The Egyptian Journal of Remote Sensing and Space Sciences 25 (2022) 673–685

HOSTED BY

Contents lists available at ScienceDirect

The Egyptian Journal of Remote Sensing and Space Sciences

journal homepage: www.sciencedirect.com

Research Paper

Short-time-series grassland mapping using Sentinel-2 imagery and deep learning-based architecture

Abolfazl Abdollahi [a], Yuxia Liu [b], Biswajeet Pradhan [a,c,d,*], Alfredo Huete [a,b], Abhirup Dikshit [a], Ngoc Nguyen Tran [b]

[a] Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering and IT, University of Technology Sydney, Ultimo NSW 2007, Australia
[b] Faculty of Science, University of Technology Sydney, Ultimo NSW 2007, Australia
[c] Center of Excellence for Climate Change Research, King Abdulaziz University, P. O. Box 80234, Jeddah 21589, Saudi Arabia
[d] Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 16 March 2022
Revised 3 June 2022
Accepted 5 June 2022
Available online 15 June 2022

Keywords:
Grassland mapping
Deep learning
Sentinel-2 image
Remote sensing

ABSTRACT

In the present work, a deep learning-based network called LeNet is applied for accurate grassland map production from Sentinel-2 data for the Greater Sydney region, Australia. First, we apply the technique to the base date Sentinel-2 data (non-seasonal) to make the vegetation maps. Then, we combine short time-series (seasonal) data and enhanced vegetation index (EVI) information to the base date imagery to improve the classification results and generate high-resolution grassland maps. The proposed model obtained an overall accuracy (OA) of 88.36% for the mono-temporal data, and 92.74% for the multi-temporal data. The experimental products proved that, by combining the short time-series images and EVI information to the base date, the classification maps' accuracy is increased by 4.38%. Moreover, the Sentinel-2 produced grassland maps are compared with the pre-existing maps such as Australian Land Use and Management (ALUM) 50 m resolution and Dynamic Land Cover Dataset (DLCD) with 250 m resolution as well as some traditional machine learning methods such as Support Vector Machine (SVM) and Random Forest (RF). The results show the effect of the LeNet network's performance and efficiency for grassland map production from short time-series data. As a result, decision-makers and urban planners can benefit from this work in terms of grassland change identification, monitoring, and planning assessment.

1. Introduction

One of the major research topics in the field of remote sensing is vegetation mapping based on the spectral properties of targets (Rapinel et al., 2019). Vegetation cover maps are undoubtedly in high demand for global change studies, natural resource monitoring and management, desertification evaluation, and many other Earth observation applications (Foley et al., 2005), (Hansen and Loveland, 2012). Due to its high spatial and temporal resolution, Sentinel-2 data has gained global attention among the remote sensing community. It is the ESA (European Space Agency) latest generation Earth observation mission that provides high temporal (ten days/five days) with a spatial resolution of 10–60 m (Cisneros et al., 2020). Also, it has gained a lot of attention in research because of its global coverage and is freely downloadable (Sonobe et al., 2018). Traditional pixel-based and object-based remote sensing imagery classification methods are commonly used to create vegetation cover maps; however, it remains a challenge for the remote sensing community to process remote sensing imagery in a time-effective manner and produce accurate maps (Gómez et al., 2016).

The emergence of Deep Learning (DL) has risen to prominence in the field of remote sensing. Feature extraction ability has been enhanced by the recent advancement of convolutional neural networks (CNN) structures. As a result, the principal issues of traditional machine learning techniques (inefficient in performance and time-consuming) have been improved with these sophisticated deep learning techniques. Furthermore, particular challenges such as (1) high temporal, structural, and spatial variety of vegetation composition, (2) spectral similarity, and (3) the small spatial extent of grasslands render grassland mapping a problematic task. There-

* Corresponding author at: Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering and IT, University of Technology Sydney, Ultimo NSW 2007, Australia.

E-mail addresses: abolfazl.abdollahi@student.uts.edu.au (A. Abdollahi), Yuxia.Liu@student.uts.edu.au (Y. Liu), Biswajeet.Pradhan@uts.edu.au (B. Pradhan), Alfredo.Huete@uts.edu.au (A. Huete), Abhirup.Dikshit@student.uts.edu.au (A. Dikshit), Ngoc.Tran@uts.edu.au (N. Nguyen Tran).

https://doi.org/10.1016/j.ejrs.2022.06.002
1110-9823/

fore, it is important to provide high-quality grassland maps to better monitor and assess the grassland changes, preserve and rehabilitate the grass areas. Previous research has shown that mapping grasslands from single-date remote sensing data is difficult due to the comparable physiognomy of some communities. Also, assessing grass cover in urban and *peri*-urban areas is difficult due to small and irregular grass patch sizes (e.g., road strips, small parks, abandoned lots, etc.). In addition, grass mapping in urban areas is currently accomplished via costly and infrequent airborne photography. Thus, in the current study, the objective is to apply a deep learning-based tool called LeNet network, which is a spectral-temporal CNN network to freely available short-time-series Sentinel-2 images of Enhanced Vegetation Index (EVI) greenness values to generate high-resolution grassland maps for the Sydney region, Australia. The model is able to classify the images more accurately than traditional methods because it can automatically learn temporal (and spectral) properties using convolutions in the temporal dimension.

