Hyperspectral change detection based on modification of UNet neural networks

Marwa S. Moustafa,* Sayed A. Mohamed, Sayed Ahmed, and Ayman H. Nasr
National Authority for Remote Sensing and Space Sciences, Cairo, Egypt

Abstract. The Earth's surface changes continuously due to several natural and humanmade factors. Efficient change detection (CD) is useful in monitoring and managing different situations. The recent rise in launched hyperspectral platforms provides a diversity of spectrum in addition to the spatial resolution required to meet recent civil applications requirements. Traditional multispectral CD algorithms hardly cope with the complex nature of hyperspectral images and their high dimensionality. To overcome these limitations, a CD deep convolutional neural network (CNN) semantic segmentation-based workflow was proposed. The proposed workflow is composed of four main stages, namely preprocessing, training, testing, and evaluation. Initially, preprocessing is performed to overcome hyperspectral image noise and the high dimensionality problem. Random oversampling (ROS), deep learning, and bagging ensemble were incorporated to handle imbalanced dataset. Also, we evaluated the generality and performance of the original UNet model and four variants of UNet, namely residual UNet, residual recurrent UNet, attention UNet, and attention residual recurrent UNet. Three hyperspectral CD datasets were employed in performance assessment for binary and multiclass change cases; all datasets suffer from class imbalance and small region of interest size. Recurrent residual UNet presented the best performance in both accuracy and inference time. Overall, the obtained results imply that deep CNN segmentation models can be utilized to implement efficient CD for hyperspectral imageries. [DOI: 10.1117/1.JRS.15.028505]

Keywords: hyperspectral imagery; semantic segmentation; change detection; deep learning.

Paper 210067 received Jan. 31, 2021; accepted for publication May 27, 2021; published online Jun. 17, 2021.

1 Introduction

Change detection (CD)[1] is an active remote sensing (RS) topic that has been adopted to monitor and understand the Earth's surface and forest areas. Very high spatial resolution imagery has been combined with modern machine learning approaches to improve the quality of CD maps. Hyperspectral image[2] contains hundreds of narrow bands that provide spectral and spatial information. Recently, HSI had been extensively used in classification and object detection tasks. Traditional CD methods such as linear transformations, classification, and abnormality analysis were proposed originally for single or multispectral imageries. However, their performance is limited when applied to HSIs due to their high dimensionality. Recently, several attempts had been introduced; these include tensor factorization,[3] orthogonal subspace mapping, multisource target feature support,[4] mixed pixel decomposition,[5] and independent component analysis.[6]

In the literature,[2,7,8] CD is a composite workflow that contains a series of comprehensive processing steps: (1) problem understanding, (2) collection of appropriate data, (3) preprocessing, (4) relevant features selection, (5) design and implementation of CD algorithm, and (6) evaluation of CD performance. The quality of the obtained change map depends on five main factors: (1) quality of CD algorithm, (2) spatial resolution, (3) temporal scale, (4) image registration (preprocessing), and (5) spectral correction. CD methods can be classified according to the number of change classes (binary, multiclass, and time series), CD algorithms (supervised,

*Address all correspondence to Marwa S. Moustafa, marwa.gis@gmail.com

semisupervised, and unsupervised), and automation (manual, semiautomated, and fully automated).

Typically, CD algorithms[9] are classified into four categories: (1) image algebra, (2) classification CD, (3) feature-based CD, and (4) machine learning-based CD. The algebra CD-based methods, such as change vector analysis, employ image difference and ratio image rules to provide robust and efficient performance. In classification CD, each image is independently classified, and then the change map is identified. Numerous classification approaches have been investigated to enhance CD accuracy. In feature learning and transformation, the learned features and distance metric are employed to distinguish changes. The features could be physically meaningful and engineered change features. Physically meaningful features are often elicited to define modifications in ground-truth types. Examples include vegetation indices, forest canopy variables, and water indices. In engineered features, the features are projected mathematically between different spaces to detect and highlight the change region. Examples include principal component analysis,[10] multivariate alteration detection,[11] subspace learning, and sparse learning. Finally, various supervised machine learning techniques had been adopted to identify land cover changes.[12] However, the limited availability of labeled datasets favors the utilization of unsupervised learning methods such as fuzzy and C-means algorithms.[13] In contrast, supervised learning methods such as support vector machines inhibit a better performance as they associated prior knowledge obtained from labeled datasets.[14]

Hyperspectral ad-hoc CD algorithms face different challenges: the availability of insufficient ground truth, data redundancy, noise existence in mixed pixels, coarse spatial resolution, and high dimensionality. In general, the limited performance of the traditional hyperspectral methods can be summarized as follows: (1) the transformation of temporal, spatial, and spectral information associated with satellite images by features engineering may cause a partial loss of data. (2) The majority of recent CD approaches depend on shallow models that lack the potential to generalize. (3) The availability of practical approaches in dimensionality reduction is limited. (4) Obtaining adequate hyperspectral labeled samples is difficult.[2]

The proliferation of sophisticated deep learning (DL) has evolved in the digital era. The availability of satellite instruments, the enormous amount of data acquired, and the availability of computational power has enabled a deeper neural network to introduce a new challenges in the earth science domain.[15,16] Recent advances in DL have demonstrated state-of-the-art results in pattern recognition tasks, mainly in image processing and speech recognition.[17,18] Modern convolutional neural network (CNN) architectures[19–21] tend to contain enormous hidden layers and millions of neurons, allowing them to concurrently learn hierarchical features for a broad class of patterns from data and achieve well-tailored models for the targeted application.[22] Recently, there has been a rapid turnover of DL frameworks to highlight land cover changes. Patch-based algorithms train temporal image patches to determine if the focal pixel is changed or not. In contrast, image-based algorithms have been utilized for training image pairs to generate a segmented change.[23] In Ref. 24, a recurrent neural network was adopted to produce the change map. The model network was fed a flattened and concatenated vector. Also, Siamese CNNs were adopted to obtain a discriminative feature map for each image. Then, the Euclidean distance metric was employed in determining the change map. These networks require a high degree of computational complexity. CD methods based on encoder–decoder segmentation techniques[25–27] were used to highlight the temporal changes in land cover. In recent years, different semantic segmentation was introduced based on CNN architectures. In modern CNN segmentation architectures, feature extraction is performed using downsampling. Deconvolutional upsampling layers were utilized to reconstruct per-pixel classification labels. A deconvolution operation is the transpose of a convolution operation and works by exchanging the forward and backward convolutional passes.[28]

Class imbalance,[29,30] which is widely observed in satellite images, hardens the identification of the minority class as the skewed distribution introduces a bias in favor of the majority class. The approaches handling class imbalance are categorized into data level and algorithm level.[29] Data level methods include data sampling [random oversampling (ROS) and random undersampling] and feature selection approaches. On the other hand, algorithm level methods include cost-sensitive and hybrid/ensemble approaches.[30] The ROS approach yields better classification performance compared with other data level approaches.

