}$  with the number of training samples  $\tau$ . As a result, after the classification model is built, the classification evaluation matrix and also the corresponding classification map can be generated by sending the training dataset  $\mathcal{T}$  into the classifier. The aim of this study is to evaluate the classification performance of three different supervised classifiers including the random forest classifier with Bayesian parameter optimization.

##### 4.2. Land Cover Classification Framework

It can be seen from Figure 2 that the framework can be divided into two main stages: classifier construction and classification performance evaluation. The classifier construction includes data pre-processing, training data labeling, and RF with Bayesian optimization, which are described in details in the following subsections.





**Figure 2.** The developed framework for land cover classification by using Sentinel-2 satellite and machine learning based classifiers.

#### 4.2.1. Remote Sensing Image Pre-Processing

Sentinel-2B level 2A product image was obtained on 4 October 2020 for the region of interest. Three different steps are adopted to pre-process the raw image including atmospheric correction, image resampling, and field subset. This image is atmospherically corrected based on Atmospheric/Topographic Correction for Satellite Imagery proposed by Richter [27]. Such a method is based on libRadtran radiative transfer model so that the image quality can be guaranteed [5]. Due to the difference of image spatial resolution in different bands, the resampling process is done so that a consistent image resolution can be guaranteed. Finally, the subset process allows for selecting the region of interest (ROI) from the downloaded large images. All of the pre-processing work is finished by Sentinel Application Platform (SNAP) software which is particularly designed for Sentinel series satellites.

#### 4.2.2. Image Labeling

According to [12], Band 10 is especially for cirrus recording and thus being omitted in this study. The remaining twelve bands are selected as features for pixel-wise classification. Ground-truth labeling is necessary in supervised learning tasks in order to build the model. Thus, this image is labeled based on manual interpretation of the original Sentinel-2 satellite image (in false-color RGB format) along with Google map images and on-site checking. The ground-truth of five classes including building (No. 1), tree (No. 2), road (No. 3), water (No. 4), and crop field (No. 5) are labeled in Matlab software (2017b) using polygons of different shapes for each class (see Figure 3) and 'Un' denotes the unlabeled data. By using the labeled pixels, the average reflectance over five different land cover classes can be compared and shown in Figure 4, which lays the foundation for discriminating various classes by various machine learning based classifiers.

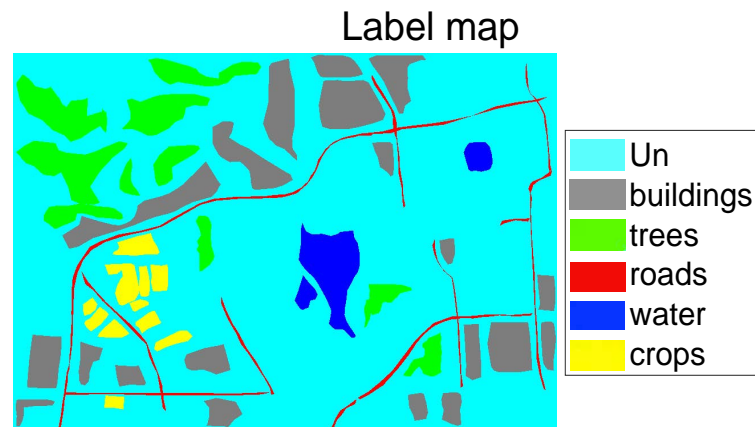


Figure 3. Labeled map for training data generation.

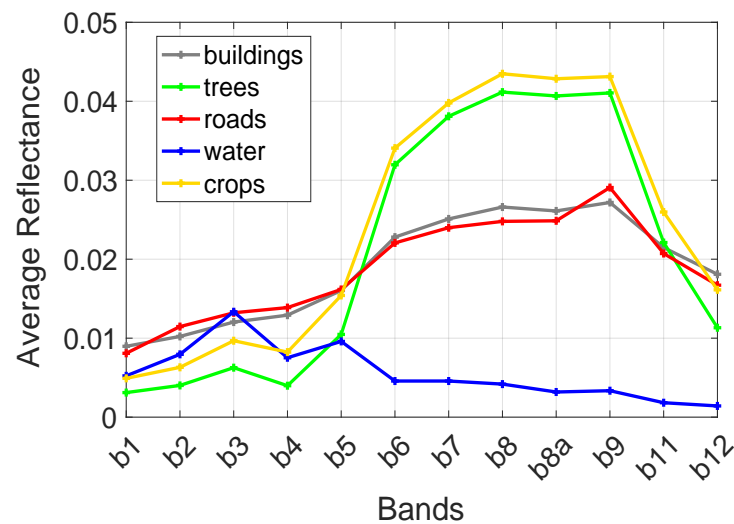


Figure 4. Band reflectance for five different classes including building, trees, roads, water, and crops.

In order to compare the discrimination ability of visible images and multispectral images [20], two training datasets are separately selected to evaluate the performance of various methods: RGB features (e.g., only Red, Green, and Blue bands) and full 12 band features. In addition, the labeled dataset is divided into the training dataset and testing dataset to avoid the problem of over-fitting, which is a common issue in machine learning based classifiers. In this study, the proportions of training data and testing data are set to be 80% and 20%, respectively.

#### 4.2.3. RF with Bayesian Optimization

An appropriate classifier can build the implicit relationship between feature information (e.g., band information) and target information (e.g., five classes in this study) by supervised learning from training datasets. Given the trained classification model, prediction can be made on unseen data to generate the corresponding class labels. A number of classifiers have been used in the literature for supervised learning, such as SVM, RF, decision trees, nearest neighbor, and neural network. It has been shown that RF possesses more advantages in avoiding over-fitting while with a relatively low computation load [28].