The main purpose of the methodology is to initially use the base date (mono-temporal) of Sentinel-2 data and implement the proposed method to classify vegetation cover and generate grassland maps. After that, we added short time series Sentinel-2 images (temporal) with EVI to the base date to improve the classification accuracy and produce high-quality grassland maps. To the best of our knowledge, the presented LeNet approach has not been implemented in the literature, and this is the first time that the network is implemented on short-time-series Sentinel-2 imagery for grassland maps production. Finally, in comparison with the pre-existing classification maps such as Australian Land Use and Management (ALUM) and Dynamic Land Cover Dataset (DLCD), the presented LeNet network could improve the results, achieving higher classification accuracy and producing high-resolution grassland maps for the region. Moreover, we compared both quantitative and qualitative results obtained by the deep learning model with some popular conventional machine learning methods such as Support Vector Machine (SVM) and Random Forest (RF). The experimental outcomes demonstrated that the LeNet model tackled the abovementioned challenges presented by traditional grassland mapping machine learning methods, improved the performance, and produced high-resolution grassland maps with high spatial consistency.

The remainder of the paper is divided into the sub-sections mentioned below. Section 2 presents related works, and Section 3 delves into the detailed descriptions of data, study area, and the specifics of the proposed LeNet network for classifying vegetation cover maps and producing grassland maps based on Sentinel-2 images. Section 4 highlights measurement metrics and experimental outcomes. Section 5 contains comprehensive qualitative and quantitative comparisons of the produced maps with the other pre-existing classification maps. Finally, in Section 6, we will discuss the conclusion and critical results.

2. Literature review

The usefulness of Sentinel-2 has been evaluated and has shown its high potential for vegetation mapping (Lefebvre et al., 2016). However, the vegetation mapping results are affected not only by the suitability of image-derived features but also by the right choice of classification approach (Lu and Weng, 2007). Diverse classification approaches for vegetation classification, and grassland generation maps based on remote sensing data have been implemented and deployed in the literature (Trisurat et al., 2000). These algorithms include parametric supervised systems, such as random forest (RF), decision tree (DT), artificial neural network (ANN), support vector machine (SVM), and k-Nearest Neighbors (KNN); to unsupervised methods such as K-means and ISODATA (Sohn and Rebello, 2002).

In the following, some of the previous works that applied various classification methods for classifying vegetation cover and generating grassland maps are discussed. For example, (Feng et al., 2015) presented a hybrid method based on texture analysis and RF algorithm to differentiate urban vegetation cover from Unmanned Aerial Vehicle (UAV) imagery. Schuster et al. (2015) performed the SVM method to classify synthetic aperture radar (TerraSAR-X) and multi-spectral (RapidEye) images and generate grassland habitats maps in northeastern Germany. Esch et al. (2014) implemented the tree C5.0 classifier to classify multi-temporal AWiFS images and identified primary grassland and crop types for the Brandenburg and Mecklenburg Western Pomerania regions, Germany. Their study demonstrated that for the separation of grassland and crops during spring time, additional data acquisitions are required. The traditional machine learning algorithms have shown some limitations when processing big satellite data for vegetation cover classification. For instance, (Zhang and Xie, 2013) stated that RF and SVM are sensitive to overtraining and noise. Also, (Pande-Chhetri et al., 2017) indicated that in terms of computational processing, the ANN approach has a high degree of complexity. Naidoo et al. (2012) specified that it is difficult to set an ideal value of k for the KNN method. According to (Prasad et al., 2006), DT has a tendency to overfit the model and is often inconsistent. Also, they reported that DT is excessively sensitive to minor changes in the training dataset.

In contrast, CNN has resulted in a number of advancements in many remote sensing applications and vegetation mapping domains. For example, Lang et al. (2019) applied a CNN model to map vegetation height using Sentinel-2 imagery. For regressing vegetation height, they collected multi-spectral Sentinel-2 imagery for a study area in Switzerland for several months and trained the CNN model to obtain appropriate textural and spectral information from reflectance images. Nijhawan et al. (2017) used a deep learning (DL) framework for vegetation cover mapping in Uttarakhand, India, that combined CNNs with local binary patterns. They mixed topographic, texture, and multi-spectral Sentinel-2 (10 m spatial resolution) data to train the model and produce the maps.

CNN is distinguished by its deep architecture (multilayer interconnected channels), ability to categorize concurrently, and capability to set parameters simultaneously, which has a high capacity to automatically learn the classifiers and features from data (Nogueira et al., 2017). Thus, in the current study, we present a deep learning-based technique named LeNet model to classify remote sensing imagery and generate grassland maps. LeNet is a popular network among CNNs that is commonly used for image classification. As a pre-processing stage, this network eliminates the requirement for manually extracted characteristics. In practice, the model can identify patterns straight from normalized input data. LeNet uses its backpropagation (BP) approach to create a self-studying processing system by integrating feature extraction and pattern recognition. Through the self-studying process, the model can determine the most effective attributes for categorization. Thus, we perform the model on the short time-series Sentinel-2 data to alleviate the limitations of traditional machine learning methods (e.g., lack the capacity to reduce salt-and-pepper class noise and lack the capability to classify specific pixels with similar spectral values accurately) and produce high-quality grassland maps.

3. Methodology

3.1. Test area and data

The test area is an urban and *peri*-urban part of Sydney which is situated on Australia's east coast at 33° 51′ 54.5148″ S and 151° 12′ 35.6400″ E (Fig. 1). For the data, short time-series Sentinel-2 satel-