In general, the demand for a cost-effective and reliable hyperspectral CD (HSICD) approach is still a major open question. The complexity of hyperspectral imageries as well as the imbalanced class problem are considered the main factors of degraded performance. Therefore, we present an efficient workflow for HSICD (HSICD_workflow) to tackle binary and multi-HSICD problems. The proposed workflow comprises four main processing phases, namely preprocessing, training, testing, and evaluation. Also, we investigate the generality and performance of the original UNet model and its four variations: residual UNet (R-UNet), residual recurrent UNet (R2-UNet), attention UNet (Att-UNet), and attention residual recurrent UNet (Att-R2-UNet) to improve the HSICD performance. The major contributions are outlined in three steps:

- We formulate the class imbalance HSICD problem to incorporate ROS in preprocessing, DL, and bagging ensemble to handle the imbalanced dataset.
- We investigate three UNet loss functions to highlight the most robust loss function for the imbalanced dataset problem.
- We conduct extensive experiments to determine the performance of the proposed workflow. The proposed workflow significantly excels and contributes to future research regarding HSI change identification.

The remainder of this paper is organized as follows: Section 2 introduces the benchmark datasets and describes the proposed HSICD workflow in depth. In Sec. 3, the performance of each architecture is presented, compared, and discussed. Finally, Sec. 4 concludes the paper.

2 Materials and Methods

2.1 *Hyperspectral Dataset*

The limited availability of benchmarks datasets for the HSICD task is considered a major limitation to the RS community. In this work, we consider three binary HSICD datasets, namely the Bay Area, Santa Barbara,[31] and multiclass Hermiston datasets,[26] as shown in Table 1. The availability of pixel-based annotated masks for each dataset enables analytical evaluation for their experimental results.

2.1.1 *Bay Area dataset*

This dataset consists of two coregistered hyperspectral images over the city of Patterson, California, of a section with (600 × 500) pixels captured by the AVIRIS sensor. Each image contains 224 spectral bands with a spatial resolution of about 30 m per pixel. The images were acquired in 2007 and 2015, respectively. The bitemporal images, as well as the ground truth, are shown in Fig. 1(a).

2.1.2 *Santa Barbara dataset*

This dataset consists of two coregistered hyperspectral images over the Santa Barbara region, California, of a section with (984 × 740) pixels collected by the AVIRIS sensor. Each image

Table 1 Specification and data distribution of HSICD benchmarks.

Dataset	#Bands	Bad bands	Rows	Columns	Class distribution (%)	Spatial resolution (m)
Bay Area	224	[98:107], [113 :128], [148: 154], [167 :170]	600	500	C1:12.81; C2: 11.40	30
Santa Barbara	224		984	740	C1: 7.16; C2: 12.81	
Hermiston	242	[1:7], [58:76], [120:132], [165:182], [221:224]	390	200	C1:55.66; C2:13.33; C3:0.79; C4:15.59; C5:14.63	30

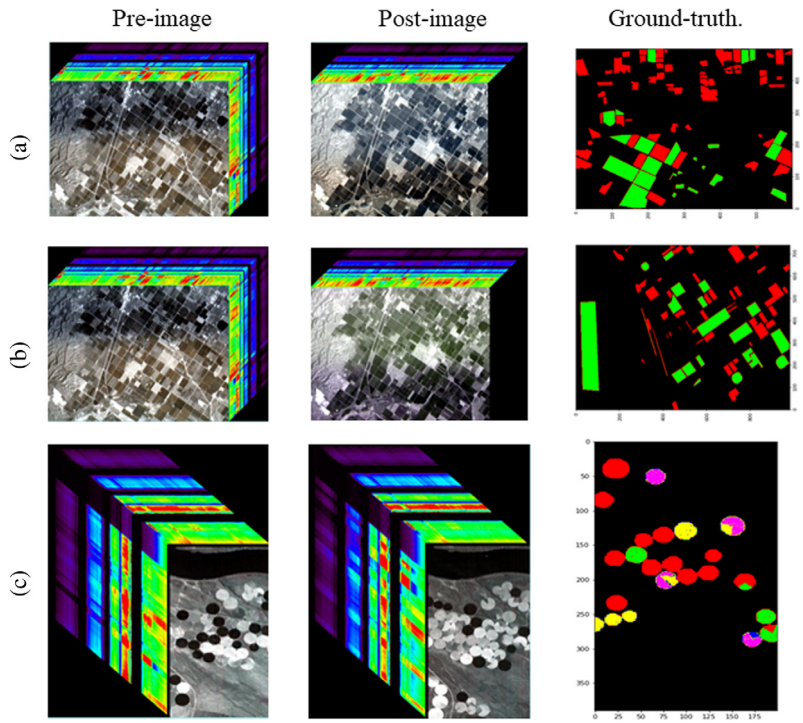

Fig. 1 Benchmark datasets: (a) Bay Area dataset, (b) Santa Barbara dataset, and (c) Hermiston dataset.

contains 224 spectral bands with a spatial resolution of about 30 m per pixel. The images were acquired in 2013 and 2014, respectively. The bitemporal images, as well as the ground truth, are shown in Fig. 1(b).

2.1.3 *Hermiston dataset*

This dataset consists of two coregistered hyperspectral images over the city of Hermiston, Oregon, of a section with (390 × 200) pixels acquired by the Hyperion sensor. Each image contains 242 spectral bands with a spatial resolution of about 30 m per pixel. The images were acquired in 2004 and 2007, respectively. The ground truth image contains five classes. The bitemporal images, as well as the ground truth, are shown in Fig. 1(c).

2.2 *Proposed Workflow*

In this paper, we present an efficient workflow for HSICD, as shown in Fig. 2, which is composed of four main phases: preprocessing, training, testing, and evaluation. The proposed workflow was inspired by semantic segmentation due to their booming performance in several applications, such as scene comprehension,[32] processing satellite images,[15,33] and object detection in satellite images.[34] UNet model,[35] which is considered a famous and effective semantic segmentation architecture, is used in the training phase to identify the change regions. In general, UNet employs the traditional encoder–decoder scheme. The input image is compressed into a dense feature vector by the encoder block. The spatial dimension of the feature vector is gradually reduced to obtain intense high discriminative representation. On the other hand, the feature vector has to spatially expand progressively to produce a segmented image. Several approaches such as bilinear interpolation and transposed convolution have been employed in the decoder block to match the original image dimensions.