RF is an ensemble learning based classification approach with a large number of decision trees constructed in the training process, where the final output integrates the outcome class of individual decision trees [15,28]. Such a method is able to avoid over-fitting and at the same time is much more robust than a single decision tree. It is also shown in the existing studies that the RF method is able to achieve a high accuracy, a

good robustness, and less computation load [29,30]. However, some hyperparameters in this method are necessary to be tuned according to the tasks of interest so that a better classification performance can be fully realized. According to [17], Bayesian optimization can be adopted to automatically tune the hyperparameters, where the details are summarized in Algorithm 1. Due to the lack of space, the basic RF algorithm is referred to the existing studies [31]. To demonstrate the advantages of the proposed RF with Bayesian hyperparameter optimization, its performance is compared against the conventional SVM and the RF with default parameters.

---

**Algorithm 1** Optimized random forest classifier

---

- (a) **Initial settings:** RF is composed of a large number of decision trees. Therefore, the number of decision trees need to be first determined (e.g., 150) because a large number of individual trees can improve the discrimination ability. At the same time, certain stopping rule needs (e.g., here the rule is maximum evaluation time) can also be defined to end the optimization iteration.
- (b) **Objective function:** Set the hyperparameters  $\mathcal{H} \in \Omega$  that should be tuned such as MinLeafSize (mls), NumPredictorsToSample (npts). The objective function in Bayesian optimization is then defined as the mean of the out-of-bag error ('oobErr') to avoid overfitting, so that the optimization problem is given by

$$\mathcal{H}_{opt} = \arg \min_{\mathcal{H} \in \Omega} oobErr(\mathcal{H}). \quad (1)$$

- (c) **Bayesian optimization:** The Bayesian optimization method aims to automatically and optimally suggest new parameters by fitting a Gaussian process model  $G$  for the existing data points  $\{\mathcal{H}_i, oobErr(\mathcal{H}_i)\}$  and find a new point after updating to minimize the objective function based on the posterior distribution function  $G$ .
  - (d) **Optimized hyperparameters:** The optimization process will end when the stopping criterion in Step (a) is satisfied and these optimized hyperparameters will be put into the random forest classifier.
  - (e) **Optimized RF classifier:** By using the optimized hyperparameters, the optimized random forest can be used for performance assessment (e.g., confusion matrix calculation) and land cover classification applications (e.g., to the whole image of interest).
- 

All algorithms (SVM, conventional RF, and RF with a Bayesian optimization method) involved in this study are implemented in Matlab of version 2017b. For the proposed method, there are a total of two hyperparameters being tuned including minimum leaf size (*mls*) and the number of predictors to sample (*npts*), where *mls* is to control the depth of the trees and *npts* determines the amount of predictors to sample at each node when growing the trees. By default, *mls* is set as 1 for classification, and *npts* is equal to the square root of the total number of variables for classification. In the proposed method settings, the prior information of *mls* is between 1 to 20. In addition, the prior of parameter *npts* is between 1 to *nf*, where *nf* means number of features. 'oobErr' is set as 'on' to store information on what observations are out of bag for each tree, and this can be used to compute the predicted class probabilities for each tree in the ensemble. The number of trees is set as 150 and the maximum objective function evaluation time (stopping rule) is set to be the default value of 30 times.

#### 4.3. Classification Performance Evaluation

In this work, 80% and 20% of the labeled pixels are randomly selected for training and testing, respectively, where the performance accuracy is calculated based on the testing dataset to avoid the problem of overfitting. In particular, True Positive (TP) denotes the correctly predicted positive values; False Positive (FP) is the value where actual class is negative and the predicted class is positive; False Negative (FN) means the scenario where the actual class is positive, but the predicted class is negative [29]. Various evaluation



metrics can be defined based on these values such as accuracy, precision, and recall as in Equations (2) and (3). In addition, confusion matrix is also commonly used to visually assess the performance of various methods. In the confusion matrix, the rows denote the output class (predicted class) and the columns represent the target class (groundtruth class). A detailed explanation of the confusion matrix will be introduced where necessary in the following parts.

The accuracy of the classification model for a particular class is defined by:

$$Accuracy = \frac{\sum TP}{All} \tag{2}$$

In order to properly assess model performance for unbalanced datasets, Precision and Recall are also usually introduced [29,32], which for a typical class are defined by

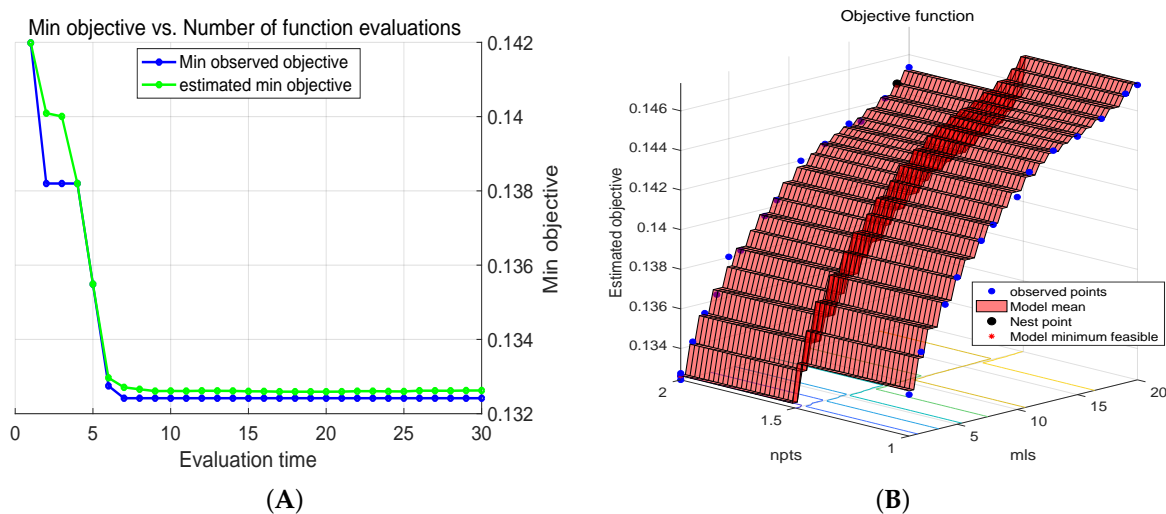
$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \tag{3}$$

### 5. Results

This section summarizes the performance evaluation results for different machine learning based classifiers with different features (RGB band features and full multispectral band features). In addition, the spatial classification maps are also generated for visual inspection wherever is necessary.