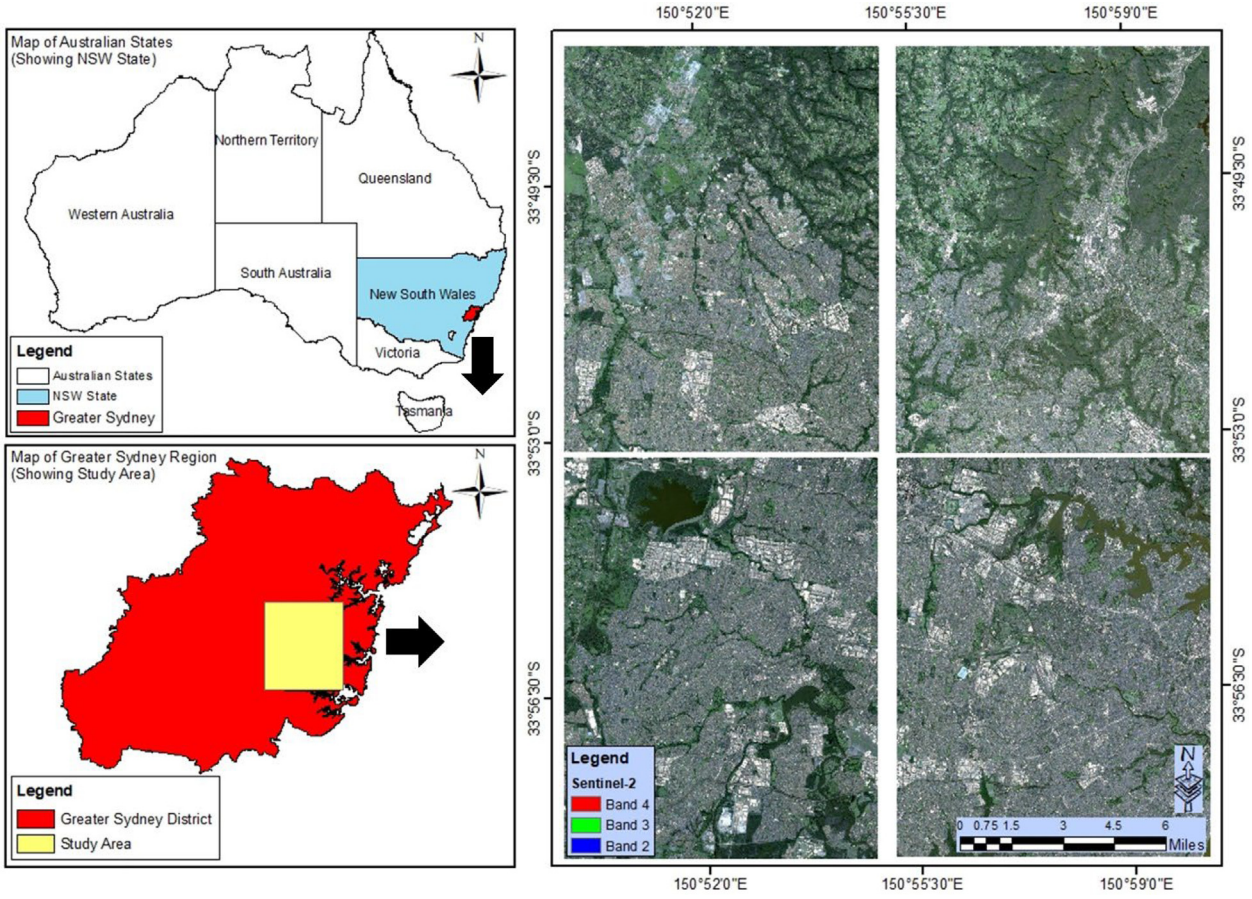

Fig. 1. Study area map of Sydney, Australia showing the Sentinel-2 data.

lite imagery (Table 1) was acquired from the Copernicus Open Access Hub to implement the classification. The product form is Level-2A, and the dataset contains 6 different cloud free dates, such as 19/02/2020, 14/04/2020, 24/04/2020, 04/05/2020, 03/06/2020 and 23/06/2020. We first selected the base date (04/05/2020) to implement the classification method and produce grassland maps. The base date was selected from a set of temporal EVI values, and the image with the highest EVI values over grassland was the specified base date. Then, we combined short-time-series EVI images including all the images with various dates with the proposed model to generate higher classification accuracy for our method. Level-2A products provide top of atmosphere (TOA) surface reflectance and they are geometrically, atmospherically and radiometrically corrected (Vasilakos et al., 2020).

Table 1
Sentinel-2 satellite data spatial and spectral resolution.

Bands	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
Band 1	443	20	60
Band 2	490	65	10
Band 3	560	35	10
Band 4	665	30	10
Band 5	705	15	20
Band 6	740	15	20
Band 7	783	20	20
Band 8	842	115	10
Band 8a	865	20	20
Band 9	945	20	60
Band 10	1380	30	60
Band 11	1610	90	20
Band 12	2190	180	20

3.2. Samples datasets

We manually analyzed the downloaded Sentinel-2 images for 6 different dates, as well as high-resolution Google Earth imagery, to collect samples for the training dataset. For compiling the training samples, the Fishnet tool in the ArcGIS 10.6 toolbox was utilized to cover the whole area of interest and to generate polygons for each class. In the next step, we used a data normalization method to boost the progress of activation functions and gradient descent optimization. Thus, we applied one of the most common methods for avoiding unusual gradients and normalizing the pixel values, which is called feature scaling or min–max normalization method. In this method, each future's maximum value is converted to a 1, the minimum value is converted to a 0, and all other values are converted to a decimal between 0 and 1. The feature scaling method can be calculated as:

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where, the minimum and maximum values in x is defined as min and max, respectively, and z is the normalized data.