The proposed workflow aims to simulate real-life scenarios in which the imbalanced class problem is a major challenge, especially in satellite imageries. Finally, the performance of the proposed workflow was measured in terms of precision, recall, F-measure, Kappa-coefficient, and overall accuracy (OA). The proposed workflow includes the following four primary phases:

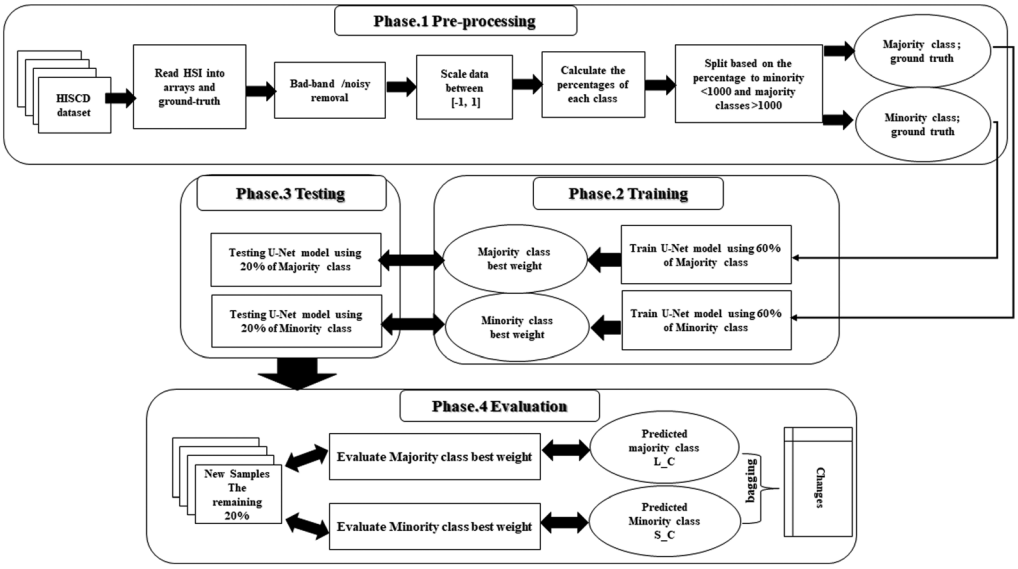

Fig. 2 The proposed phases for segmentation-based HSICD (HSICD_workflow).

1. Preprocessing: The bitemporal hyperspectral images were atmospherically corrected, and the bad and noisy bands were removed. The resulting images were scaled between $[-1, 1]$. The classes' distribution was computed to identify the majority and minority classes based on a threshold (>1000). We utilized ROS to handle the imbalanced class problem.
2. Training: To handle the class imbalance problem, we favor the algorithm level solution. We trained two semantic segmentation model with the same architecture, the first one using 60% of the majority class and the other one trained by 60% of the minority class. We adopted bagging ensemble to aggregate both models to generate the change map.
3. Testing: We iteratively carryied out 10-fold cross validation to obtain the best model weights and fine-tuned parameters for majority and minority classes. Both models were tested with 20% of the majority and minority classes to compute the model change identification performance.
4. Evaluation: The learned weights in the previous phase were employed to produce change patches for the bitemporal input image. First, for the remaining new samples 20% of bitemporal HIS were scaled to the $[-1, 1]$ range using the same procedure as in the training phase. Then, each input patch was fed to the learned model to obtain change patches by bagging the results to be merged. The final change map was produced, and for the overlapped regions between the adjacent patches, averaging was used to obtain the final pixel value.

In this work, we employed UNet model and four of its variants in the proposed workflow of HSICD, namely traditional UNet,[35] residual UNet (R UNet), attention UNet (Att- UNet), recurrent residual UNet (R2 UNet), and attention recurrent residual UNet (Att-R2 UNet).Traditional UNet is shown in Fig. 3(a). The first variant of UNet is Residual UNet,[38] which was introduced as an extension to benefit from the residual learning as shown in Fig. 3(b). The second variant is attention UNet,[37] which incorporates attention gates (AGs) to produce soft region proposals to highlight salient region of interest (ROI) features and suppress feature activations from irrelevant backgrounds. AG was plugged after the standard convolutional block in the decoder. The AG architecture[37] is shown in Fig. 3(c). The third variant is the recurrent residual UNet architecture,[36] shown in Fig. 3(d), in which the recurrent convolutional operation is measured at a discrete time. The last variant incorporated residual, recurrent, and AGs[37] into each encoder and decoder block to enrich information flow and enforce a semantic discriminative intermediate feature map at every scale.

Typically, the loss functions applied in segmentation are categorized into distribution-based losses (minimize dissimilarity between two distributions) ad region-based losses (minimize the

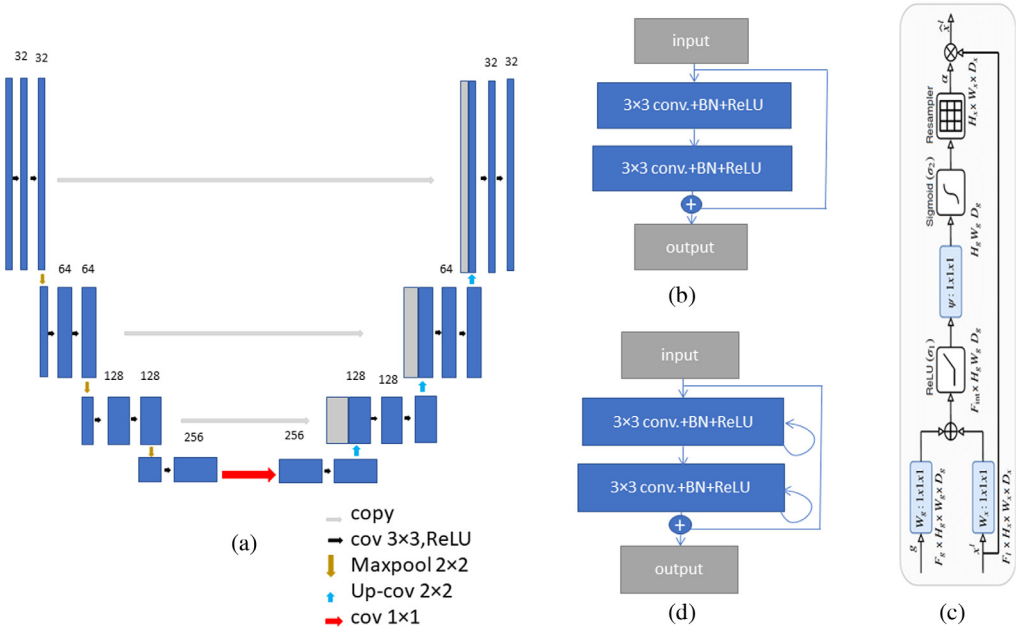

Fig. 3 Variant UNet architectures: (a) traditional UNet, after,[35] (b) residual UNet block,[36] (c) attention UNet block,[37] and (d) recurrent residual UNet block.

mismatch or maximize the overlap regions between the two images); details are given in Refs. 39,40. A common practice is to evaluate small subset of available loss functions to avoid the impracticability of experimenting on all available loss functions. In this work, we compared the performance of five widely used loss functions, namely cross-entropy loss, focal loss, Tversky loss, dice loss, and contrastive loss, to evaluate their performance in imbalanced HIS datasets.

