Object-Based Supervised Change Detection Using Extracted Features from VHR Imagery

Farzaneh Naeimiasl [1] iD *, Reza Shah-Hosseini [2]✉* iD

1. School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. E-mail: *farzanenaeimi@ut.ac.ir*

2. Corresponding author, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. E-mail: *rshahosseini@ut.ac.ir*

Article Info	ABSTRACT
	This study aims to tackle challenges in change detection within Very High Resolution (VHR) satellite imagery, such as complex feature interactions, noise sensitivity, and detailed land use alterations, by developing a robust detection methodology using bi-temporal images. *We propose a novel approach integrating object-based, feature-based, and dual learning-based techniques. Initially, VHR images are segmented separately to reduce complexity and preserve pixel relationships. Spectral bands are then enhanced with textural, mathematical, and geometrical features extracted from paired images for a thorough conceptual analysis. Finally, we apply two supervised classifier categories—individual and ensemble—to produce a binary change map distinguishing changed and unchanged regions.* *Experimental evaluation on two VHR datasets shows our method significantly surpasses traditional techniques. It achieves an F1-score of 99.02% and an Intersection over Union (IoU) of 96.78%. The generated change maps feature numerous homogeneous, detailed areas, confirming effective detection performance.* *The integration of object-based, feature-based, and learning-based approaches results in a comprehensive feature extraction framework that enhances change detection accuracy. Our method contributes detailed and reliable change maps, representing a significant advancement in the field of VHR satellite image analysis.*

1. Introduction

Identifying land cover evolutions is a crucial undertaking that provides invaluable insights into environmental dynamics. Various types of changes, including shifts in land use, alterations in climate patterns, and the occurrence of natural disasters have profound implications for ecosystems, human societies, and global well-being. Monitoring these changes allows us to assess the impact of human activities on the environment, identify emerging threats, and implement proactive measures to mitigate potential risks. It also facilitates the tracking of long-term trends, aiding us in promoting environmental conservation (Z. Lv et al., 2022).

Remote Sensing (RS) technology provides a wide range of data with extensive coverage and short revisit times, which has evolved into a primary resource for continuous environmental monitoring and detecting changes on the land surface (Parelius, 2023).

Change Detection (CD) is the process of identifying differences by analyzing the variability in a set of images captured at different times of the same geographical area (H. Chen et al., 2022). The goal is to monitor and understand changes in land surface or other phenomena by observing the differences in multi-temporal images.

The utilization of remote sensing data in change detection has found widespread applications including, but not limited to, land cover mapping, environmental monitoring, disaster damage assessment, urban expansion investigation (Zhang et al., 2020), vegetation change detection, evaluation of desertification and deforestation, and fire detection (Zerrouki et al., 2021), and other land management practices (Shafique et al., 2022). These applications also extend to recognizing climate change indicators such as heatwaves, droughts, floods, hurricanes, and more (L. Khelifi & M. Mignotte, 2020).

Advancements in the RS imaging framework and increased accessibility to Very High Resolution (VHR) satellite images have significantly broadened the scope for employing change detection in high-resolution bi-temporal images across diverse domains. With their exceptionally high spatial resolution, VHR images are particularly well-suited for precisely capturing Earth's surface phenomena and monitoring changes over time. However, despite their efficiency, working with VHR images presents challenges; as spatial resolution increases to capture intricate details, the reflectance variability of individual objects also rises. This heightened level of detail leads to an increase within-class variances, which hampers the successful application of traditional supervised classification methods, complicating the analysis of such high-resolution images. Consequently, numerous efforts and diverse methods have been introduced to streamline the handling of these images (Teng et al., 2023).

One significant limitation of many existing methods is their reliance on direct comparisons of pixel values, which fail to detect relationships between pixels (Zhang et al., 2020). While these methods generate image difference maps and apply thresholds to classify pixels as changed or unchanged, they do not account for the semantic relationships that may exist between them. Common arithmetic operations, such as image differencing (Singh, 1986), image rationing (Todd, 1977), and image regression (Jackson, 1983), are typically employed for image comparison. However, these approaches may struggle to effectively analyze the complexities introduced by the high spatial resolution of VHR images, highlighting the need for more advanced techniques that can capture the intricate changes in the data more effectively.

To address the limitations of pixel-based methods, Object-Based Change Detection (OBCD) methods are used, employing objects as processing units to enhance the completeness and accuracy of the final results. The object-based approach relies on objects rather than pixels, allowing effective use of spectral and spatial information to match the features of Earth's objects. In OBCD, images are divided into meaningful, homogeneous areas, and further processing is conducted on these areas. Consequently, relationships between pixels are established, leading to improved results. Since pixel-based change detection methods neglect spatial contextual information, OBCD analysis extracts features from segmented image-objects to identify changes in the state of objects. One advantage of this method is its effectiveness in addressing the challenge associated with VHR images. By smoothing small changes in the size of each object, the high spectral variability is reduced (Javed et al., 2020).

Using a post-classification OBCD approach, (Mitkari et al., 2018) analyzed glacier surface changes in the Indian Himalayas over a three-year period using high spatial resolution images from WorldView-2 and Linear Imaging Self-Scanning System (LISS) IV. By employing multi-resolution segmentation and spectral and spatial features such as mean value, standard deviation, brightness temperature, slope, and spectral ratios, they improved their overall accuracy to 91.41%.

Building on the advancements provided by OBCD, another approach to tackling the challenges of VHR images is to extract advanced features, such as textural and geometrical attributes, which enhance the method's ability to understand and detect complex changes (Luo et al., 2018). (Khurana & Saxena, 2017) utilized three sets of texture features—mean, variance, homogeneity, contrast, and dissimilarity—extracted from the Gray Level Co-occurrence Matrix (GLCM) method while investigating how these combinations affect the accuracy of change detection, keeping all other parameters constant. However, the sequential combination of OBCD and feature extraction

requires further investigation to fully understand their potential benefits and effectiveness.

The application of various methodologies is essential for effectively interpreting complex datasets. In this context, machine learning techniques, including both supervised and unsupervised methods, have been developed, many of which rely on image transformation algorithms to classify multi-temporal images. The performance of unsupervised methods can be negatively impacted by external factors such as changes in atmospheric conditions, variations in illumination, and inadequate sensor calibration that often occur during image acquisition at different dates. On the other hand, supervised methods are more robust in handling various atmospheric and illumination conditions that may vary across different acquisition times, providing a more reliable approach for image analysis and change detection (Pacifici et al., 2007).