3.3. The architecture of LeNet

Figs. 2 and 3 illustrate the applied LeNet network structure for grassland map production. Fig. 2 shows the applied model's architecture on the base date images (mono-multi-temporal imagery), while Fig. 3 depicts the model's structure for the multi-temporal images. In the first structure (Fig. 2), we only used the spectral information of the base date as an input, while in this structure, we added short-time-series Sentinel-2 images (spectral information for various multi-temporal images along with EVI information used as an input) to the model to improve the classification accuracy. As Figs. 2 and 3 show, the proposed model consists of two convolutional layers (C1 and C3) with 6 and 16 feature maps with a kernel size of 3 × 3 followed by two pooling layers (S2 and S4) with 2 × 2 kernel size and then two fully connected layers (F5 and F6) with 120 and 84 feature maps, respectively. An activation function, which is a type of transformation function, is used with a convolution mechanism. Presume $x_k(ii,jj)$ is an input to the neural network's activation function, which is the convolution process's output. The bias vector is b, and the weight vector is w. The following is the description of the activation function:

$$Z(x_k(ii,jj)) = f\left(\sum_{k=1}^{k} x_k(ii,jj).w_k + b_k\right) \Longleftrightarrow Z = f(X \cdot W + b) \tag{2}$$

There are different functions $f(\cdot)$ such as tanh, sigmoid, rectified function, etc. The Rectified Linear Unit function (ReLU) was used in this study as an activation function that was described as:

$$A(x_k(ii,jj)) = \max(0, z(x_k(ii,jj))) \tag{3}$$

In order to compose semantically similar features in one, pooling operations are often used. For carrying out the spatial subsampling, a pooling operation uses a pooling window to consider the maximum or average value. In this work, we applied an average pooling operator with a stride of s_p to the prior activation function output $A(X_k(ii,jj))$, which is calculated as:

$$X_k\ (ii_p, jj_p) = \underset{0 \leqslant ii_p \leqslant h_p - 1, 0 \leqslant jj_p \leqslant w_p - 1}{average} A(x_k(ii,jj)) \tag{4}$$

In the average pooling method, the k-th channel input $((W - w_f)/s_f + 1) \times ((W - w_f)/s_f + 1)$ is reduced to the $\left((W - w_f)/s_{f.sp} + 1\right) \times \left((W - w_f)/s_{f.sp} + 1\right)$. A classifier function was utilized after dense layers to predict class possibility. The most common transformation function is a Softmax classifier for multi-class prediction. The classifier is denoted as a multi-channel patch $\widehat{m}$ of the input data patch n and the ground truth data patch $\widetilde{m}$. Suppose the number of channels for the output data patch $\widehat{m}$ is defined as K and the reshaped LeNet model output is denoted as $_m \times H_m \times K$. Then, the pixel value in the dense layer's output is expressed as $x = [x_1, \ldots, x_k]^T$ and every × is converted into the probability vector $\widehat{m} = [\widehat{m}_1, \ldots, \widehat{m}_k]^T$ based on the Softmax classifier as follows:

$$\widehat{m} = \tag{5}$$

We also utilized the categorical cross-entropy (CCE) loss function that is for multi-class classification to train the proposed network for vegetation classification and grassland maps generation. The CCE loss function is calculated as:

$$L_{CCE}(g, f(o), \theta_1) = \sum_{i=1}^{S} \sum_{j=1}^{P} \sum_{c=1}^{C} -1\left(g_i^j = C\right) \log l_c\left(I_i^j\right) \tag{6}$$

where, the output of the last convolutional layer at the pixel I_i^j is $f\left(I_i^j\right)$, the ground truth label is g_i^j, the number of classes is C, the jth pixel in the ith patch is I_i^j, the size of the batch is S, the number of pixels in each patch is P, the network parameters are δ_1, and each class probability of pixel I_i^j is $l_c\left(I_i^j\right)$ that is denoted as:

$$l_c\left(I_i^j\right) = \frac{\exp\left(f_c\left(I_i^j\right)\right)}{\sum_{l=1}^{c} exp\left(f_l\left(I_i^j\right)\right)}. \tag{7}$$

3.4. Experimental setting

In this study, we used the Adaptive Moment Estimation (Adam) optimizer with a learning rate of 0.001 to optimize the loss function. The presented LeNet network was trained with a batch size of 64, and the trained network was then used to classify vegetation cover and generate grassland maps on test data. A dropout layer of 0.25 (Shi et al., 2018) was added to the network's deeper convolutional layers to overcome the overfitting issue. This strategy can provide a model regularization that is both computationally affordable and efficient. We performed the suggested network with the Keras 2.0 system and Tensorflow backend.

3.5. Evaluation of performance

The measurement variables like precision, recall, F1 score, and overall accuracy (OA) were used to test the proposed method's efficacy, which is shown in equations (8–11). Precision is defined as the proportion of grass pixels that are correctly classified among all anticipated pixels. The ratio of grass pixels that are correctly identified among all actual grass pixels is referred to as recall. OA is the precision of grass and other pixels, and the F1 score is a combination of recall and precision.

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

$$F1\_score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

$$OA = \frac{TP + TN}{N} \tag{11}$$

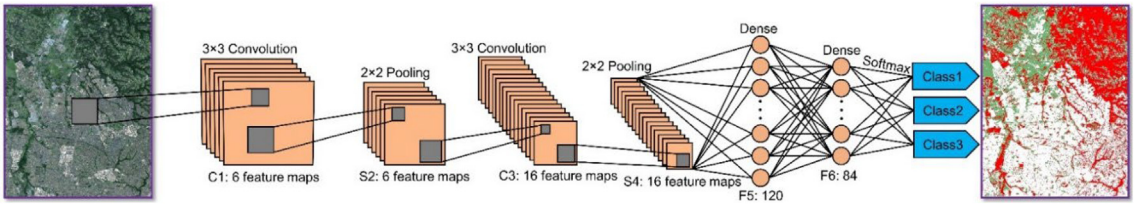

Fig. 2. LeNet network structure for mono-temporal data or base date imagery (04/05/2020) with spectral and spatial values.