3 Experimental Results and Analysis

3.1 *Evaluation Metrics*

The proposed HSICD workflow performance was evaluated based on precision, recall, F-measure, kappa coefficient, and OA.

Precision computed by Eq. (1) indicates the average of images that are correctly identified to the total number of images that are correctly and noncorrectly identified with the reference input:

$$\text{Percision}(P) = \frac{T_p}{T_p + F_p}, \tag{1}$$

where T_p and F_p represent the true positive images and the false positive images, respectively.

Recall, depicted in Eq. (2), is defined as the average number of images that are correctly identified out of the total number of images that are correctly and noncorrectly identified:

$$\text{Recall}(R) = \frac{T_p}{T_p + F_N}; \tag{2}$$

where F_N represents the false negative.

$F1$ score is defined by Eq. (3). If the obtained value reaches 1, it is classified as best, and if it reaches 0, as worst:

$$F\text{-measure} = \frac{2PR}{P+R}. \tag{3}$$

Kappa coefficient is calculated as

$$\text{Kappa} = \frac{\text{PCC} - \text{PRE}}{1 - \text{PRE}}, \tag{4}$$

where $\text{PRE} = \frac{(\text{TP}+\text{FP}).\text{MC}+(\text{FN}+\text{TN}).\text{MU}}{(\text{TP}+\text{FP}+\text{FN}+\text{TN})^2}$, $\text{PCC} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{FP}+\text{FN}+\text{TN})}$.

Finally, OA represents the proportion of correctly identified pixels to the total number of pixels.

3.2 *Experiment Setup*

In all experiments, each dataset was separated into three subsets, namely, training (60%), testing (20%), and evaluation (20%). We implemented a 10-fold cross-validation strategy to ensure balanced outcomes; patches in training and testing subsets are nonoverlapped. In the training and evaluation phases, the mutually exclusive dataset ensures an event that does not split the training and testing datasets.

For all variant UNet models, we eliminated one layer from the original UNet architecture and implemented a three-layer ($\#L = 3$) UNet version to cope with small input patches (16×16 pixels) as shallower architectures are easier to train due to the relatively smaller number of hyperparameters to be optimized. The encoder is preceded by a bridge layer and a three-layer, skip-linked decoding path. The adaptive moment estimation (Adam)[41] was selected to train the models due to its minimal tuning parameters requirement. The models were trained with a mini-batch size of 16, and the number of epochs and the learning rate were set to 20 and 0.0001, respectively. These parameters were chosen based on their empirically adequate performance. We conducted all experiments using an Intel (R) Core i7 3.40 GHz CPU with NVIDIA GeForce GTX 1080-Ti. Due to the computing resources limitations, the optimization of thetraining algorithm parameters may further improve the performance.

3.3 *Results and Discussion*

We conducted ample experiments to thoroughly analyze each UNet model's performance and inference time. The Bay Area dataset obtained results are shown in Table 2, which compares the performance of the five implemented models (UNet, R-UNet, Att-UNet, R2-UNet, and Att-R2-UNet). The performance of the results was calculated from the test set results over the 10-fold cross validation. The implemented Att-R2-UNet architecture performed better on semantic segmentation with respect to several metrics, with the highest OA equals to 94.99% and a maximum precision score of 93.23%. The lowest OA score was reported for the traditional UNet model and is equal to 91.5%.

For the residual UNet and recurrent residual UNet architectures, the obtained OA results are auspicious (0.93 and 0.92), despite their naive architectures. Finally, the attention UNet and attention recurrent residual UNet architectures present higher performance in comparison with the

Table 2 Results obtained for HSICD using variant UNet models on the Bay Area dataset.

	UNet	R-UNet	Att-UNet	R2-UNet	Att-R2-UNet
Accuracy	0.915088	0.930299	0.937725	0.919060	0.949942
Precision	0.828569	0.928306	0.884508	0.879393	0.932373
Recall	0.928973	0.826899	0.910247	0.818650	0.893517
*F*1 score	0.867821	0.870907	0.896927	0.841313	0.911859
Kappa (Cohen's kappa)	0.795920	0.800804	0.838752	0.776492	0.863853

traditional UNet and residual UNet models (0.93 and 0.95) since the spatial pyramid pooling outcome is combined with recurrent and recalibrated features from the encoder blocks. Overall, the obtained results revealed that the more naive decoder's architecture leads to lower test accuracy.

Furthermore, the diversity and limitations of the performance of the proposed workflow can also be confirmed by the results in Table 3 obtained on the Santa Barbara dataset. It can be observed that there is a trade-off between the simplicity of network architecture and the obtained accuracy. More specifically, the UNet and residual UNet architectures present a relatively comparable accuracy. The same can be observed for the Att-UNet architecture; nonetheless, the OA was improved at the cost of the Cohen's kappa metric. Thus, the R2-UNet model and the Att-R2 UNet can be considered to be the most effective since the OA for both of them are very close. Overall, almost all UNet models denoted comparable accuracies. The Att-R2 UNet model achieved the best performance numerically and visually in both the Santa Barbara and Bay Area datasets.

Figures 4 and 5 show the visual results of the obtained change maps from variant UNet segmentation models. The residual UNet model presents adequate performance for both benchmark datasets. On the contrary, the traditional UNet model demonstrated the lowest accuracy in identifying positive and negative changes. Moreover, the traditional UNet model fails to generate a change map that correctly captures change and no-change regions. The accuracy is significantly improved based on the visual results by integrating recurrent and residual learning.

Table 3 Results obtained for HSICD using variant UNet models on the Santa Barbara dataset.

	UNet	R UNet	Att-UNet	R2 UNet	Att-R2 UNet
Accuracy	0.944399	0.947131	0.953430	0.961011	0.954615
Precision	0.899405	0.910179	0.891454	0.920812	0.926634
Recall	0.885115	0.878233	0.934358	0.933101	0.895326
*F*1 score	0.890932	0.891838	0.911839	0.926548	0.910192
Kappa (Cohen's kappa)	0.816347	0.820643	0.854150	0.875530	0.848560