#### 5.1. RGB Band Features

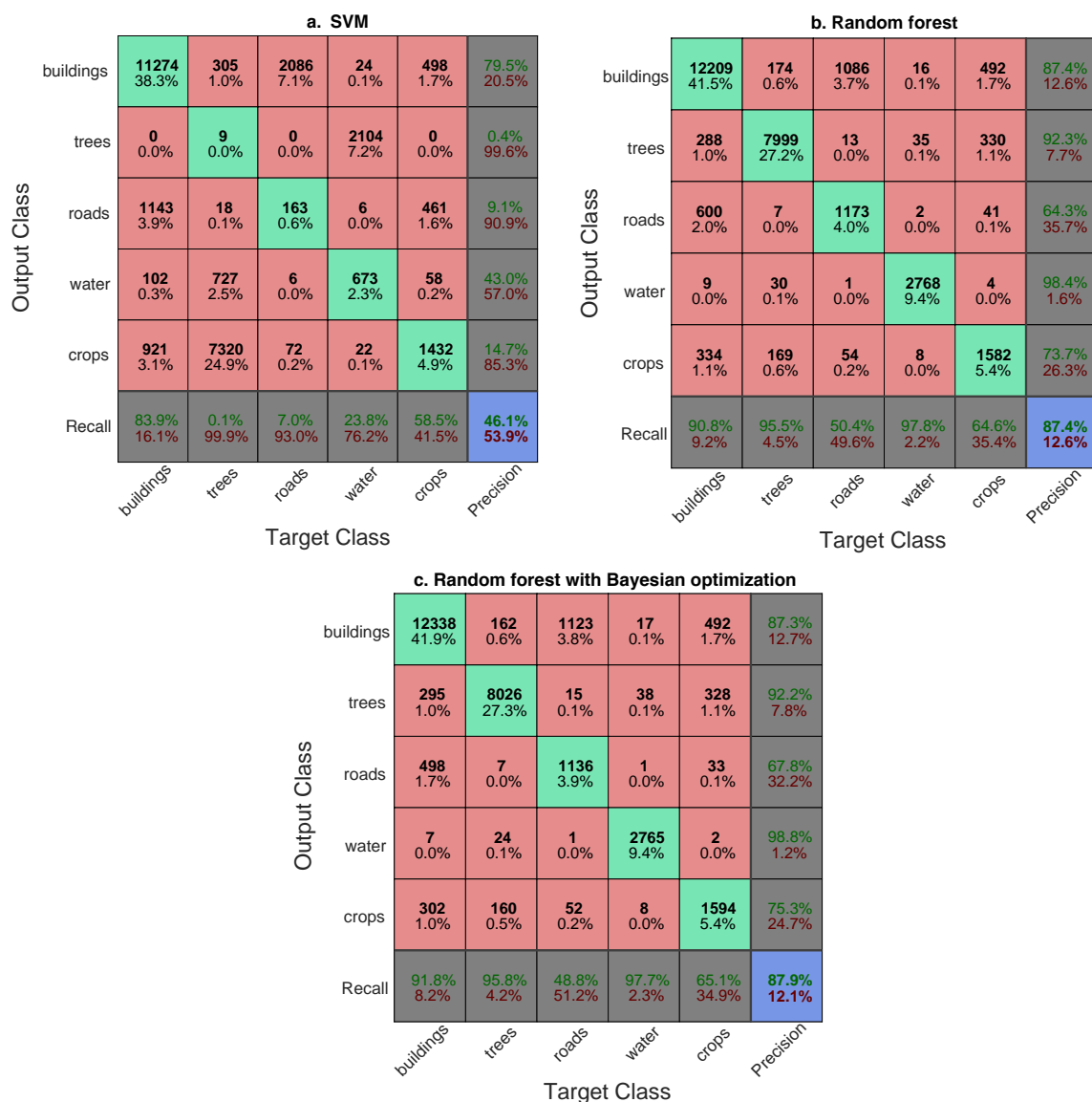
In the first set of models, RGB band features are selected for the three methods including SVM, RF with default parameters, and the RF with Bayesian hyperparameter optimization. The Bayesian hyperparameter optimization results are shown in Figure 5, where subplot A shows the estimated minimum objective over evaluation time, and subplot B shows the estimated objective over different hyperparameter combinations. It can be observed that the estimated objective function achieves equilibrium after a few evaluations and is close to the observed objective, and the minimum objective function is achieved by the optimized hyperparameters vector  $mls = 1$  and  $npts = 2$  (default parameters:  $mls = 1$ ,  $npts = 1$ ).



**Figure 5.** (A) Minimum objectives over evaluation time by using RGB features; (B) estimated fitness function values over the combinations of  $npts$  and  $mls$  using RGB features.

The confusion matrices for the three machine learning based classifiers are displayed in Figure 6. In these matrices, target classes denote the truth labels, whereas the output classes mean the classifier predicted labels. The diagonal cell in green shows the number

and the corresponding percentage for correctly classified pixels and the off-diagonal cell indicates the misclassified pixels. Taking the proposed algorithm as an example, for the “building” class, 12,338 pixels in green is TP, another 1794 (162 + 1123 + 17 + 492) pixels in red in the first row is FP, and 1102 (295 + 498 + 7 + 302) pixels in red in the first column is FN. Thus, Precision for “buildings” class is  $12,338 / (12,338 + 1794) = 87.3\%$  and similarly Recall for “buildings” is  $12,338 / (12,338 + 1102) = 91.8\%$ . The overall accuracy is 87.9%. In comparison to SVM and the RF with default parameters, the proposed method obtains the best classification performance, which marginally improved against the conventional RF by 0.5%. However, the overall accuracy of SVM algorithm is only 46.1%, which is much less than random forest classifiers. The main reason is the inappropriateness of the SVM algorithm for the land cover classification problem with only RGB band information. A comparison for different methods is shown in Table 3 showing that the optimized random forest method achieves the highest OA and kappa value.