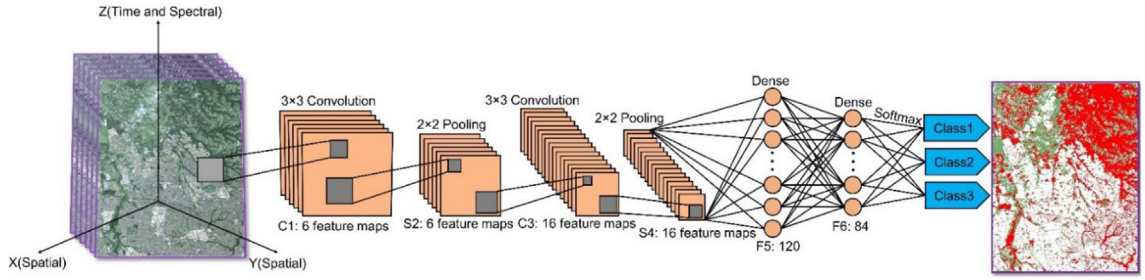

Fig. 3. LeNet network structure for multi-temporal data (short time-series imagery with all the dates), which includes both spectral, spatial and temporal information.

where, FP denotes false-positive pixels, FN denotes false-negative pixels, TP denotes true-positive pixels, TN denotes true-negative pixels, and N denotes the number of pixels.

3.6. Comparing methods

We compared the generated grassland maps by the proposed LeNet network from Sentinel-2 imagery with the pre-existing grassland maps for Australia such as DLCD with 250 m resolution and ALUM with 50 m resolution as well as traditional machine learning methods such as SVM and RF to further investigate the benefit of the presented method and Sentinel-2 data. We compared both quantitative and quantitative products for the grassland maps derived from Sentinel-2 (LeNet, SVM, and RF), DLCD and ALUM and note that the results for the other classification maps (DLCD and ALUM) were taken from the published works, while the presented network was applied for creating grassland maps using experiential Sentinel-2 dataset.

4. Results

We utilized several images from complex backgrounds to validate the presented model for classifying vegetation cover and generating grassland maps from Sentinel-2 images. Fig. 4 depicts the proposed network's vegetation cover classification results based on the base date (mono-temporal data). We selected 04/05/2020 date as the base date because during the month of May, peak vegetation growth is achieved. Also, we applied the models on some other dates (e.g., 19/02/2020 and 23/06/2020) and noticed that the proposed model achieved better results on this date. Fig. 5 illustrates the results of the proposed model for dates 19/02/2020 and 23/06/2020. Additionally, Fig. 5 shows that the model predicted more FPs and other pixels on the images captured for the above dates compared to the images captured on 04/05/2020, which is selected as the base date. In contrast, Fig. 6 illustrates the results for the base date image after adding short time-series images and EVI details. The network could achieve sat-

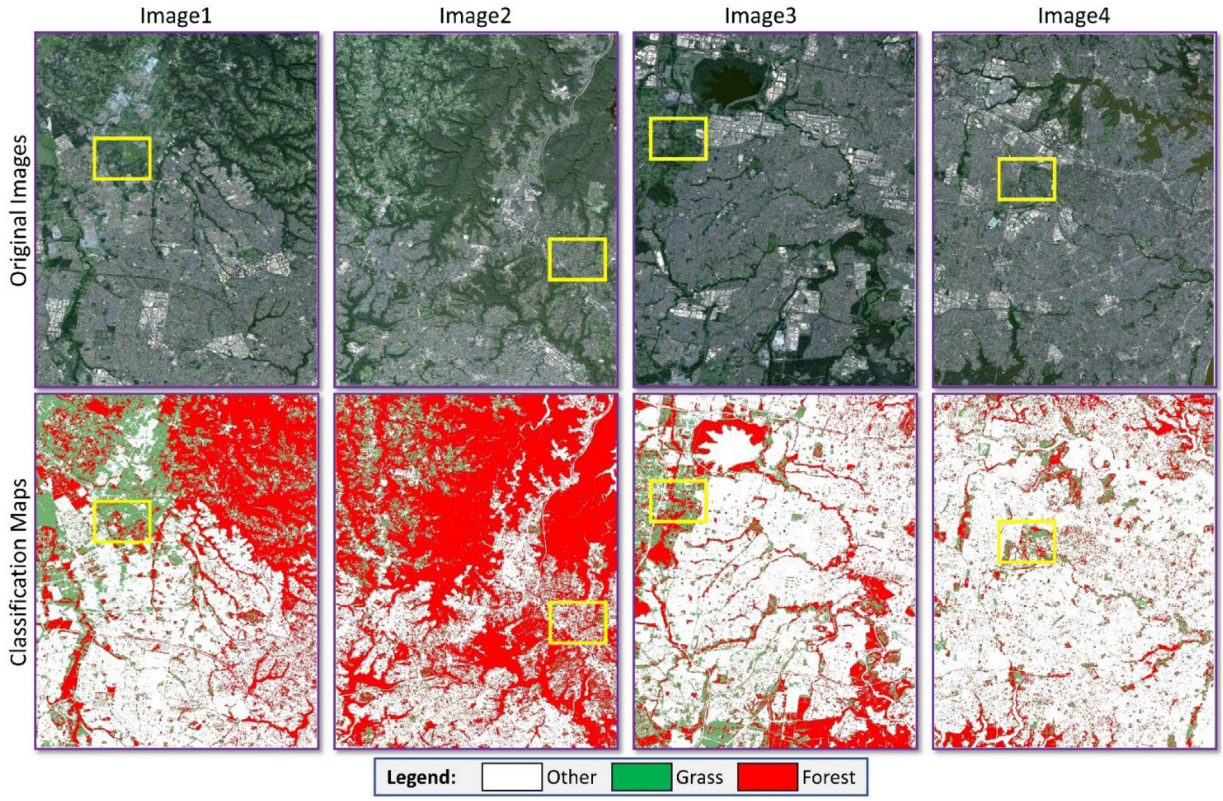

Fig. 4. Qualitative outcomes of the presented approach for vegetation mapping from mono-temporal data or base date imagery (04/05/2020). The first row displays the original imagery, while the second row shows the results for vegetation cover classification with background, grass and forest areas. The yellow box presents the FPs' predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