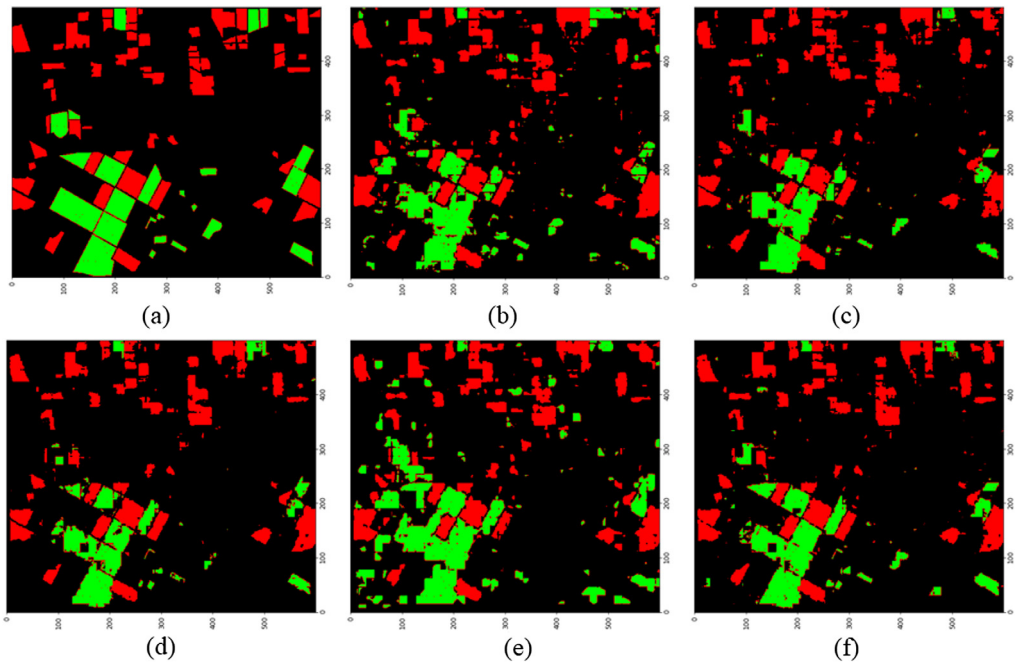

Fig. 4 Bay Area dataset change map: (a) ground-truth, (b) traditional UNet, (c) R-UNet, (d) Att- UNet, (e) R2-UNet, and (f) Att-R2-UNet.

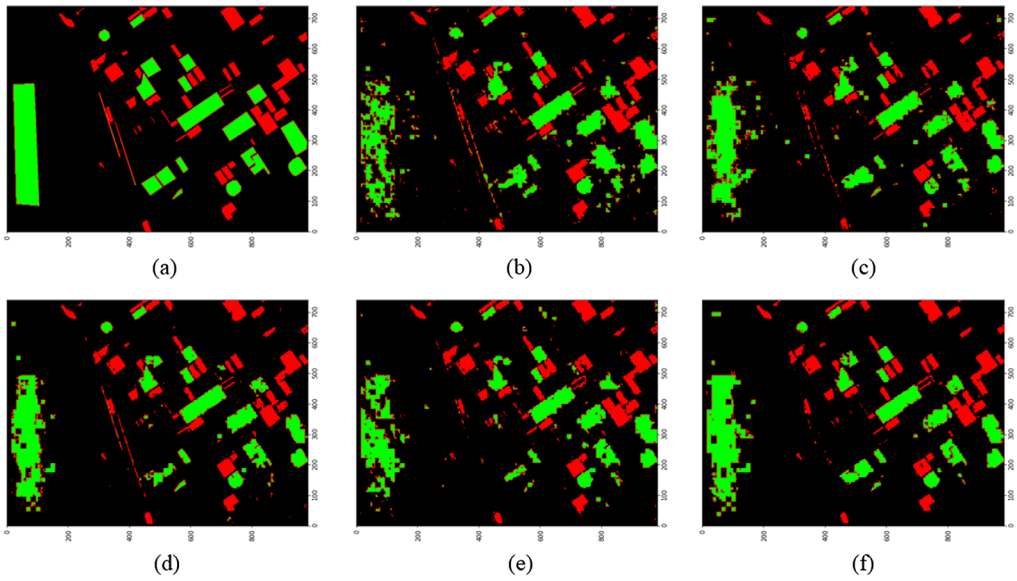

Fig. 5 Santa Barbara dataset change map (a) ground-truth, (b) traditional UNet, (c) R UNet, (d) Att- UNet, (e) R2 UNet, and (f) Att-R2 UNet.

Furthermore, the rich salient regions obtained from AGs in the Att-UNet and R2 UNet models tend to be more robust in binary change identification; however, some false positives were reported.

Next, we evaluated the efficiency of the proposed workflow at detecting multiclass changes using the Hermiston dataset as illustrated in Table 4. In particular, the obtained OA for traditional UNet demonstrated the worst value 0.94 (OA). However, the utilization of residual and recurrent blocks enhanced the accuracy of R-UNet (0.989) and R2-UNet (0.953). Visually, Figs. 6(a)–6(c) demonstrated the enhancement of change map when incorporating residual and recurrent blocks. Moreover, attention mechanism shows more robust results, especially in small ROIs. Att-R2 UNet achieved the highest OA (0.991) on the pixels compared with all other UNet architectures. The obtained results showed that all UNet models could learn effectively change features from hyperspectral images in multiclass change cases. To sum up, all CD experiments confirmed that the integration of residual, recurrent, and attention mechanism facilitates a spectral–spatial–temporal change feature to be constructed effectively.

Furthermore, we carried out comparison between the deployed models to analyze the execution time and memory required for inference. The average inference time and the number of parameters of each model are given in Table 5. Overall, traditional UNet among all variant models presents the fastest in terms of inference time. The R-UNet and R2-UNet models demonstrate a higher inference time in spite of the number of parameters being lower compared with the traditional UNet model. This is justifiable because of the residual operations employed in both models at the encoding and decoding stages. The attention residual recurrent model presents the highest number of parameters allocation and displays the best performance for both binary and

Table 4 Results obtained for HSICD using variant UNet models on the Hermiston dataset.

	UNet	R UNet	Att- UNet	R2 UNet	Att-R2 UNet
Accuracy	0.945470	0.989402	0.986232	0. 953387	0.991611
Precision	0.935675	0.948417	0.900143	0.978676	0.958538
Recall	0.951087	0.923722	0.908169	0.920067	0.946333
*F*1 score	0.942151	0.935821	0.893870	0.919009	0.952342
Kappa (Cohen's kappa)	0.950427	0.945470	0.930937	0.900139	0.957096

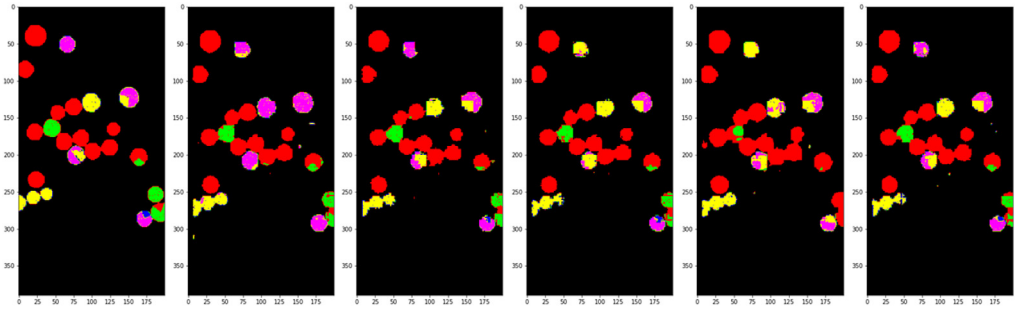

Fig. 6 Hermiston dataset change map (a) ground-truth, (b) traditional UNet, (c) R-UNet, (d) Att-UNet, (e) R2-UNet, and (f) Att-R2-UNet.