(Amini et al., 2022) applied the Random Forest (RF) supervised classifier, using Landsat time-series data, to analyze land use and land cover changes in Isfahan from 1985 to 2019. Their proposed algorithm achieved an overall accuracy that was 10% higher than the Copernicus Global Land Cover Layers (CGLCL) map. (Wang et al., 2018) utilized relevant conceptual features and identified urban changes using a weighted voting group learning approach with various classifiers, such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Extreme Learning Machine (ELM), and Random Forest (RF). Their experiments conducted on Gaofen-1 and Ziyuan-3 High Resolution (HR) images demonstrated an overall accuracy of approximately 99%. (Kumar et al., 2022) presented an optimal integration of multi-sensor datasets, including ASTER, PALSAR, and Sentinel-1, using various input features such as spectral, morphological, and textural data to classify different rock types in Chhatarpur district, India, with Machine Learning methods. Their results indicated that the Support Vector Machine (SVM) achieved a better classification accuracy of 77.78%, which is around 15% higher than that obtained using ASTER spectral data alone.

In summary, while the combination of object-based methods and contextual information can enhance change detection in high-resolution remote sensing images; however, the results may still be influenced by several challenges. These include the selection of training samples, urban scene characteristics, the choice of classifier, and variations in illumination, shadows, or seasonal changes. Additionally, many existing models struggle with complex feature interactions and are sensitive to noise, which can lead to misclassifications in altered areas. Furthermore, traditional methods often do not fully leverage the potential of diverse feature extraction techniques, limiting their effectiveness in detecting subtle changes.

In this paper, we address these gaps by integrating object-based, feature-based, and dual learning-based

approaches on pairs of Very High-Resolution (VHR) bi-temporal images, which enhances our ability to capture complex feature interactions. Firstly, the Iterative Self-Organizing Data Analysis Technique (ISODATA) is applied to generate segmentation maps of the image pairs. Subsequently, small homogeneous areas are processed, and essential semantic features—including textural, mathematical, and geometrical features—are extracted from these areas, improving robustness against noise and variability caused by illumination and seasonal changes. The next step involves obtaining a difference image by calculating the difference vectors of the feature sets from the image pairs. The difference image is then input into two categories of supervised learning-based classification methods: 1) individual learning and 2) ensemble learning. Finally, change maps are generated for each supervised method, allowing for a thorough comparison alongside further comparative analysis.

Our proposed method detects existing changes with optimal accuracy, leveraging high processing power through the combination of object-based, feature-based and learning-based approaches. The primary objective of this research is to develop a change detection approach specifically tailored for very high-resolution remote sensing images, emphasizing the establishment of communication among various components within the images. Based on this, the advancements of this article are summarized in the following cases:

- *Sequential combination of three methods based on objects, features, and dual learning and investigating their effects.*
- *Implementation of semantic segmentation with an emphasis on pixel relationships.*
- *Extraction of a wide range of features for semantic analysis to address the complexities of the very high-resolution (VHR) images.*
- *Comprehensive implementation and comparison of individual and ensemble learning-based methods.*

The rest of the paper is organized as follows: Section 2 presents the proposed change detection methodology, including the preprocessing step, image segmentation, an explanation of the different types of features extracted, and an introduction to the supervised classifiers. Section 3 outlines the experimental setup, detailing the datasets used and presenting both visual and numerical results. In Section 4, we provide a discussion that includes further analysis and an ablation study. Finally, Section 5 concludes the paper, summarizing key findings and implications.

2. Methodology

In this section, we first briefly state the general steps of our proposed method, then present a detailed description of

each step and introduce the procedure depicted as a flowchart in Figure 1. The main steps are as follows:

1. *The preprocessing stage is carried out to facilitate accurate comparisons between bi-temporal images.*
2. *The Iterative Self-Organizing Data Analysis Technique (ISODATA) is utilized for image segmentation on VHR images to clarify objects with specific characteristics.*
3. *Textural, mathematical, and geometrical (related to shape) features are extracted from the segmented pair-images. The obtained features are fused with each other to create the Feature Fusion Layer (FFL).*
4. *The difference image is calculated using the vector difference of the two FFLs obtained from time 1 and time 2 images.*
5. *The difference image is analyzed using individual and ensemble supervision classification methods, resulting in the production of final change maps*

2.1. Image Preprocessing

Basic preprocessing techniques for change detection (CD) images are applied to align bi-temporal images acquired from the same or different sensors and acquisition dates. These techniques include geometric and radiometric corrections, denoising to reduce noise and preserve information (Afaq & Manocha, 2021), pan-sharpening to improve spatial resolution (Bovolo et al., 2010), and orthorectification to eliminate terrain distortion (Im et al., 2008), enhancing the overall quality and accuracy of subsequent analysis.
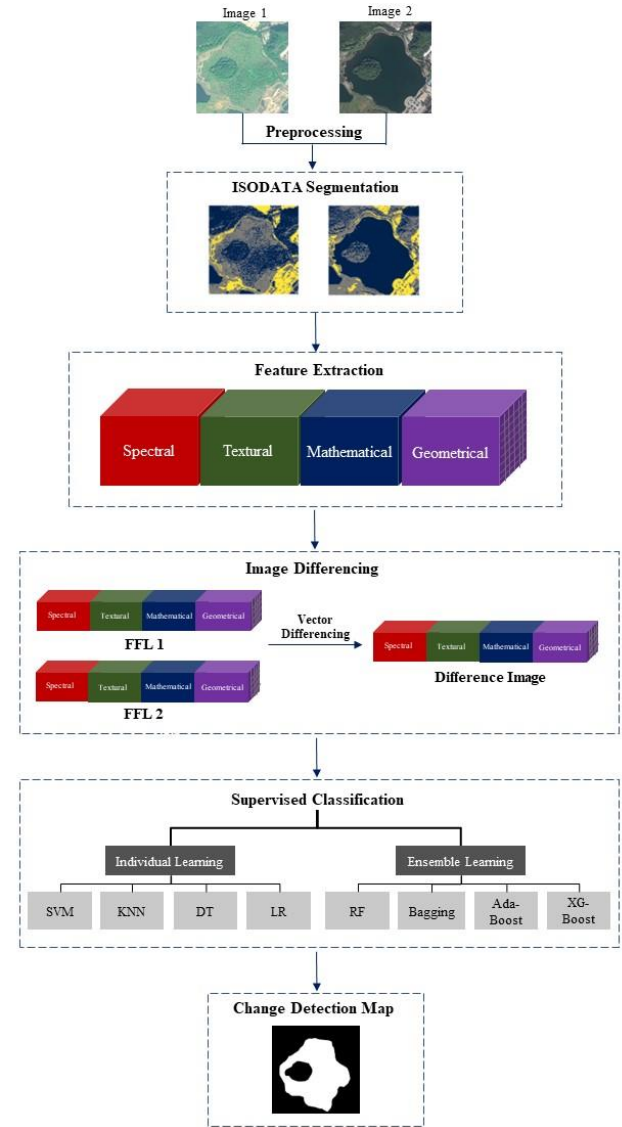