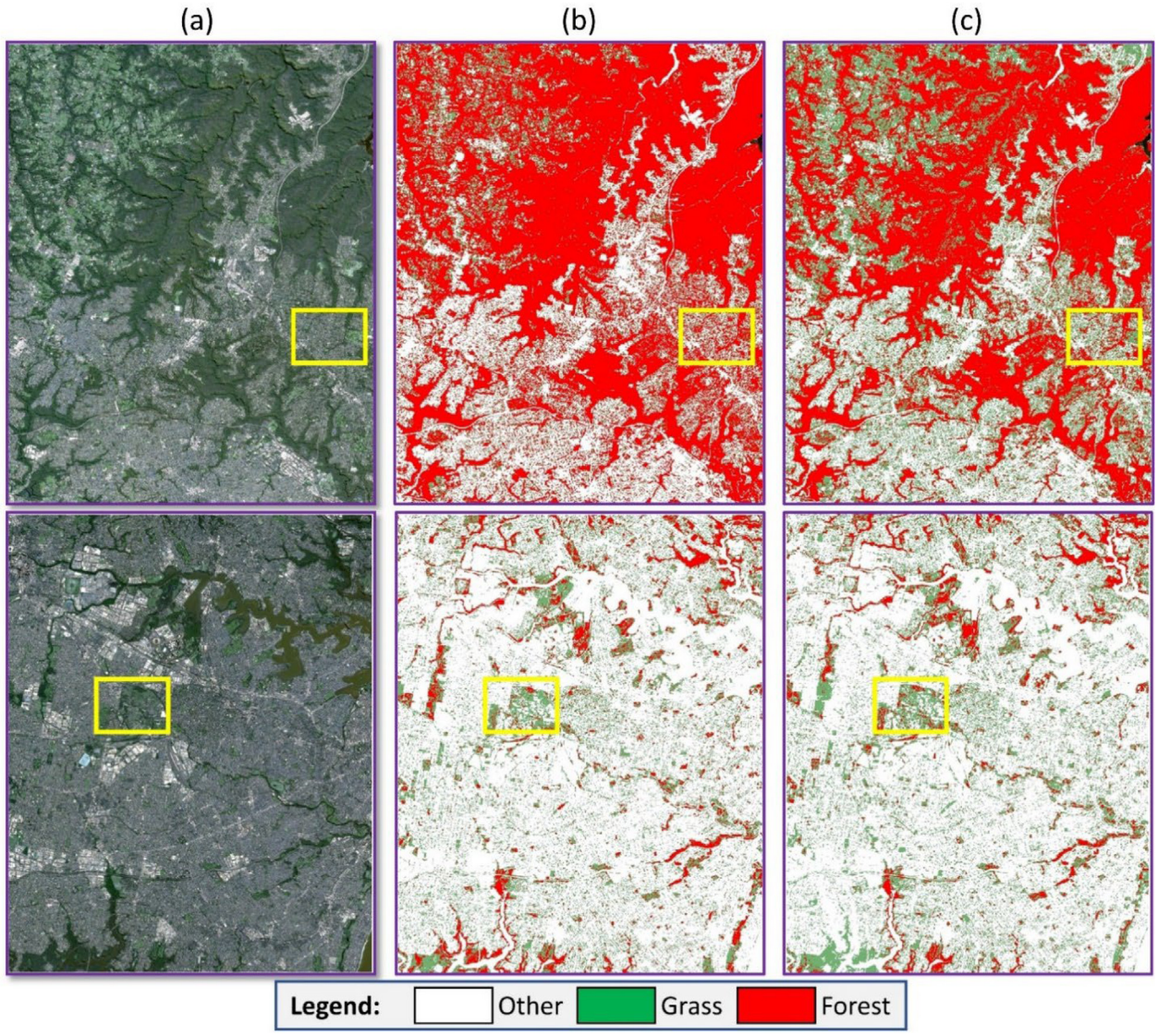

Fig. 5. Qualitative outcomes of the presented approach for vegetation mapping from other mono-temporal data or images. The first row displays the original imagery, while the second and third row show the results for vegetation cover classification with background, grass and forest areas for dates 19/02/2020 and 23/06/2020, respectively. The yellow box presents the FPs' predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

isfactory qualitative results for vegetation cover mapping from both mono-temporal (one date) and multi-temporal (short time-series) images. However, the model obtained more accurate results for multi-temporal data than for mono-temporal data. In fact, the model misclassified pixels (predicted more FPs) that present similar spectral values when we only used base date imagery, which leads to obtaining less accuracy for vegetation mapping. For example, the model misclassified grass pixels as other or forest pixels, specifically for the complex background such as urban areas, where buildings, shadows, etc., cover grass pixels.

Figs. 7 and 8 display the presented approach's performance accuracy on training and validation datasets (several images), using either mono-temporal or multi-temporal training data, respectively. Based on the decrease in model loss and increase in model accuracy over time, the method has learned efficient features for classifying the images into vegetation cover maps with various class labels. In fact, the training and validation accuracy are close together in the learning curve for both mono-temporal and multi-temporal datasets, and the model reduced over-fitting, and the variance of the method is negligible. However, we discovered that adding short time-series Sentinel-2 data (seasonal) to the base date resulted in improving the training and validation accuracy.