Table 5 Comparison of variant UNet models in terms of mean inference time and number of parameters.

Model	Inference time	#Parameters
UNet	102.491962	7,787,270
R-UNet	105.8280697	1,358,502
Att-UNet	126.4024163	23,886,470
R2-UNet	108.850071	7,916,969
Att-R2-UNet	130.6800033	24,016,169

multichange identification cases. In conclusion, the recurrent residual UNet model was an ideal solution for binary and multichange identification for the hyperspectral problem with its high performance and relatively comparable inference speed.

Next, we conducted various experiments to evaluate the following loss functions: focal loss, Dice loss, Tversky loss, and contrastive loss on the three HSI benchmarks using the standard UNet architecture described above. Based on the results in Table 6 from the Bay Area dataset, we select the contrastive loss, Dice loss, and focal loss as the top three performing loss functions. As shown in Table 6, in Hermiston dataset, the focal loss was associated with the best recall–precision balance, and it outperformed the contrastive loss and dice loss in recall and precision scores.

Finally, Figs. 7, and 8 present Bland–Altman plots and linear regression plots for area (segmented) and area (truth) for the Bay Area dataset and Santa Barbara dataset, respectively. This experiment was conduct using the standard UNet to visualize the robustness of the proposed

Table 6 UNet performance using variant loss functions on hyperspectral datasets. Numbers in boldface denote the highest values for each metric.

	Bay Area dataset			Santa Barbara dataset			Hermiston dataset		
Loss function	OA	Precision	Recall	OA	Precision	Recall	OA	Precision	Recall
Focal	0.935	0.886	0.918	0.942	0.909	0.926	**0.965**	**0.907**	**0.939**
Dice	0.941	0.876	**0.929**	0.947	0.922	**0.936**	0.944	0.878	0.916
Tversky	0.909	0.875	0.916	0.932	0.988	0.914	0.924	0.899	0.921
Contrastive	**0.961**	**0.891**	0.817	**0.956**	**0.927**	0.933	0.953	0.861	0.892
Cross-entropy	0.915	0.828	0.928	0.944	0.899	0.885	0.945	0.935	0.931

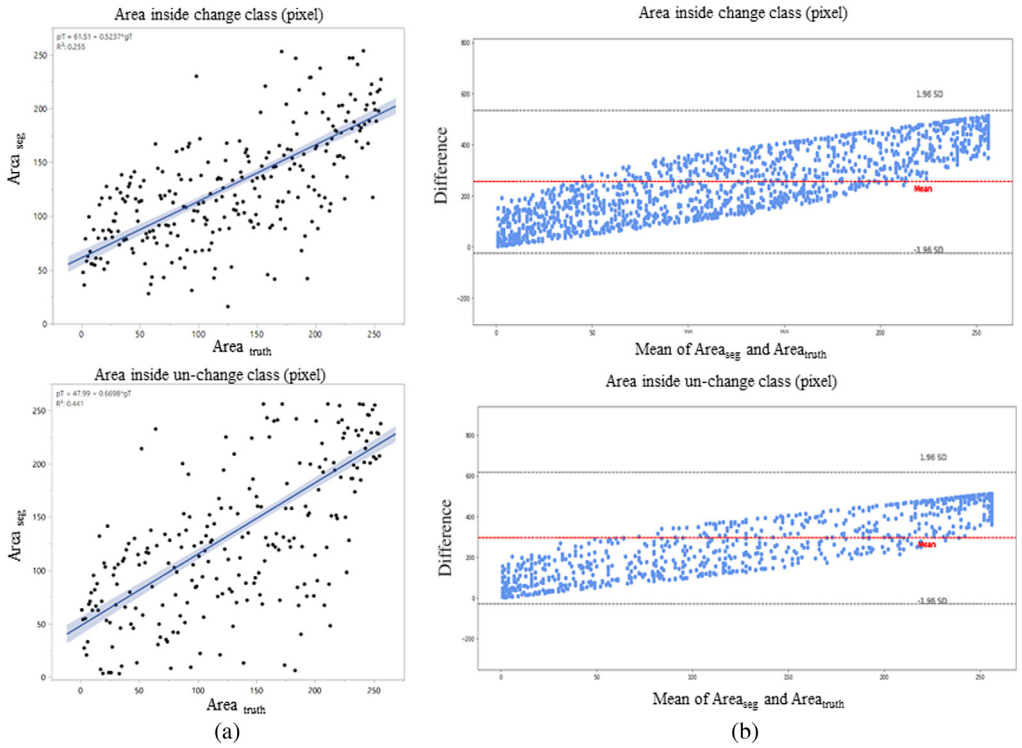

Fig. 7 (a) Linear regression results and (b) Bland–Altman plots for the comparison of change and no-change areas detected by the proposed workflow and corresponding ground truth for the Bay Area dataset.

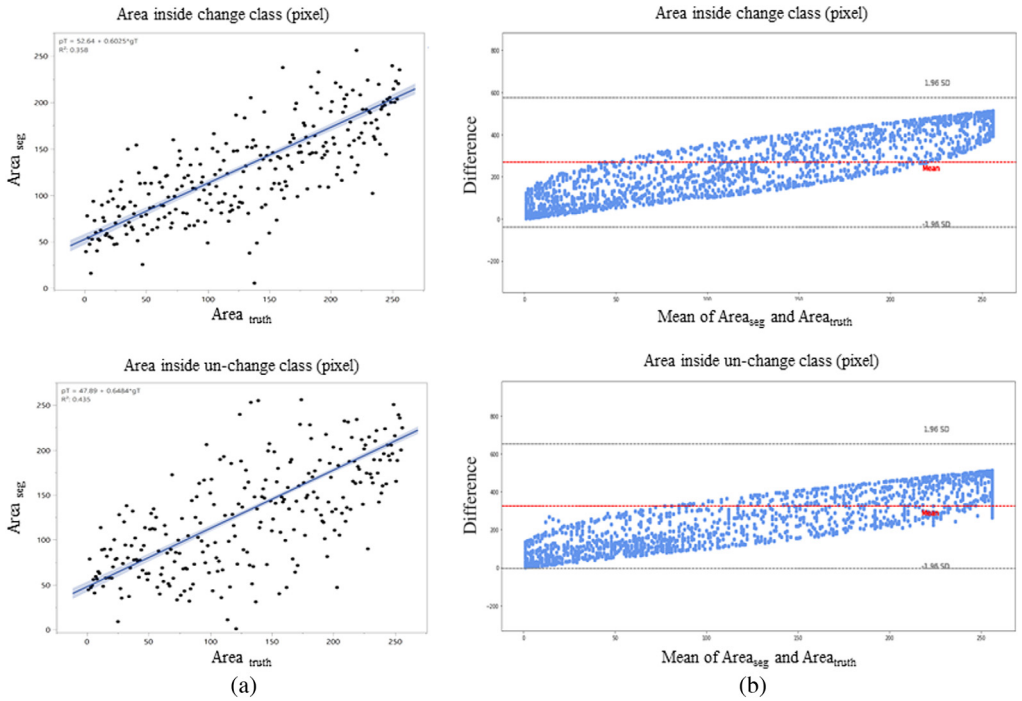

Fig. 8 (a) Linear regression results and (b) Bland–Altman plots for the comparison of change and no-change areas detected by the proposed workflow and corresponding ground truth for the Santa Barbara dataset.