Figure 1. The flowchart of proposed method

2.2. Image Segmentation

Image segmentation divides an image into homogeneous areas based on physical and semantic characteristics, using segmentation algorithms (Ez-zahouani et al., 2023). The ISODATA algorithm, an unsupervised technique and a modification of k-means clustering, groups pixels into objects while addressing k-means limitations. It identifies cluster structures and refines cluster centers to minimize the sum of squared distances between data points and their nearest center. By utilizing an average response pattern to represent a group of patterns, the algorithm iteratively refines these patterns to enhance precision of the clustering

process (*Ball & Hall, 1967*). We use the ISODATA algorithm within the OBCD framework to segment bi-temporal VHR images, enabling detailed analysis and extraction of valuable information about changes and patterns over time in the identified objects.

2.3. Feature Extraction

Feature extraction is crucial for recognition, as advantageous features enhance intra-class aggregation and inter-class separation, improving the change detection process (*Gao et al., 2019*). In this step, three types of features— textural, mathematical, and geometrical—are derived from the segmented image to construct the FFL for bi-temporal images.

2.3.1. Textural Features

Textural features pertain to the visual patterns and arrangements of pixels in an image, capturing spatial distribution and relationships, revealing surface characteristics like texture and smoothness. One method for calculating texture is the Gray-Level Co-occurrence Matrix (GLCM), which provides a statistical representation of texture using co-occurring pixel values and includes features such as mean, standard deviation, contrast, and entropy (*Haralick et al., 1973*). Another method involves using a Gabor filter, which captures both spatial and frequency information through convolution and is defined by parameters such as orientation and wavelength (*Mehrotra et al., 1992*). Additionally, the structure tensor offers a statistical summary of intensity gradients in local neighborhoods (*D. Chen et al., 2021*).

2.3.2. Mathematical Features

Mathematical features involve techniques to analyze and extract information from images using mathematical computations. Common methods include the Gaussian filter, which smooths pixel values, as well as its derivatives: the Difference of Gaussians (DoG), which highlights intensity differences, and the Laplacian of Gaussian (LoG), which enhances features by highlighting rapid intensity changes. Morphological profiles, using opening and closing operations with structuring elements, are also employed to extract spatial features at various scales. Advanced opening and closing by reconstruction techniques improve information extraction and shape preservation (*Hu et al., 2018*). This research utilized 15 morphological operations, detailed in Table 1.

2.3.3. Geometrical Features

Geometrical features pertain to the shape characteristics of objects in an image, offering insights into their structure and edges. This study utilizes common shape features including the shape index and specific edge detectors, as detailed in Table 1, to extract valuable information about object details.

Table 1. List of features extracted using VHR Imagery

Mathematical		Textural		Geometric
Morphological Operation	erosion	*GLCM*	mean	shape index
	dilation		std	Canny
	opening		contrast	Roberts
	closing		Dissimilarity	Sobel
	morphological gradient		Homogeneity	Scharr
	hit or miss		ASM	Prewitt
	thin lines removed		energy	
	block average		max	
Gaussian		entropy		
DoG		Gabor filter		
LoG		structure tensor		

After extracting features from two image pairs, the values are normalized using Min-Max scaling to standardize them on a similar scale (0 to 1), facilitating consistent analysis during fusion. The normalized features are combined in a stacking layer to create a Feature Fusion Layer (FFL) for each image, integrating complementary information. Image differencing is then performed by calculating difference vectors from the two FFLs, which serve as input for classifiers. Mathematically, the difference vector (ΔD) is defined as (*Wang et al., 2018*):

$$\Delta D = T - S \tag{1}$$

where $T = (t_1, t_2, ..., t_n)$ and $S = (s_1, s_2, ..., s_n)$ represent two single FFL of images, and n is the depth of the FFL. Pixels with significant difference vector (ΔD) values are indicative of transitions in land cover and desired changes. Conversely, unchanged pixels are expected to have $\Delta D \approx 0$, while pixels showing substantial deviations from 0 in at least one feature are more likely to be linked to land cover change.

2.4. Supervised Classification

Supervised classification, a subset of Machine Learning (ML), uses labeled data to learn functions that predict outcomes for new data, making it foundational in the change detection (CD) process (*Thanh Noi & Kappas, 2018*). Change vector from FFLs is input for supervised classifiers to distinguish between changed and unchanged areas. This approach includes individual learning, where a single

model makes predictions, and ensemble learning, which combines predictions from multiple models (*Mienye & Sun, 2022*). This paper utilizes both types of classifiers, as detailed below.

2.4.1. Individual Learning

In this study, individual learning methods are used to train a single model on labeled data for accurate predictions, including:

1. *Support Vector Machine (SVM): A nonparametric classifier that separates data using an n-dimensional hyperplane, with Gaussian radial basis function as the kernel function (Thanh Noi & Kappas, 2018).*
2. *K-Nearest Neighbor (KNN): A supervised technique that classifies data by examining the k nearest neighbors and using majority voting, achieving effective classification with minimal training data (Uddin et al., 2022).*
3. *Decision Tree (DT): This classifier partitions training data hierarchically by splitting attribute values iteratively until leaf nodes contain a specified number of records for classification (Berhane et al., 2018).*
4. *Logistic Regression (LR): A statistical method for modeling the relationship between a categorical dependent variable and one or more independent variables, commonly used in binary classification (Mondal & Mandal, 2018).*

2.4.2. Ensemble Learning

Ensemble learning enhances classification performance by combining predictions from multiple models, categorized into parallel and sequential ensembles. Parallel methods, like Bagging and Random Forest (RF), train base classifiers independently and combine their predictions, promoting diversity. In contrast, sequential methods, like Boosting and Stacking, train models iteratively to correct errors from previous ones, improving overall performance (*Mienye & Sun, 2022*). The ensemble learning methods used in this research include:

1. *Random Forest (RF): This algorithm constructs multiple decision trees from bootstrapped samples and combines their predictions through majority voting, which reduces overfitting and improves classification accuracy (Mienye & Sun, 2022).*
2. *Bootstrap Aggregating (Bagging): Developed by (Breiman, 1996), Bagging improves classification performance by combining outputs from models trained on different bootstrapped samples, effectively reducing variance and computational time for large datasets.*
3. *Adaptive Boosting (AdaBoost): Developed by (Freund & Schapire, 1999), AdaBoost trains weak learners on weighted training data, adjusting weights based on classification errors to create a robust classifier. It is flexible, requiring minimal hyperparameter tuning.*
4. *Extreme Gradient Boosting (XGBoost): Introduced by (T. Chen & Guestrin, 2016), XGBoost is a scalable, accurate ensemble method based on gradient boosting, featuring regularization to reduce overfitting and effectively handle data normalization and missing values.*

The eight models (four from individual learning and four from ensemble learning) were trained independently and evaluated across two image pairs in two experiments, which will be described in the next section. A comparative analysis will follow to assess the effectiveness of the models in determining the final change detection outcome for each pixel.

3. Implementation Experiments and Results

3.1. Experiment 1

In Experiment 1, a pair of GaoFen-2 (GF-2) multispectral images is used. The size of these images is 512×512 and they are taken from Crop Land Change Detection (CLCD) dataset (*M. Liu et al., 2022*). In the CLCD dataset, the bi-temporal images were acquired within Guangdong Province, China (Figure 2), during 2017 for time 1 and 2019 for time 2, with a spatial resolution ranging from 0.5 to 2 meters. Each group of samples includes two before-and-after images and a binary label indicating changes in cropland. Time 1 and Time 2 images, along with the label image related to Experiment 1 are shown in Figure 3.

Figure 4 depicts the image segmentation results obtained using the ISODATA algorithm. When running this algorithm, several parameters need to be set, such as the Number of Clusters (K), Minimum Class Size, color weight, shape weight, smoothness, and compactness weights, which are set to 3, 20, 0.8, 0.2, 0.5, and 0.5, respectively. This segmentation process aims to evaluate the effectiveness of utilizing multiple features and the classification step in the change detection process.

Figure 2. CLCD dataset study area, with Guangdong province highlighted by the red circle.
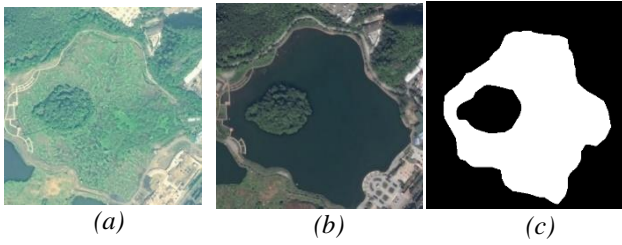

(a)   *(b)*   *(c)*

Figure 3. The VHR images acquired by GF-2 satellite; (a) image 1 acquired on 2017; (b) image 2 acquired on 2019; (c) reference change map.
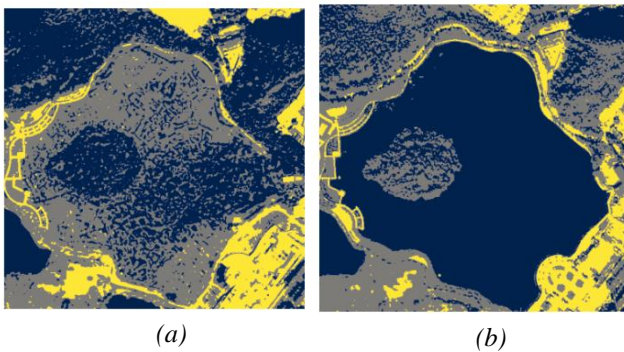

(a)   *(b)*

Figure 4. The segmented images by ISODATA algorithm; (a) segmented image of time 1; (b) segmented image of time 2.

A total of 45 features are extracted from the segmented images, including textural, mathematical, and geometrical features, with enhancements made to Gabor and morphological features using various kernels. These features are combined to create FFLs for each VHR image, leading to a difference image with dimensions of

$512 \times 512 \times 45$, which improves the capture of underlying patterns. Both individual and ensemble classifiers are employed, with a test ratio of 0.4, allowing effective use of the available labeled reference data for training while ensuring sufficient data for reliable evaluation of model performance. Notable settings include a maximum iteration of 100 for SVM, 50 neighbors for KNN, 100 estimators for RF, and a learning rate of 0.01 for XGBoost. Change maps resulting from these classifiers are illustrated in Figure 5.

Table 2 presents the numerical results of the eight supervised classifiers, indicating various metrics achieved by each. The highest accuracies for each metric are highlighted in bold. Mean accuracies for individual and ensemble learners are calculated separately, showing that ensemble methods generally outperform individual methods. Among ensemble methods, the Random Forest (RF) method shows the best performance, while the Decision Tree (DT) method leads among individual methods. These results provide a valuable assessment of different models' efficacy in change detection and facilitate a consistent comparison between individual and ensemble learning approaches.

(SVM)   *(KNN)*   *(DT)*

(LR)   *(RF)*   *(Bagging)*
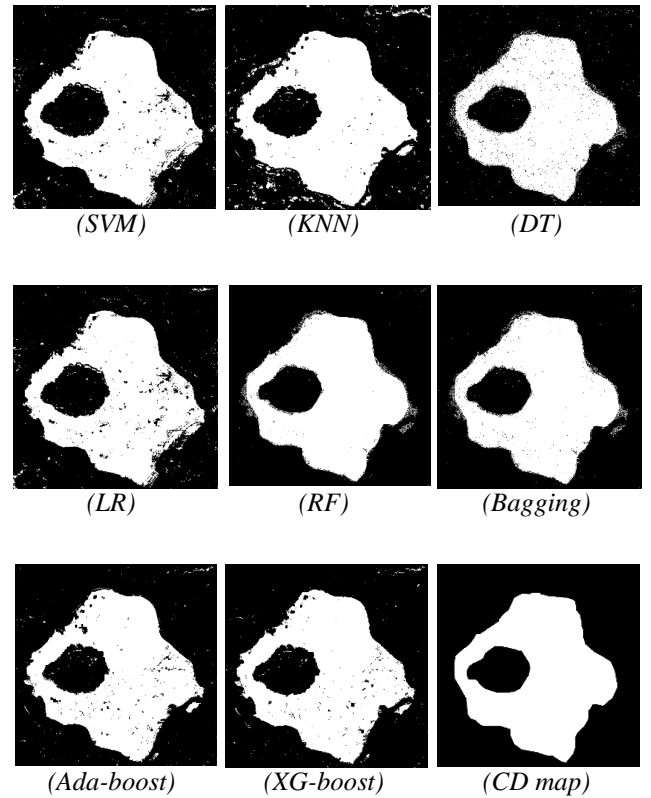
(Ada-boost)   *(XG-boost)*   *(CD map)*

Figure 5. Binary change maps obtained from Individual classifiers, Ensemble classifiers, and Reference change detection map (CD map).