Also, we calculated the specified evaluation metrics to assess the quantitative outcomes attained by the presented approach for vegetation mapping. Table 2 presents the average quantitative results on other mono-temporal data (19/02/2020 and 23/06/2020), whereas, Tables 3 and 4 present the average quantitative results for all the images on both mono-temporal (base date) and multi-temporal data, respectively. As shown in Table 2, the average OA achieved by the presented method for classifying vegetation cover maps for date 19/02/2020 and 23/06/2020 is 87.19% and 85.01, respectively, while the model reached the average OA of 88.36% for the base date data. Also, by looking at the results of F1_score achieved by the model for grass areas, it can be seen that the model could improve the results to 5.36% and 8.41% compared to the dates of 19/02/2020 and 23/06/2020, respectively. This confirms that the proposed model showed better results for classification and grassland mapping from the based date images with high vegetation growth. Also, the accuracy of the proposed model

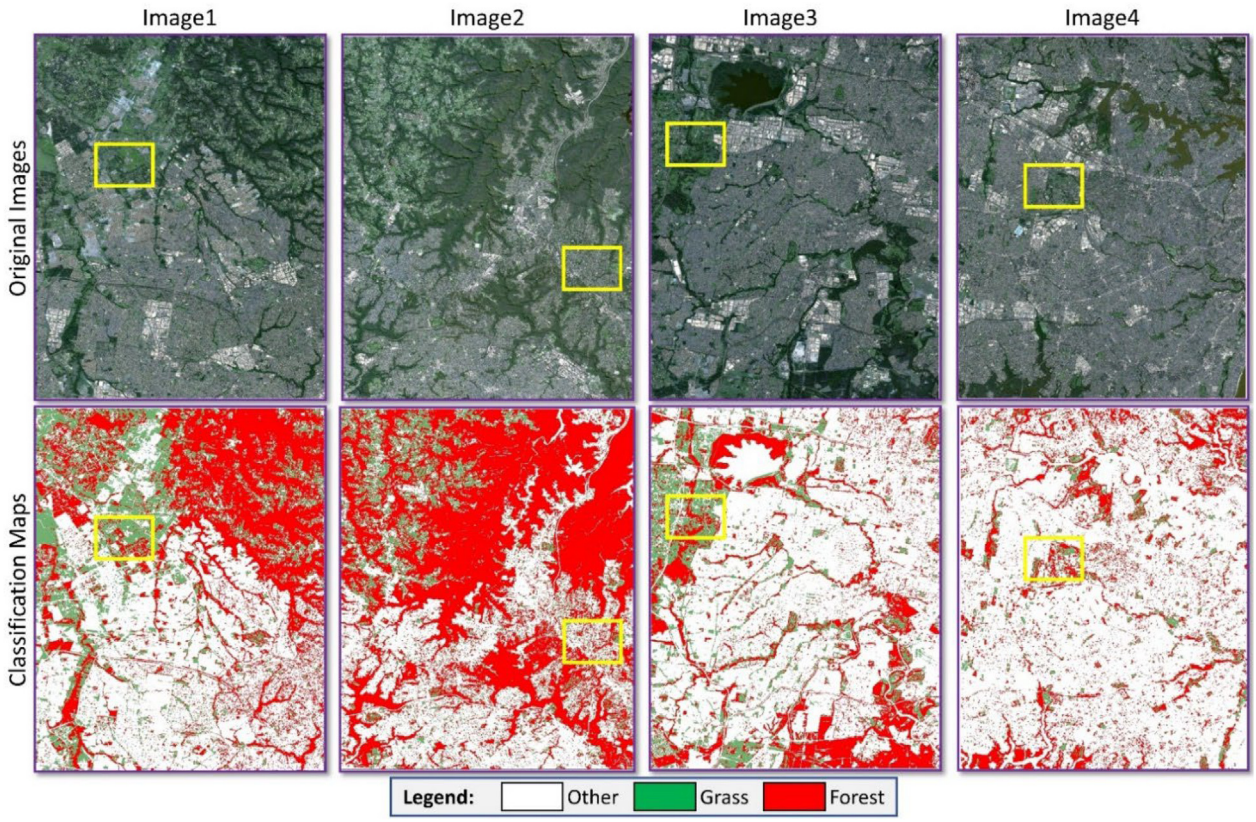

Fig. 6. Qualitative outcomes of the presented approach for vegetation mapping from multi-temporal data (short time-series imagery with all the dates). The original images are shown in the first row, while the results for vegetation cover classification with background, grass and forest areas are illustrated in the second row. The yellow box presents the FPs' predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