**Figure 6.** Confusion matrices of three machine learning based classifiers (a) SVM, (b) Random forest, (c) Random forest with Bayesian optimization) by using RGB band features.

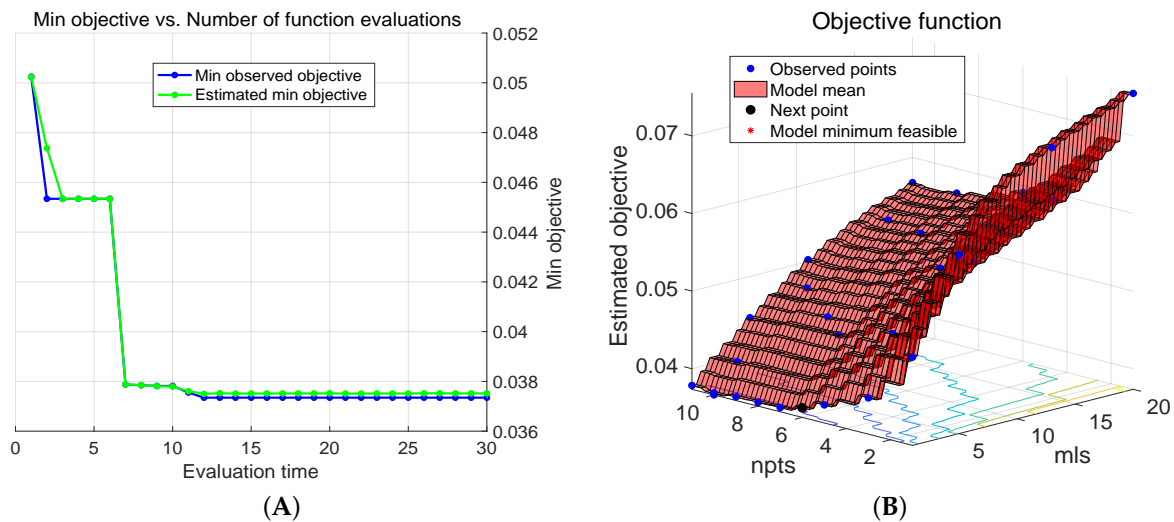
**Table 3.** A comparison between different methods using OA and kappa value (RGB features).

Methods	OA	Kappa Value
SVM	0.4605	0.2526
Random forest	0.8744	0.8152
Optimized random forest	0.8788	0.8210

5.2. Full Multispectral Band Features

5.2.1. Classification Performance Evaluation

It can be seen from Figure 4 that, in addition to the commonly used RGB bands, other bands (e.g., NIR, SWIR) can also provide vital discrimination information for land cover classification and therefore all multispectral band features are also assessed for the three machine learning based models in this subsection. Similar to the case of RGB band features in Section 5.1, the results of Bayesian hyperparameter optimization for full multispectral band features are displayed in Figure 7 including the minimum objectives over time and the estimated objective function values over different combinations of *mls* and *npts*. The optimized hyperparameters vector shows that *mls* = 1 and *npts* = 10 (default parameters: *mls* = 1, *npts* = 6). In addition, the out-of-bag error over the number of trees is also displayed in Figure 8. The smaller the out-of-bag error is, the more accurate the classifier will be. It can be seen that the error using Bayesian optimization is smaller than that of the RF with default parameters when the same number of trees is used. Therefore, the RF with Bayesian optimization possesses better performance over the one with default hyperparameters. Under this hyperparameter setting, the confusion matrix of the optimized RF is shown against the ones for SVM and RF with default parameters in Figure 9.



**Figure 7.** (A) Minimum objectives over evaluation time by using full multispectral band features; (B) estimated fitness function values over the combinations of *npts* and *mls* by using full multispectral band features.

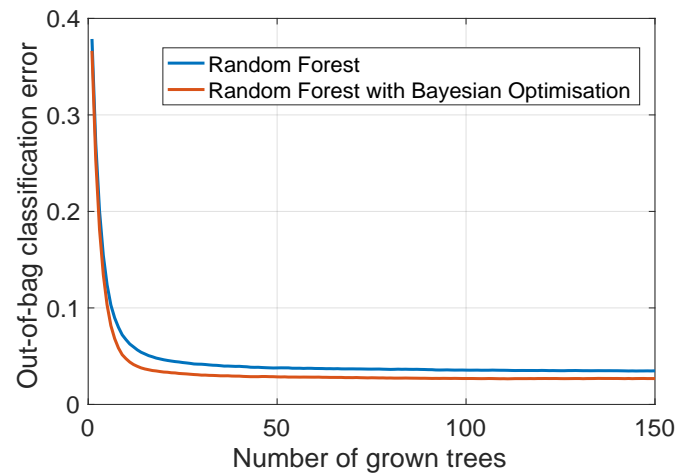


Figure 8. The out-of-bag error over the number of grown trees using different methods.