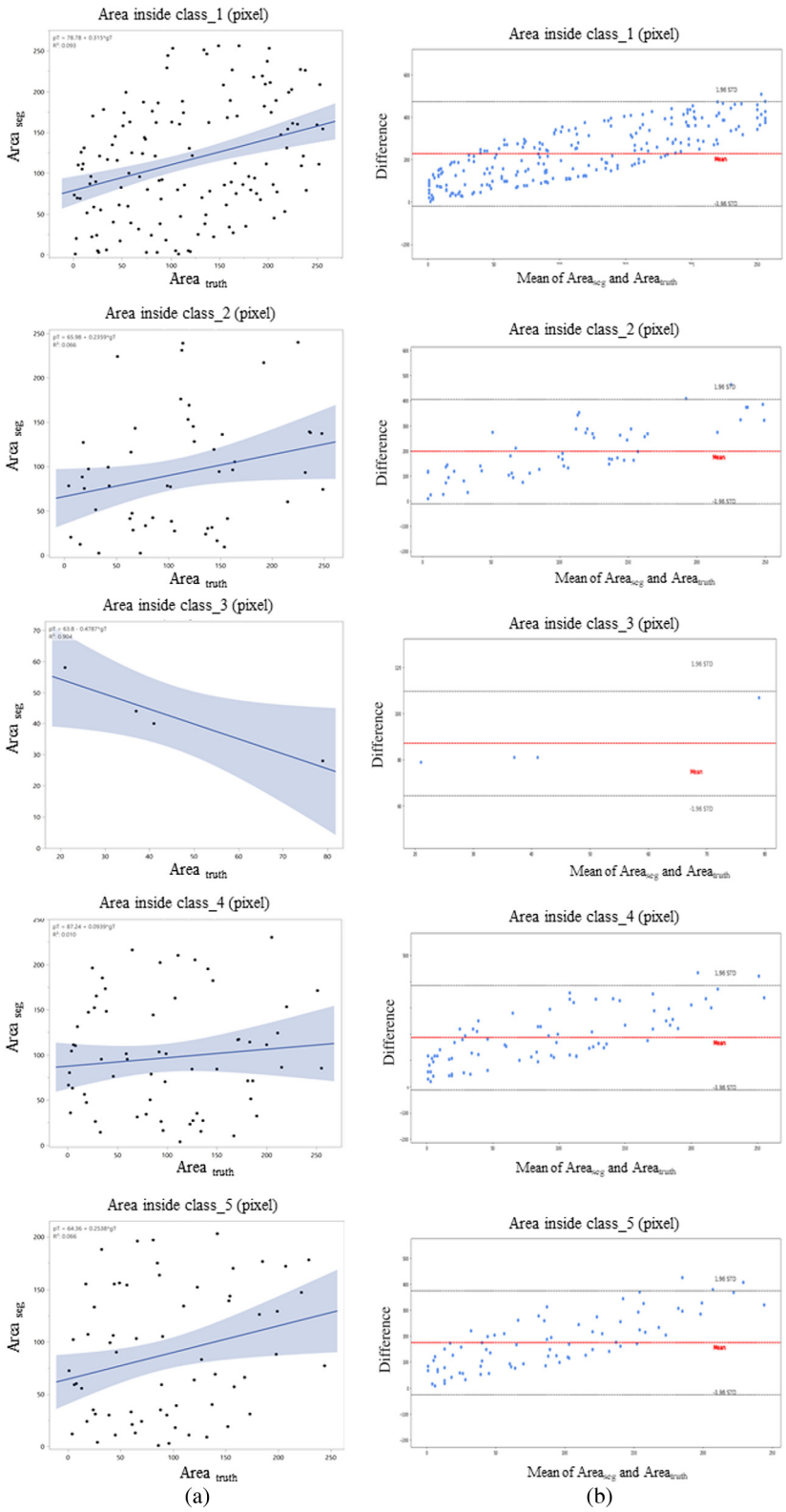

Fig. 9 (a) Linear regression results and (b) Bland–Altman plots for the comparison of five change class detected by the proposed workflow and corresponding ground truth for the Hermiston dataset.

workflow to identify the change and no-change zones. Specifically, the linear regression analysis (Figs. 7 and 8) indicates correlation with $R2 = 0.255$ and 0.358 for the identification of change zones for the Bay Area and Santa Barbara datasets, respectively. On the other hand, $R2 = 0.441$ and 0.435 for the unchanged zones. Bland–Altman plots indicate a slight bias for detecting change zones detection. Figure 9 shows Bland–Altman plots and linear regression plots for each class in the Hermiston dataset.

4 Conclusions

This paper proposes a CD workflow for bitemporal hyperspectral datasets based on DL segmentation. The workflow is composed of four phases, namely preprocessing, training, testing, and evaluation. We incorporate ROS in preprocessing, DL, and bagging ensemble to handle imbalanced dataset. The obtained results imply that the proposed workflow contributes significantly to future research activity regarding change identification in hyperspectral imageries. The contributions of this work can be summarized as follows:

- Four variant UNet models, namely residual UNet (R-UNet), residual recurrent UNet (R2-UNet), attention UNet (Att-UNet), and attention residual recurrent UNet (Att-R2-UNet), were implemented. We compared these models with traditional UNet's ability to segment and classify change and no-change regions.
- Extensive analytical experiments were conducted on three hyperspectral benchmark datasets. The imbalanced class distribution was addressed in the proposed workflow while training the DL models.
- The UNet-based CD algorithm accurately reveals the changed and unchanged areas using convolutional layers.

The obtained results show that the proposed workflow attention residual recurrent UNet (Att_R2_UNet)-based CD architecture successfully highlights the change and no change areas. Furthermore, the attention residual recurrent model presents the highest number of parameters allocation and displays the best performance for both binary and multichange identification cases. Therefore, the recurrent residual UNet model was an ideal solution for binary and multichange identification for hyperspectral problem with its high performance and relatively comparable inference speed. This study strengthens the idea that deep neural networks can learn highly complicated features, and when combined with HSI data they might have potential to improve HSI CD.