Table 2. Evaluation of the performance of Individual and Ensemble learners through some evaluation metrics in the dataset of Experiment 1.

Model	Precision (%)	Recall (%)	F1_ score (%)	Specificity (%)	IoU (%)
SVM	93.95	96.55	95.23	90.22	85.58
KNN	92.73	98.18	95.38	88.78	86.48
DT	**98.53**	**98.49**	**98.51**	**97.52**	**95.10**
LR	94.23	96.07	95.14	90.56	85.19
Mean	94.86	97.32	96.07	91.77	88.09

Model	Precision (%)	Recall (%)	F1_ score (%)	Specificity (%)	IoU (%)
RF	98.73	**99.31**	**99.02**	97.88	**96.78**
Bagging	**98.85**	98.98	98.91	**98.06**	96.41
Ada-boost	95.48	97.48	96.47	92.63	89.05
XG-boost	95.92	97.64	96.77	93.31	89.90
Mean	**97.25**	**98.35**	**97.80**	**95.47**	**93.03**

(SVM)   *(KNN)*   *(DT)*

(LR)   *(RF)*   *(Bagging)*

(Ada-boost)   *(XG-boost)*   *(CD map)*

Figure 8. Binary change maps obtained from Individual classifiers, Ensemble classifiers, and Reference change detection map (CD map).

3.2. Experiment 2

Experiment 2 used a pair of GaoFen-2 (GF-2) multispectral images from the Crop Land Change Detection (CLCD) dataset, depicting Guangdong Province, China, in 2017 and 2019. The images, with a size of 512×512 and a spatial resolution of 0.5 to 2 m, are shown in Figure 6. Image segmentation results from the ISODATA algorithm are illustrated in Figure 7, using parameters identical to Experiment 1: K=3, Minimum Class Size=20, color weight=0.8, shape weight=0.2, smoothness weight=0.5, and compactness weight=0.5. This method identifies homogeneous areas utilizing spatial information, enhancing change detection accuracy.
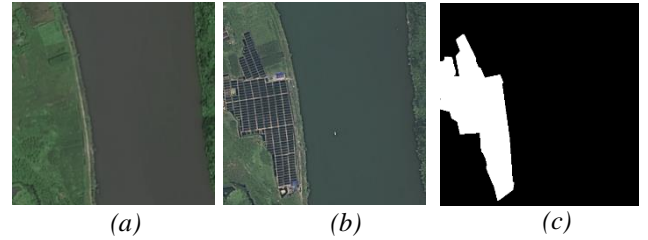

(a)   *(b)*   *(c)*

Figure 6. The VHR images acquired by GF-2 satellite; (a) image 1 acquired on 2017; (b) image 2 acquired on 2019; (c) reference change map.
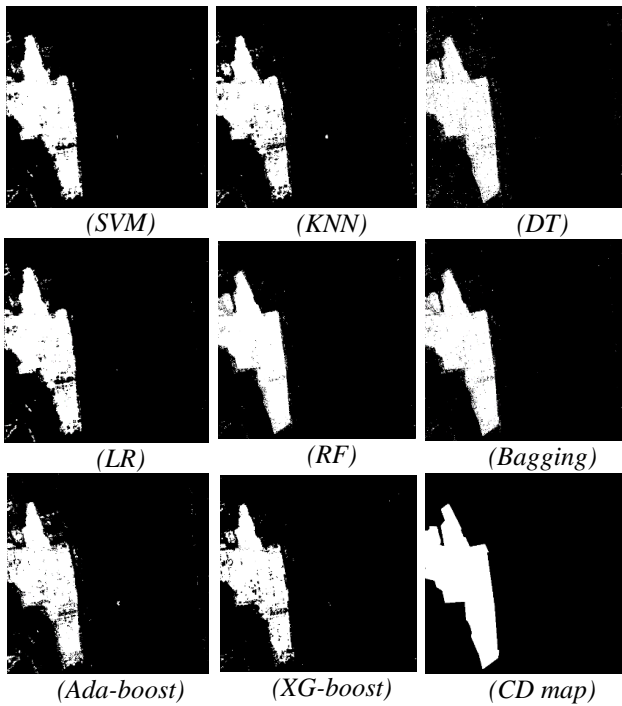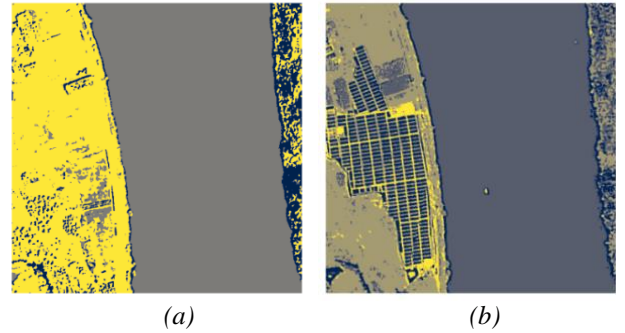

(a)   *(b)*

Figure 7. The segmented images by ISODATA algorithm; (a) segmented image of time 1; (b) segmented image of time 2.

A total of 45 textural, mathematical, and geometrical features are extracted from the segmented images, with specific features enhanced using various kernels. These features are combined to create FFLs, and vector contrasts yield a difference image of dimensions $512 \times 512 \times 45$. Similar to Experiment 1, both individual and ensemble classifiers are employed with a test ratio of 0.4. The configurations include a maximum iteration of 100 for SVM, 50 neighbors for KNN, 100 estimators for RF, and a learning rate of 0.01 for XGBoost. The classifiers are applied to the difference image, yielding change maps shown in Figure 8.