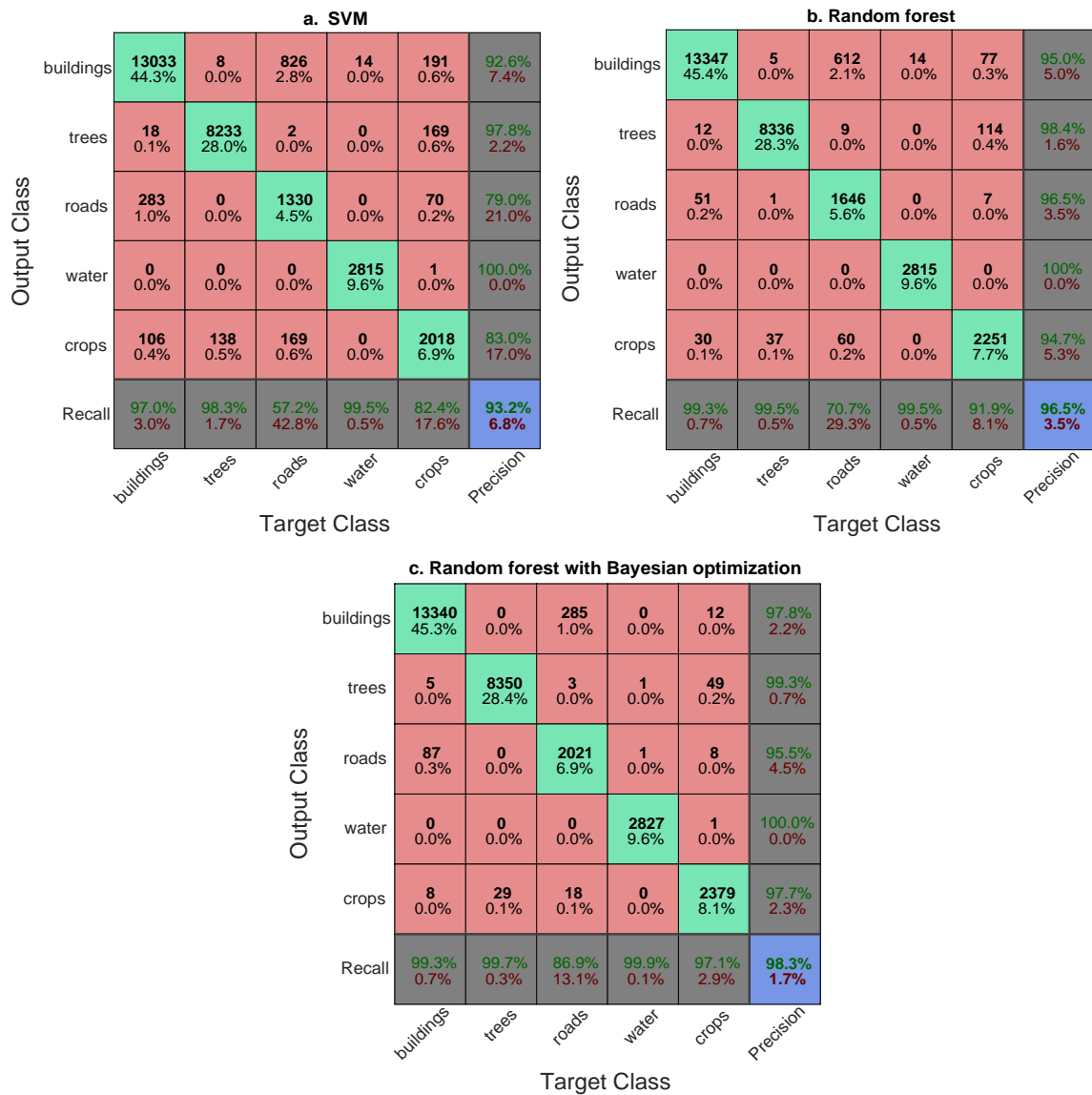


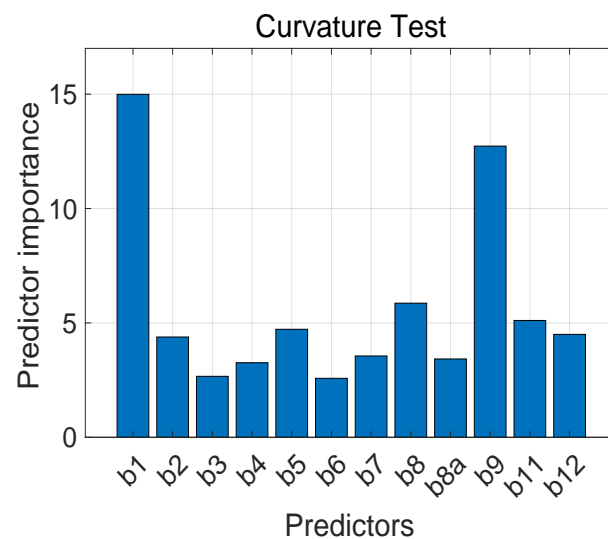
Figure 9. Confusion matrix of the classifiers (a) SVM, (b) Random forest, (c) Random forest with Bayesian optimization) with full multispectral band features.

The confusion matrices for the three methods are shown in Figure 9. It can be seen that incorporating more band information in the range of NIR and SWIR of the Sentinel-2A/B satellite can significantly improve the land cover classification performance. For example, the classification accuracy change of SVM is 46.1%→93.2%, RF is 87.4%→96.5%, and RF with Bayesian optimization is 87.9%→98.3%. This observation clearly demonstrates that incorporating more related band information can significantly improve the classification performance, as it can be seen from Figure 4 that, in addition to RGB bands, NIR bands and SWIR bands also have a strong discrimination capability. On the other hand, in comparison with SVM and the RF with default parameters, the RF with Bayesian hyperparameter optimization shows the best performance in terms of Precision (user’s accuracy), Recall (producer’s accuracy), and Overall Accuracy [33]. The comparison between different methods is displayed in Table 4. This again shows the advantages of optimizing the hyperparameters of RF classifiers.

**Table 4.** A comparison between different methods using OA and kappa value (full bands features).

Methods	OA	Kappa Value
SVM	0.9322	0.9002
Random forest	0.9650	0.9485
Optimized random forest	0.9834	0.9751

In addition, the curvature test [34] is capable of evaluating feature scores to reflect their contribution and usefulness in the classification task. The curvature test result for the RF with Bayesian optimization is displayed in Figure 10, where the usefulness of different bands is shown with a high value meaning a higher predictor importance.