References

1. T. Krauß and J. Tian, "Automatic change detection from high-resolution satellite imagery," in *Remote Sensing for Archaeology and Cultural Landscapes*, D. Hadjimitsis et al., Eds., pp. 47–58, Springer, Cham, Switzerland (2020).
2. S. Liu et al., "A review of change detection in multitemporal hyperspectral images: current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.* **7**(2), 140–158 (2019).
3. F. Huang, Y. Yu, and T. Feng, "Hyperspectral remote sensing image change detection based on tensor and deep learning," *J. Vis. Commun. Image Represent.* **58**, 233–244 (2019).
4. V. Ignatiev et al., "Targeted change detection in remote sensing images," in *Eleventh Int. Conf. Mach. Vision (ICMV 2018)*, p. 110412H (2019).
5. H. Jafarzadeh and M. Hasanlou, "An unsupervised binary and multiple change detection approach for hyperspectral imagery based on spectral unmixing," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **12**, 4888–4906 (2019).
6. C. Wu et al., "Unsupervised deep slow feature analysis for change detection in multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.* **57**(12), 9976–9992 (2019).
7. L. Khelifi and M. Mignotte, "Deep learning for change detection in remote sensing images: comprehensive review and meta-analysis," *IEEE Access* **8**, 126385–126400 (2020).

8. M. Yu et al., "Spatiotemporal event detection: a review," *Int. J. Digital Earth* **13**, 1339–1365 (2020).
9. W. Shi et al., "Change detection based on artificial intelligence: state-of-the-art and challenges," *Remote Sens.* **12**(10), 1688 (2020).
10. G. Destefanis et al., "The use of principal component analysis (PCA) to characterize beef," *Meat Sci.* **56**(3), 255–259 (2000).
11. B. Wang et al., "Application of IR-MAD using synthetically fused images for change detection in hyperspectral data," *Remote Sens. Lett.* **6**(8), 578–586 (2015).
12. S. Talukdar et al., "Land-use land-cover classification by machine learning classifiers for satellite observations—a review," *Remote Sens.* **12**(7), 1135 (2020).
13. L. Li et al., "Semi-supervised fuzzy clustering with feature discrimination," *PloS One* **10**(9), e0131160 (2015).
14. J. Cervantes et al., "A comprehensive survey on support vector machine classification: applications, challenges and trends," *Neurocomputing* **408**, 189–215 (2020).
15. M. M. Elkholy et al., "Hyperspectral unmixing using deep convolutional autoencoder," *Int. J. Remote Sens.* **41**(12), 4799–4819 (2020).
16. M. S. Moustafa, S. Ahmed, and A. A. Hamed, "Learning to hash with convolutional network for multi-label remote sensing image retrieval," *Int. J. Intell. Eng. Syst.* **13**(5), 539–548 (2020).
17. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436 (2015).
18. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press (2016).
19. C. Dong et al., "Learning a deep convolutional network for image super-resolution," in *Eur. Conf. Comput. Vision*, pp. 184–199 (2014).
20. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, pp. 1097–1105 (2012).
21. Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," arXiv:1510.03820 (2015).
22. A. Khan et al., "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.* **53**(8), 5455–5516 (2020).
23. L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.* **57**(2), 924–935 (2018).
24. H. Lyu, H. Lu, and L. Mou, "Learning a transferable change rule from a recurrent neural network for land cover change detection," *Remote Sens.* **8**(6), 506 (2016).
25. N. Venugopal, "Automatic semantic segmentation with DeepLab dilated learning network for change detection in remote sensing images," *Neural Process. Lett.* **51**, 2355–2377 (2020).
26. J. López-Fandiño et al., "Stacked autoencoders for multiclass change detection in hyperspectral images," in *IGARSS 2018-2018 IEEE Int. Geosci. Remote Sens. Symp.*, pp. 1906–1909 (2018).
27. X. Zhang et al., "Two-phase object-based deep learning for multi-temporal SAR image change detection," *Remote Sens.* **12**(3), 548 (2020).
28. H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 1520–1528 (2015).
29. F. Thabtah et al., "Data imbalance in classification: experimental evaluation," *Inf. Sci.* **513**, 429–441 (2020).
30. H. G. Zefrehi and H. Altınçay, "Imbalance learning using heterogeneous ensembles," *Expert Syst. Appl.* **142**, 113005 (2020).
31. J. López-Fandiño et al., "GPU framework for change detection in multitemporal hyperspectral images," *Int. J. Parallel Programm.* **47**(2), 272–292 (2019).
32. V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
33. C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.* **15**(12), 1867–1871 (2018).

34. A. Mahmoud et al., "Object detection using adaptive mask RCNN in optical remote sensing images," *Int. J. Intell. Eng. Syst* **13**, 65–76 (2020).
35. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
36. M. Z. Alom et al., "Recurrent residual U-Net for medical image segmentation," *J. Med. Imaging* **6**(1), 014006 (2019).
37. O. Oktay et al., "Attention U-Net: learning where to look for the pancreas," in *1st Conf. Med. Imaging with Deep Learn. (MIDL 2018)*, Amsterdam, Netherlands (2018).
38. Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.* **15**(5), 749–753 (2018).
39. J. Ma et al., "Loss odyssey in medical image segmentation," *Med. Image Anal.* **71**, 102035 (2021).
40. M. Yeung et al., "A mixed focal loss function for handling class imbalanced medical image segmentation," arXiv:2102.04525 (2021).
41. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Int. Conf. Learn. Represent. (ICLR)* (2015).

Marwa S. Moustafa received her PhD in computer and information sciences from Ain Shams University, Egypt, in 2016. She is currently working as a researcher at the National Authority for Remote Sensing and Space, Egypt. She has expertise in the domains of image processing, machine learning, and computational intelligence.

Sayed A. Mohamed received the PhD in engineering science of electronic and system engineering from the Faculty of Electronic Engineering of Menofia University in 2015. He is currently a manager of the Ground Receiving Station and researcher at the National Authority for Remote Sensing and Space Sciences, Egypt. He has published various scientific papers in national and international journals. He has over 20 years of experience in the field of receiving and processing of remotely sensed data.

Sayed Ahmed received his BS degree in computer sciences from Asyut University, Egypt, in 2007 and his MSc degree in computer sciences from Cairo University, Egypt, in 2019. Currently, he is pursuing his PhD in computer sciences. His primary research interests are GIS and its applications, cloud computing, mobile computing, big data, statistical analysis, software engineering, and data sciences.

Ayman H. Nasr is an emeritus professor in NARSS. He received his BSc and MSc degrees and his PhD degrees in electronics and communications, Faculty of Engineering, Cairo University, Egypt. His research expertise includes image processing and GIS in remote sensing fields. He has published 55 papers in various international peer-reviewed journals and conferences. He has contributed to the publication of four atlases and participated in 60 research projects. He was also a member of the NARSS Board of Administration.