Table 3. Evaluation of the performance of Individual and Ensemble learners through some evaluation metrics in the dataset of Experiment 2.

Model	Precision (%)	Recall (%)	F1_score (%)	Specificity (%)	IoU (%)
SVM	98.46	97.94	98.20	90.57	80.34
KNN	98.75	97.72	98.23	92.10	80.36
DT	**99.32**	**99.32**	**99.32**	**95.96**	**92.24**
LR	98.48	97.79	98.14	90.59	79.63
Mean	98.75	98.19	98.47	92.30	83.14

Model	Precision (%)	Recall (%)	F1_score (%)	Specificity (%)	IoU (%)
RF	99.58	**99.57**	**99.58**	97.53	**95.11**
Bagging	**99.59**	99.36	99.48	**97.56**	93.94
Ada-boost	98.39	97.89	98.14	90.17	79.81
XG-boost	99.23	97.38	98.29	94.82	80.46
Mean	**99.20**	**98.55**	**98.87**	**95.02**	**87.33**

Table 3 shows the quantitative outcomes of the eight classifiers, highlighting the highest accuracies in bold. Mean accuracies for individual and ensemble learners reveal that ensemble methods generally achieve better results, with RF achieving the best performance among ensemble classifiers and DT leading among individual classifiers.

4. Discussion

While effective, VHR images increase variability in object reflection due to their higher spatial resolution. This research utilized VHR images to identify both general and specific changes in Earth's surface phenomena through a comprehensive approach that included object detection, feature extraction, and both individual and ensemble learning techniques. A total of 45 features categorized as textural, mathematical, and geometrical were extracted. The Random Forest (RF) classifier was employed to calculate feature importance scores based on Gini impurity reductions (*Disha & Waheed, 2022*), allowing for ranking according to their impact on change detection. The structure tensor (Ayy) ranked highest, followed by the GLCM (entropy) feature, while the spectral bands ranked 27th, 29th, and 31st, indicating their lesser importance compared to other features.

The composition of feature categories was analyzed in Figure 9, showing that mathematical and textural features each contributed 40%, while geometrical features and spectral bands contributed 13% and 7%, respectively. In

terms of average of scores, the textural category outperformed others with an average score of 39.6%, followed by mathematical (34.3%), spectral bands (14.9%), and geometrical (11.2%) categories, highlighting the significance of these features in bi-temporal image classification over raw image data.
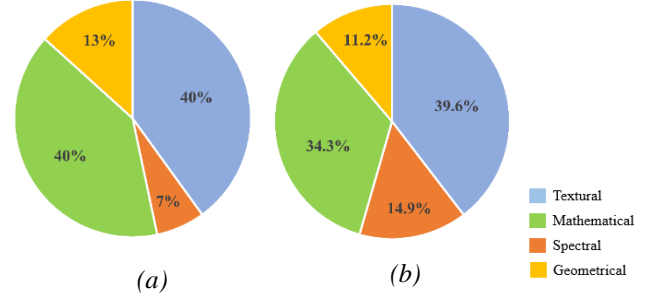
Figure 9. Pie charts illustrating the participation percentage (a) and mean score percentage (b) of feature categories based on their level of participation and influence.

The study employed individual and ensemble classifiers, with RF and Bagging achieving the highest performance, followed by Decision Tree (DT) classifier, and XGBoost and AdaBoost in subsequent ranks. Ensemble learners demonstrated superior performance compared to individual learners, supported by comparative results in Tables 2 and 3. Higher accuracies were found in Experiment 1 due to more distinct objects; however, a similar trend across experiments indicates the proposed method's generalizability in classifier performance, as shown in Figure 10.
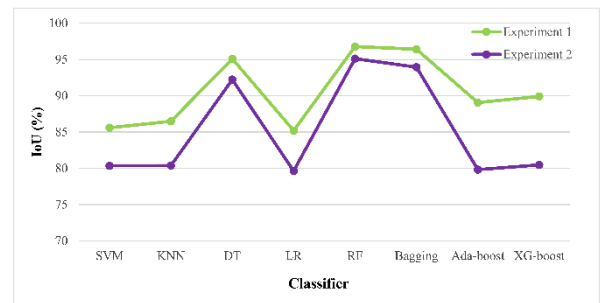
Figure 10. Chart of performance of learners in experiment 1 and 2.

To evaluate the effectiveness of the proposed approach compared to other methods used on similar dataset, Table 4 is presented, which compares the results obtained from (*Soheili et al., 2023*) on the similar CLCD dataset using the PCA-KMeans method with those obtained from our method. As shown, our proposed approach outperformed the PCA-

KMeans method, achieving an increase of 16% in Recall and 8% in F1-score.

Table 4. Comparison of proposed approach and PCA-KMeans method.

Approach	Recall (%)	F1_score (%)
PCA-KMeans	83.00	91.00
Proposed approach	99.57	99.58

To further demonstrate the efficiency of the segmentation stage in change detection, the proposed model was executed once without the segmentation stage, relying solely on the initial image (pixel-based), and the results are presented in Table 5. This table presents the average evaluation metrics for individual and ensemble learners, comparing the pixel-based method with the object-based method in both experiments. The analysis revealed that the object-based method consistently outperformed the pixel-based approach. In Experiment 1, the object-based method achieved an IoU of 88.09% for individual learners and 93.03% for ensemble learners, representing enhancements of 3% and 4%, respectively, compared to pixel-based methods. In Experiment 2, although there was a slight decline in recall for individual learners, the ensemble learners using object-based methods outperformed the pixel-based ones in all metrics. Overall, the maximum improvement observed with the object-based method exceeded 4%, highlighting its effectiveness in enhancing object detection tasks compared to the pixel-based method, while still demonstrating the significant effectiveness of ensemble learning.

To examine the impact of different feature sets on the final results, an ablation study was conducted on the first dataset, across various feature categories. In this study, mathematical, textural, and geometrical feature types were utilized separately in the models, and average results for both individual and ensemble learners were recorded in Table 6.

Table 5. Comparison of proposed object-based and pixel-based methods.