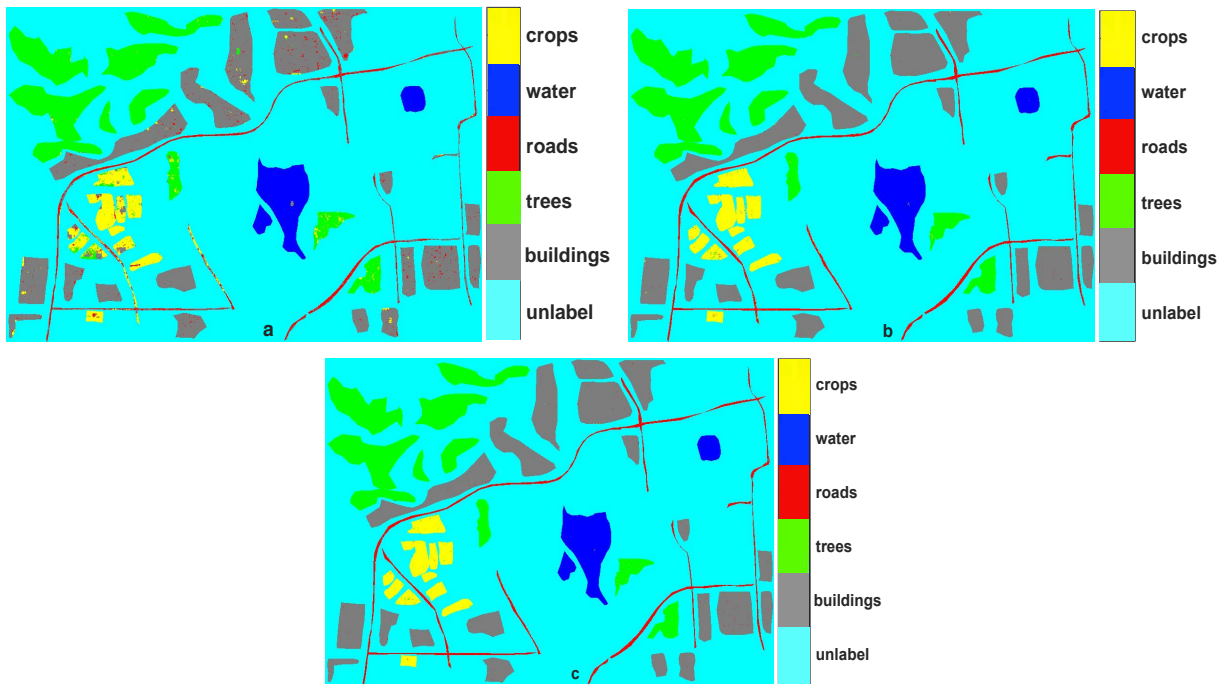


**Figure 10.** Curvature test result for the features in the random forest classifier with Bayesian optimization.

### 5.2.2. Classification Maps

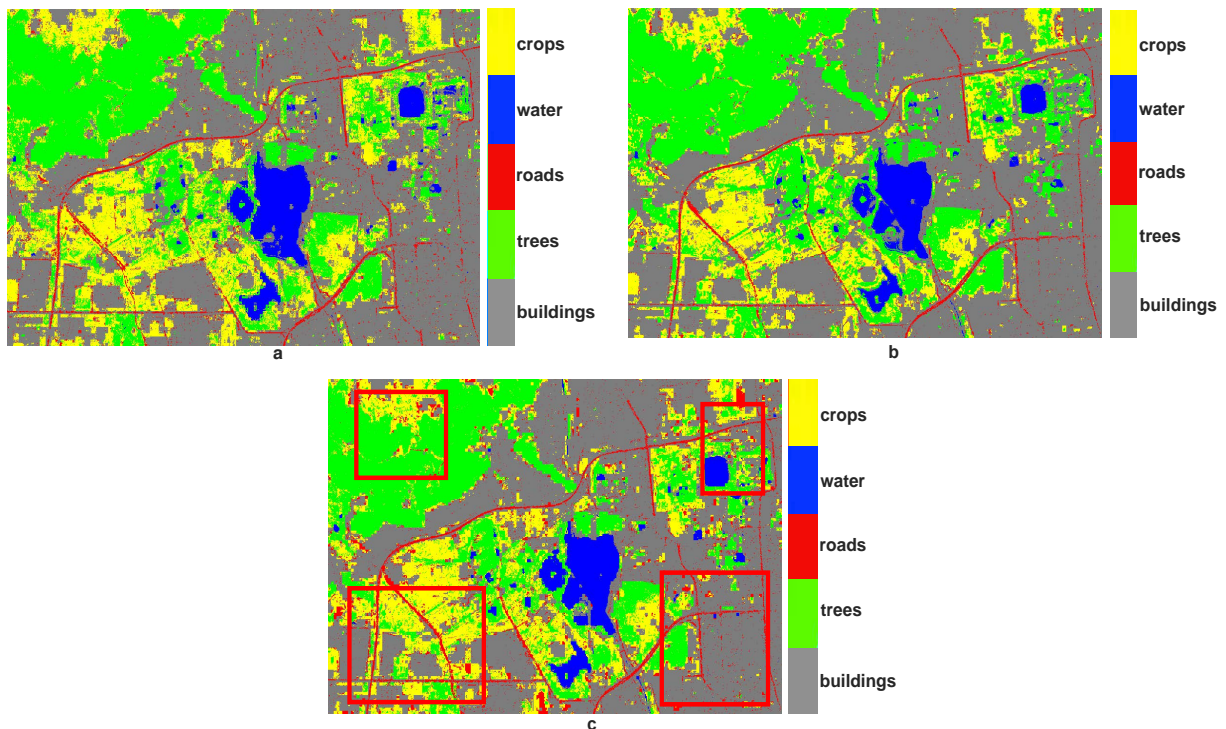
Quantitative results are very useful to compare the performance of different models, it would also be visually useful to assess the spatial classification maps by different methods. To this end, the three trained models are applied to both the labeled areas and the whole images, respectively. The trained models with all multispectral band features are first applied to the labeled areas, where the spatial classification maps are shown in Figure 11. It can be seen that all three models generate satisfying spatial classification maps; however, the RF with Bayesian optimization has the fewest wrongly classified pixels and noises.





**Figure 11.** Classification map for labeled areas by three models with all band features: (a) SVM; (b) conventional RF; (c) RF with Bayesian optimization.

In addition to applying the trained models to the labeled areas, it would also be interesting and useful to see the classification results on the whole satellite images for the purpose of urban land cover analysis. Based on the five labeled classes, the classification maps by using the three different models with full band features are shown in Figure 12. It can be seen that the RF with a Bayesian optimization approach again generates the best land cover classification result, which has fewer noises by comparing the areas highlighted by red rectangles.