Experiment 1						
Model	mode	Precision (%)	Recall (%)	F1_score (%)	Specificity (%)	IoU (%)
Individual learners	Pixel-Based	93.62	96.77	95.16	89.84	85.47
	Object-Based	**94.86**	**97.32**	**96.07**	**91.77**	**88.09**
Ensemble learners	Pixel-Based	95.59	97.22	96.39	92.82	88.88
	Object-Based	**97.25**	**98.35**	**97.80**	**95.47**	**93.03**

Experiment 2						
Model	mode	Precision (%)	Recall (%)	F1_score (%)	Specificity (%)	IoU (%)
Individual learners	Pixel-Based	97.88	**98.62**	98.25	87.67	81.67
	Object-Based	**98.75**	98.19	**98.47**	**92.30**	**83.14**
Ensemble learners	Pixel-Based	98.93	98.42	98.68	93.43	85.34
	Object-Based	**99.20**	**98.55**	**98.87**	**95.02**	**87.33**

For mathematical features, ensemble learners achieved an F1 score of 94.99%, indicating a strong balance between precision and recall, while the Intersection over Union (IoU) was 84.35%. This suggests that mathematical features effectively contribute to model performance, highlighting their ability to represent complex relationships within the data. In contrast, textural features yielded an even higher F1 score of 96.34% for ensemble learners, along with an IoU of 90.38%. This exceptional performance demonstrates that textural features are particularly beneficial for applications requiring detailed recognition of patterns and textures. When examining geometrical features, the metrics show a different trend. Ensemble learners recorded a precision of 98.64%, but the F1 score was only 63.24%. These unstable metrics suggest that geometrical features alone lack the capability to effectively detect relevant patterns, primarily due to their small number of only six features. The introduction of geometrical features aimed mainly at improving edge detection of objects, while the primary responsibilities for accurate identification rest with the other feature types.

When considering all features collectively, ensemble learners achieved a commendable F1 score of 95.47% and an IoU of 93.03%. This suggests that combining multiple feature types enhances performance by offsetting potential weaknesses in any single feature type.

Table 6 Ablation study results showing the performance of different types of features in individual and ensemble learning models.

Feature type	Model	Precision (%)	Recall (%)	F1_score (%)	Specificity (%)	IoU (%)
Mathematical	Individual learners	90.89	94.51	92.63	85.86	79.21
	Ensemble learners	95.15	94.84	94.99	91.77	84.35
Textural	Individual learners	93.47	96.53	94.97	89.65	85.19
	Ensemble learners	95.81	97.91	96.84	93.30	90.38
Geometrical	Individual learners	**99.27**	63.59	77.51	55.59	22.26
	Ensemble learners	**98.64**	63.29	77.10	63.24	33.04
All features (proposed)	Individual learners	94.86	**97.32**	**96.07**	**91.77**	**88.09**
	Ensemble learners	97.25	**98.35**	**97.80**	**95.47**	**93.03**

5. Conclusions

This study presents a practical change detection scheme that leverages object detection through the segmentation of very high-resolution (VHR) images, complemented by feature extraction and various supervised learning techniques. Object-based segmentation enhances image homogeneity by preserving pixel relationships and reducing noise sensitivity, significantly outperforming traditional pixel-based methods. The introduction of textural features proves particularly effective, while ensemble classifiers demonstrate superior performance over individual models, underscoring their ability to generate accurate and reliable change maps even with minimal prior knowledge of the study area.

The contributions of this work lie in developing a comprehensive change detection framework that combines advanced segmentation techniques with the extraction of a diverse range of textural, mathematical, and geometrical features. This framework effectively mitigates issues associated with noise sensitivity and within-class variability inherent in traditional methods. It also reduces the need for deep learning methods, which require extensive data, training time, and costs.

Despite these advancements, some limitations remain, including the reliance on supervised learning and manual tuning of segmentation parameters. Future research should prioritize unsupervised or semi-supervised methods, automate segmentation processes, incorporate temporal data, and improve parameter optimization and feature weighting to enhance the framework's scalability and applicability across various remote sensing challenges.

References

Afaq, Y., & Manocha, A. (2021). Analysis on change detection techniques for remote sensing applications: A review. *Ecological Informatics, 63*, 101310. *https://doi.org/10.1016/j.ecoinf.2021.101310*

Amini, S., Saber, M., Rabiei-Dastjerdi, H., & Homayouni, S. (2022). Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series. *Remote Sensing, 14*(11). *https://doi.org/10.3390/rs14112654*

Ball, G. H., & Hall, D. J. (1967). A clustering technique for summarizing multivariate data. *Behavioral Science, 12*(2), 153–155.

Berhane, T. M., Lane, C. R., Wu, Q., Autrey, B. C., Anenkhonov, O. A., Chepinoga, V. V., & Liu, H. (2018). Decision-Tree, Rule-Based, and Random Forest Classification of High-Resolution Multispectral Imagery for Wetland Mapping and Inventory. *Remote Sensing, 10*(4), Article 4. *https://doi.org/10.3390/rs10040580*

Bovolo, F., Bruzzone, L., Capobianco, L., Garzelli, A., Marchesi, S., & Nencini, F. (2010). Analysis of the Effects of Pansharpening in Change Detection on VHR Images. *IEEE Geoscience and Remote Sensing Letters, 7*(1), 53–57. IEEE Geoscience and Remote Sensing Letters. *https://doi.org/10.1109/LGRS.2009.2029248*

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140. *https://doi.org/10.1007/BF00058655*

Chen, D., Li, J., Di, S., Peethambaran, J., Xiang, G., Wan, L., & Li, X. (2021). Critical Points Extraction from Building Façades by Analyzing Gradient Structure Tensor. *Remote Sensing, 13*(16), Article 16. *https://doi.org/10.3390/rs13163146*

Chen, H., Qi, Z., & Shi, Z. (2022). Remote Sensing Image Change Detection With Transformers. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–14. IEEE Transactions on Geoscience and Remote Sensing. *https://doi.org/10.1109/TGRS.2021.3095166*

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. *https://doi.org/10.1145/2939672.2939785*

Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. *Cybersecurity, 5*(1), 1. *https://doi.org/10.1186/s42400-021-00103-8*

Ez-zahouani, B., Teodoro, A., El Kharki, O., Jianhua, L., Kotaridis, I., Yuan, X., & Ma, L. (2023). Remote sensing imagery segmentation in object-based analysis: A review of methods, optimization, and quality evaluation over the

past 20 years. Remote Sensing Applications: Society and Environment, 32, 101031. https://doi.org/10.1016/j.rsase.2023.101031

Freund, Y., & Schapire, R. (1999). A Short Introduction to Boosting. https://www.semanticscholar.org/paper/A-Short-Introduction-to-Boosting-Freund-Schapire/c834bddd5e75a64ca9bb80c195cf84345c38bb9b