**Figure 12.** Classification maps using different trained models to the original image: (a) SVM; (b) conventional RF; (c) RF with Bayesian optimization (red rectangles mean obvious improvements for different regions among various classifiers).

## 6. Discussion

The RF with Bayesian hyperparameter optimization method presented in Section 4 shows better performance in terms of precision, recall, and accuracy of an urban land cover classification example. The RGB band features and full multispectral band features have been discussed and examined by three different classifiers including SVM, RF, and RF with Bayesian optimization. In order to evaluate the performance of different models, pixel-wise classification is used in this paper. Then, Equations (2) and (3) are used to estimate the evaluation values (accuracy, precision, and recall) from the confusion matrices. From the confusion matrices, all three classifiers show improved performance by incorporating more related band information. The classification accuracy of SVM is increased by 47.1%, RF is improved by 9.1%, while the RF with Bayesian optimization is enhanced by 10.4%. Compared with RGB bands, NIR bands and SWIR bands (multispectral band features) provide more precise results (the minimum overall precision provided by SVM is over 93.2%).

Meanwhile, in terms of precision, recall, and accuracy, the RF with Bayesian hyperparameter optimization gives the best results. For the overall precision of an urban land cover classification example given in Section 3, the RF with Bayesian optimization is 0.5% higher than RF and 41.8% higher than SVM by using RGB band features, respectively. Simultaneously, the RF with Bayesian optimization is 1.8% higher than RF and 5.1% higher than SVM by using multispectral band features, individually. Moreover, quantitative results are also presented to compare the differences of the performance provided by different models through the classification map for labeled areas with all band features and full band features. Both classification maps show that the RF with a Bayesian hyperparameter optimization model generates the best land cover visualization results with less noises and error classified pixels.

There are also a number of issues that are worth investigation when the proposed method is to be applied in real-world applications. For instance, the spatial resolution of Sentinel-2 satellite is about 10 m; as a result, some pixels (in particular the ones at boundaries of surface classes) are actually mixed pixels involving different surface classes. The classification performance is therefore not accurate enough for these mixed pixels and a better result may be obtained with a higher spatial resolution. In addition, the cloud may have adverse effects on classification performance. This can be partially addressed by either taking the satellite image with a low cloud coverage or taking the median value of the satellite images within a time interval.

## 7. Conclusions and Future Work

This paper investigates the problem of urban land cover classification by using Sentinel-2 satellite remote sensing imageries and machine learning based classifiers. In particular, Bayesian optimization is drawn to automatically tune the hyperparameters of random forest classifiers so that its performance can be improved. An urban land cover classification example in Beijing, China is drawn to demonstrate the performance of optimized random forest classifier against the random forest with default parameters and the classical support vector machine (SVM) classifier. In performance evaluation, RGB band features of Sentinel-2 satellite are firstly considered by employing three different methods. The results show that the optimized random forest classifier achieves the best performance with overall accuracy (OA: 0.879), kappa coefficient (0.8210), whereas SVM achieves a low OA and kappa value of 0.461 and 0.2526, respectively. Then, full band features are evaluated by the three methods, and it is shown that the optimized random forest still possesses the highest value of OA (0.983) and kappa value (0.9751). In addition, the classifiers with more useful band information outperform the ones with only RGB band information. Therefore, the developed random forest classifier with Bayesian hyperparameter optimization is expected to provide better urban land cover classification performance so that city managements can be achieved in a more precise manner. Meanwhile, with a suitable training dataset, this method can also find a wide range of applications in land resource management, green infrastructure monitoring, disaster management, and agricultural applications.

Although the results in this study are quite promising, there is still much room for further improvement. For instance, due to the logistics issues, only a small set of training dataset is used to assess the performance of the algorithms. With the advent of a more labeled dataset, the performance can be evaluated in a more accurate manner. Moreover, this study is mainly focused on applying spectral information for land cover classification, and spatial information can also provide vital information. In addition to machine learning methods, the popular deep learning approaches such as a Convolutional Neural Network (CNN) can also be drawn to simultaneously learn the spectral and spatial information in an end-to-end manner and possible improved performance.

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

**Funding:** This work was supported by the Fundamental Research Funds for the China Central Universities of USTB (FRF-DF-19-002), Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB (BK20BE014).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

## References

1. Luo, X.; Tong, X.; Pan, H. Integrating Multiresolution and Multitemporal Sentinel-2 Imagery for Land-Cover Mapping in the Xiongan New Area, China. *IEEE Trans. Geosci. Remote Sens.* **2020**. [[CrossRef](#)]
2. Lin, L.; Hao, Z.; Post, C.J.; Mikhailova, E.A.; Yu, K.; Yang, L.; Liu, J. Monitoring Land Cover Change on a Rapidly Urbanizing Island Using Google Earth Engine. *Appl. Sci.* **2020**, *10*, 7336. [[CrossRef](#)]
3. 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](#)] [[PubMed](#)]
4. Tong, X.Y.; Xia, G.S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [[CrossRef](#)]
5. Zhang, T.X.; Su, J.Y.; Liu, C.J.; Chen, W.H. Potential bands of sentinel-2A satellite for classification problems in precision agriculture. *Int. J. Autom. Comput.* **2019**, *16*, 16–26. [[CrossRef](#)]
6. Kranjčić, N.; Medak, D.; Župan, R.; Rezo, M. Machine learning methods for classification of the green infrastructure in city areas. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 463. [[CrossRef](#)]
7. Acharya, T.D.; Yang, I.T.; Lee, D.H. Land cover classification using a KOMPSAT-3A multi-spectral satellite image. *Appl. Sci.* **2016**, *6*, 371. [[CrossRef](#)]
8. Van Der Werff, H.; Van Der Meer, F. Sentinel-2A MSI and Landsat 8 OLI provide data continuity for geological remote sensing. *Remote Sens.* **2016**, *8*, 883. [[CrossRef](#)]
9. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [[CrossRef](#)]
10. Martimort, P.; Arino, O.; Berger, M.; Biasutti, R.; Carnicero, B.; Del Bello, U.; Fernandez, V.; Gascon, F.; Silvestrin, P.; Spoto, F.; et al. Sentinel-2 optical high resolution mission for GMES operational services. In Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium, Barcelona, Spain, 23–28 July 2007; pp. 2677–2680.
11. Kumar, P.; Prasad, R.; Gupta, D.; Mishra, V.; Vishwakarma, A.; Yadav, V.; Bala, R.; Choudhary, A.; Avtar, R. Estimation of winter wheat crop growth parameters using time series Sentinel-1A SAR data. *Geocarto Int.* **2018**, *33*, 942–956. [[CrossRef](#)]
12. Zhang, T.; Su, J.; Liu, C.; Chen, W.H.; Liu, H.; Liu, G. Band selection in Sentinel-2 satellite for agriculture applications. In Proceedings of the 2017 23rd International Conference on Automation and Computing (ICAC), Huddersfield, UK, 7–8 September 2017; pp. 1–6.
13. Hramov, A.E.; Maksimenko, V.A.; Pchelintseva, S.V.; Runnova, A.E.; Grubov, V.V.; Musatov, V.Y.; Zhuravlev, M.O.; Koronovskii, A.A.; Pisarchik, A.N. Classifying the perceptual interpretations of a bistable image using EEG and artificial neural networks. *Front. Neurosci.* **2017**, *11*, 674. [[CrossRef](#)] [[PubMed](#)]