Gao, L., Zhang, X., Gao, J., & You, S. (2019). Fusion Image Based Radar Signal Feature Extraction and Modulation Recognition. IEEE Access, 7, 13135–13148. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2892526

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6), 610–621. IEEE Transactions on Systems, Man, and Cybernetics. https://doi.org/10.1109/TSMC.1973.4309314

Hu, J., Ghamisi, P., & Zhu, X. X. (2018). Feature Extraction and Selection of Sentinel-1 Dual-Pol Data for Global-Scale Local Climate Zone Classification. ISPRS International Journal of Geo-Information, 7(9), Article 9. https://doi.org/10.3390/ijgi7090379

Im, J., Jensen, J. R., & Tullis, J. A. (2008). Object- based change detection using correlation image analysis and image segmentation. International Journal of Remote Sensing, 29(2), 399–423. https://doi.org/10.1080/01431160601075582

Jackson, R. D. (1983). Spectral indices in n-space. Remote Sensing of Environment, 13(5), 409–421.

Javed, A., Jung, S., Lee, W. H., & Han, Y. (2020). Object-Based Building Change Detection by Fusing Pixel-Level Change Detection Results Generated from Morphological Building Index. Remote Sensing, 12(18), Article 18. https://doi.org/10.3390/rs12182952

Khurana, M., & Saxena, V. (2017). Exploring the effectiveness of various texture features for change detection in remote sensing images. 2017 International Conference on Computer, Communications and Electronics (Comptelix), 96–100. https://doi.org/10.1109/COMPTELIX.2017.8003945

Kumar, C., Chatterjee, S., Oommen, T., Guha, A., & Mukherjee, A. (2022). Multi-sensor datasets-based optimal integration of spectral, textural, and morphological characteristics of rocks for lithological classification using machine learning models. Geocarto International, 37(20), 6004–6032. https://doi.org/10.1080/10106049.2021.1920632

L. Khelifi & M. Mignotte. (2020). Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis. IEEE Access, 8, 126385–126400. https://doi.org/10.1109/ACCESS.2020.3008036

Luo, H., Liu, C., Wu, C., & Guo, X. (2018). Urban Change Detection Based on Dempster–Shafer Theory for Multitemporal Very High-Resolution Imagery. Remote Sensing, 10(7). https://doi.org/10.3390/rs10070980

M. Liu, Z. Chai, H. Deng, & R. Liu. (2022). A CNN-Transformer Network With Multiscale Context Aggregation for Fine-Grained Cropland Change Detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 4297–4306. https://doi.org/10.1109/JSTARS.2022.3177235

Mehrotra, R., Namuduri, K. R., & Ranganathan, N. (1992). Gabor filter-based edge detection. Pattern Recognition, 25(12), 1479–1494. https://doi.org/10.1016/0031-3203(92)90121-X

Mienye, I. D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access, 10, 99129–99149. IEEE Access. https://doi.org/10.1109/ACCESS.2022.3207287

Mitkari, K. V., Arora, M. K., & Tiwari, R. K. (2018). Detecting Glacier Surface Changes Using Object-Based Change Detection. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 5180–5183. https://doi.org/10.1109/IGARSS.2018.8519230

Mondal, S., & Mandal, S. (2018). RS & GIS-based landslide susceptibility mapping of the Balason River basin, Darjeeling Himalaya, using logistic regression (LR) model. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 12(1), 29–44. https://doi.org/10.1080/17499518.2017.1347949

Pacifici, F., Del Frate, F., Solimini, C., & Emery, W. J. (2007). An Innovative Neural-Net Method to Detect Temporal Changes in High-Resolution Optical Satellite Imagery. IEEE Transactions on Geoscience and Remote Sensing, 45(9), 2940–2952. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2007.902824

Parelius, E. J. (2023). A Review of Deep-Learning Methods for Change Detection in Multispectral Remote Sensing Images. Remote Sensing, 15(8), Article 8. https://doi.org/10.3390/rs15082092

Shafique, A., Cao, G., Khan, Z., Asad, M., & Aslam, M. (2022). Deep Learning-Based Change Detection in Remote Sensing Images: A Review. Remote Sensing, 14(4), Article 4. https://doi.org/10.3390/rs14040871

Singh, A. (1986). Change detection in the tropical forest environment of northeastern India using Landsat. Remote Sensing and Tropical Land Management, 44, 273–254.

Soheili, N., Hasanlou, M., & Gomroki, M. (2023). Unsupervised building change detection in VHR images using improved PCA-KMeans algorithm. Earth Observation and Geomatics Engineering, 7(1). https://doi.org/10.22059/eoge.2023.365406.1142

Teng, Y., Liu, S., Sun, W., Yang, H., Wang, B., & Jia, J. (2023). A VHR Bi-Temporal Remote-Sensing Image Change Detection Network Based on Swin Transformer. Remote Sensing, 15(10). https://doi.org/10.3390/rs15102645

Thanh Noi, P., & Kappas, M. (2018). Comparison of Random Forest, k-Nearest Neighbor, and Support

Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors, 18(1), Article 1. https://doi.org/10.3390/s18010018

Todd, W. J. (1977). Urban and regional land use change detected by using Landsat data. J. Res. US Geol. Surv, 5(5), 529–534.

Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Scientific Reports, 12(1), 6256. https://doi.org/10.1038/s41598-022-10358-x

Wang, X., Liu, S., Du, P., Liang, H., Xia, J., & Li, Y. (2018). Object-Based Change Detection in Urban Areas from High Spatial Resolution Images Based on Multiple Features and Ensemble Learning. Remote Sensing, 10(2), Article 2. https://doi.org/10.3390/rs10020276

Z. Lv, T. Liu, J. A. Benediktsson, & N. Falco. (2022). Land Cover Change Detection Techniques: Very-high-resolution optical images: A review. IEEE Geoscience and Remote Sensing Magazine, 10(1), 44–63. https://doi.org/10.1109/MGRS.2021.3088865

Zerrouki, Y., Harrou, F., Zerrouki, N., Dairi, A., & Sun, Y. (2021). Desertification Detection Using an Improved Variational Autoencoder-Based Approach Through ETM-Landsat Satellite Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 202–213. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. https://doi.org/10.1109/JSTARS.2020.3042760

Zhang, C., Yue, P., Tapete, D., Jiang, L., Shangguan, B., Huang, L., & Liu, G. (2020). A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 166, 183–200. https://doi.org/10.1016/j.isprsjprs.2020.06.003