14. Su, J.; Coombes, M.; Liu, C.; Guo, L.; Fang, S.; Chen, W.H. Machine Learning Based Crop Drought Mapping System by UAV Remote Sensing RGB Imagery. *Unmanned Syst.* **2020**. [[CrossRef](#)]
15. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.
16. Dobrinić, D.; Medak, D.; Gašparović, M. Integration of Multitemporal SENTINEL-1 and SENTINEL-2 Imagery for Land-Cover Classification Using Machine Learning Methods. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *43*, 91–98. [[CrossRef](#)]
17. Zhang, T.; Su, J.; Liu, C.; Chen, W.H. Bayesian calibration of AquaCrop model for winter wheat by assimilating UAV multi-spectral images. *Comput. Electron. Agric.* **2019**, *167*, 105052. [[CrossRef](#)]
18. Friedl, M.A.; Brodley, C.E. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* **1997**, *61*, 399–409. [[CrossRef](#)]
19. Li, Y.; Cheng, B. An improved k-nearest neighbor algorithm and its application to high resolution remote sensing image classification. In Proceedings of the 2009 17th International Conference on Geoinformatics, Fairfax, VA, USA, 12–14 August 2009; pp. 1–4.
20. Su, J.; Yi, D.; Su, B.; Mi, Z.; Liu, C.; Hu, X.; Xu, X.; Guo, L.; Chen, W.H. Aerial Visual Perception in Smart Farming: Field Study of Wheat Yellow Rust Monitoring. *IEEE Trans. Ind. Informat.* **2020**. [[CrossRef](#)]
21. Clevers, J.; Kooistra, L.; Van Den Brande, M. Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sens.* **2017**, *9*, 405. [[CrossRef](#)]
22. Li, H.; Chen, Z.X.; Jiang, Z.W.; Wu, W.B.; Ren, J.Q.; Liu, B.; Tuya, H. Comparative analysis of GF-1, HJ-1, and Landsat-8 data for estimating the leaf area index of winter wheat. *J. Integr. Agric.* **2017**, *16*, 266–285. [[CrossRef](#)]
23. Roy, D.P.; Wulder, M.; Loveland, T.R.; Woodcock, C.; Allen, R.; Anderson, M.; Helder, D.; Irons, J.; Johnson, D.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [[CrossRef](#)]
24. Tavares, P.; Beltrão, N.; Guimarães, U.; Teodoro, A. Integration of Sentinel-1 and Sentinel-2 for Classification and LULC Mapping in the Urban Area of Belém, Eastern Brazilian Amazon. *Sensors* **2019**, *19*, 1140. [[CrossRef](#)] [[PubMed](#)]
25. Akanwa, A.O.; Okeke, F.I.; Nnodu, V.C.; Iortyom, E.T. Quarrying and its effect on vegetation cover for a sustainable development using high-resolution satellite image and GIS. *Environ. Earth Ence* **2017**, *76*, 505. [[CrossRef](#)]
26. Shoko, C.; Mutanga, O. Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 32–40. [[CrossRef](#)]
27. Richter, R.; Schläpfer, D. *Atmospheric/Topographic Correction For Satellite Imagery*; DLR Report DLR-IB; DLR: Wessling, Germany, 2005; pp. 565–601.
28. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [[CrossRef](#)]
29. Su, J.; Liu, C.; Coombes, M.; Hu, X.; Wang, C.; Xu, X.; Li, Q.; Guo, L.; Chen, W.H. Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. *Comput. Electron. Agric.* **2018**, *155*, 157–166. [[CrossRef](#)]
30. Su, J.; Liu, C.; Hu, X.; Xu, X.; Guo, L.; Chen, W.H. Spatio-temporal monitoring of wheat yellow rust using UAV multispectral imagery. *Comput. Electron. Agric.* **2019**, *167*, 105035. [[CrossRef](#)]
31. Reis, I.; Baron, D.; Shahaf, S. Probabilistic random forest: A machine learning algorithm for noisy data sets. *Astron. J.* **2018**, *157*, 16. [[CrossRef](#)]
32. Yi, D.; Su, J.; Liu, C.; Chen, W.H. Personalized driver workload inference by learning from vehicle related measurements. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**. [[CrossRef](#)]
33. Barsi, Á.; Kugler, Z.; László, I.; Szabó, G.; Abdulmutalib, H. Accuracy Dimensions in Remote Sensing. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*. [[CrossRef](#)]
34. Loh, W.Y.; Shih, Y.S. Split selection methods for classification trees. *Stat. Sin.* **1997**, *7*, 815–840.