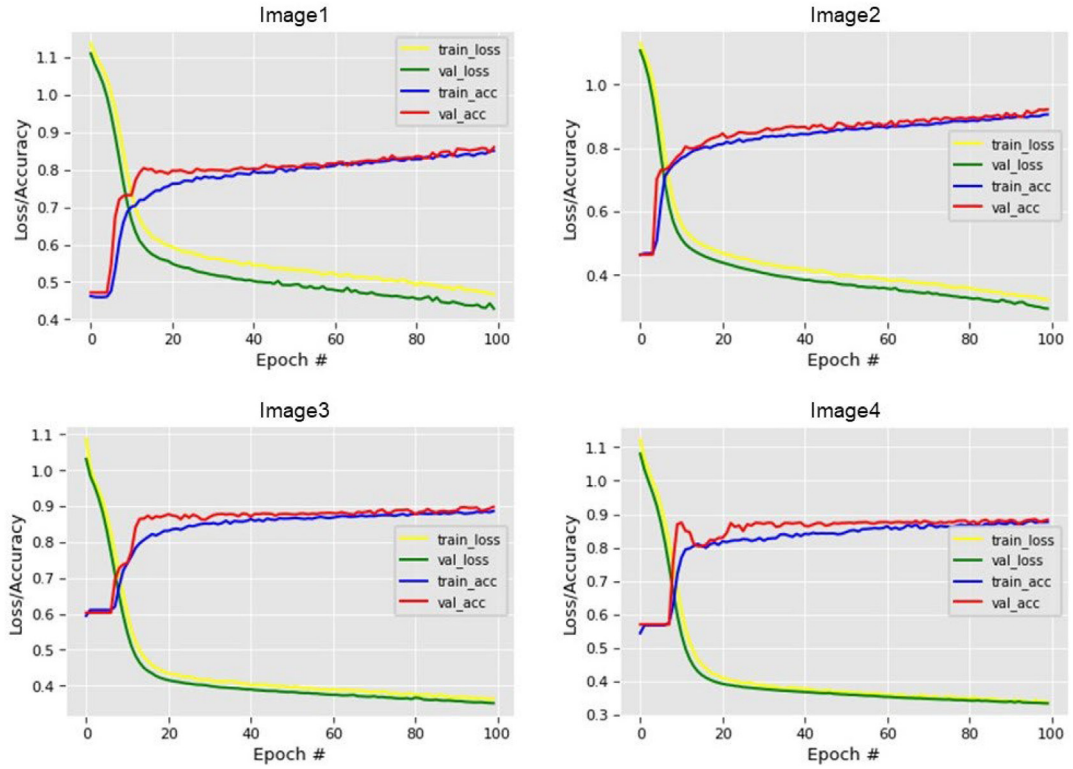

Fig. 7. Vegetation cover classification accuracy on train and validation results with mono-temporal data or base date information (04/05/2020) through training epochs. Images 1, 2, 3 and 4 present the training and validation accuracy/loss for the four Sentinel-2 sub-tiles, respectively.

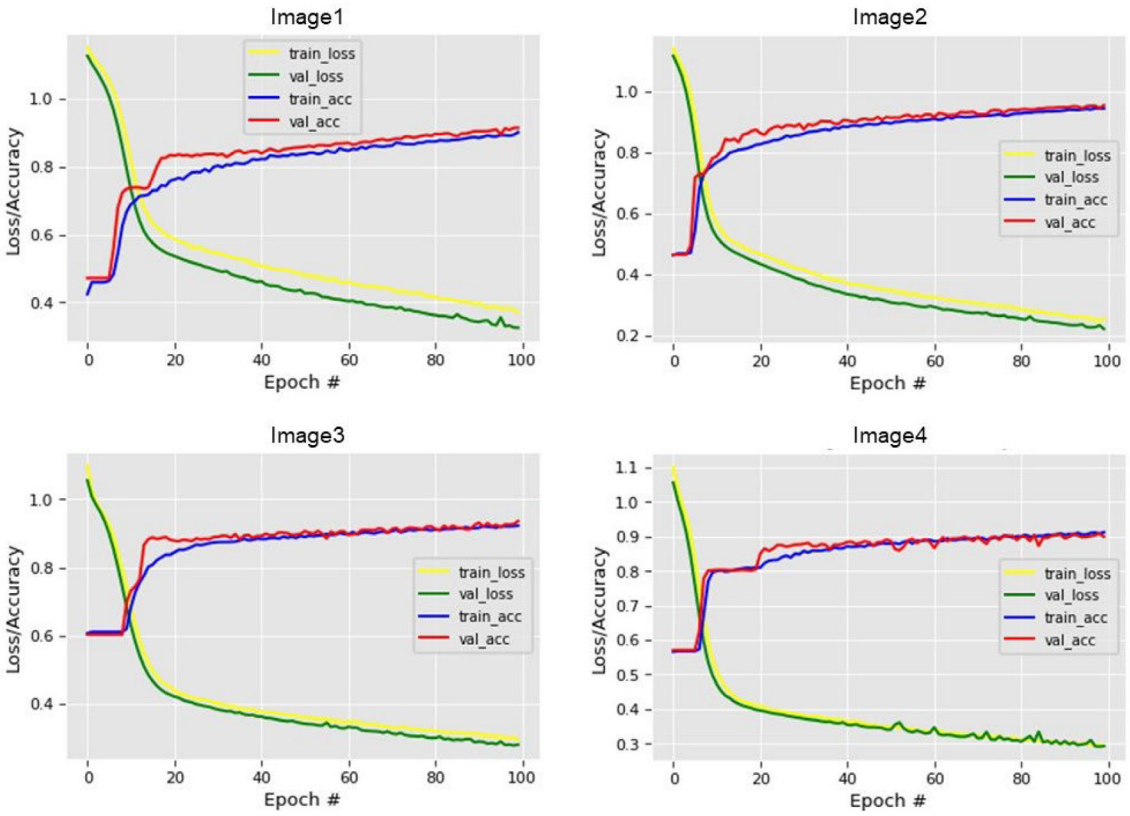

Fig. 8. Vegetation cover classification accuracy on train and validation results with short time-series data (multi-temporal information with all date's information) through training epochs. Images 1, 2, 3 and 4 present the training and validation accuracy/loss for the four Sentinel-2 sub-tiles, respectively.

Table 2
The average percentage of assessment metrics such as Precision, Recall, F1_score and Overall accuracy (OA) attained on testing data sets for other dates.

Dates	Classes	Precision	Recall	F1_score	OA
	Other	98.30	98.67	98.48	**87.19**
19/02/2020	Grass	75	69.96	72.39	
	Forest	68.85	73.52	71.11	
	Other	98.82	98.96	98.89	
23/06/2020	Grass	67.86	70.90	69.34	**85.01**
	Forest	64.04	60.39	62.16	

Table 3
The average percentage of assessment metrics such as Precision, Recall, F1_score and Overall accuracy (OA) attained on testing data sets for mono-temporal data (04/05/2020).

	Precision	Recall	F1_score	OA
Other	97	97.25	97.25	**88.36**
Grass	77.25	79	77.75	
Forest	80	76	77.5	

Table 4
The average percentage of assessment metrics such as Precision, Recall, F1_score and Overall accuracy (OA) attained on testing data sets for multi-temporal data.

	Precision	Recall	F1_score	OA
Other	97.75	97.5	97.75	**92.74**
Grass	85.25	87	86	
Forest	88.5	86.5	87.5	

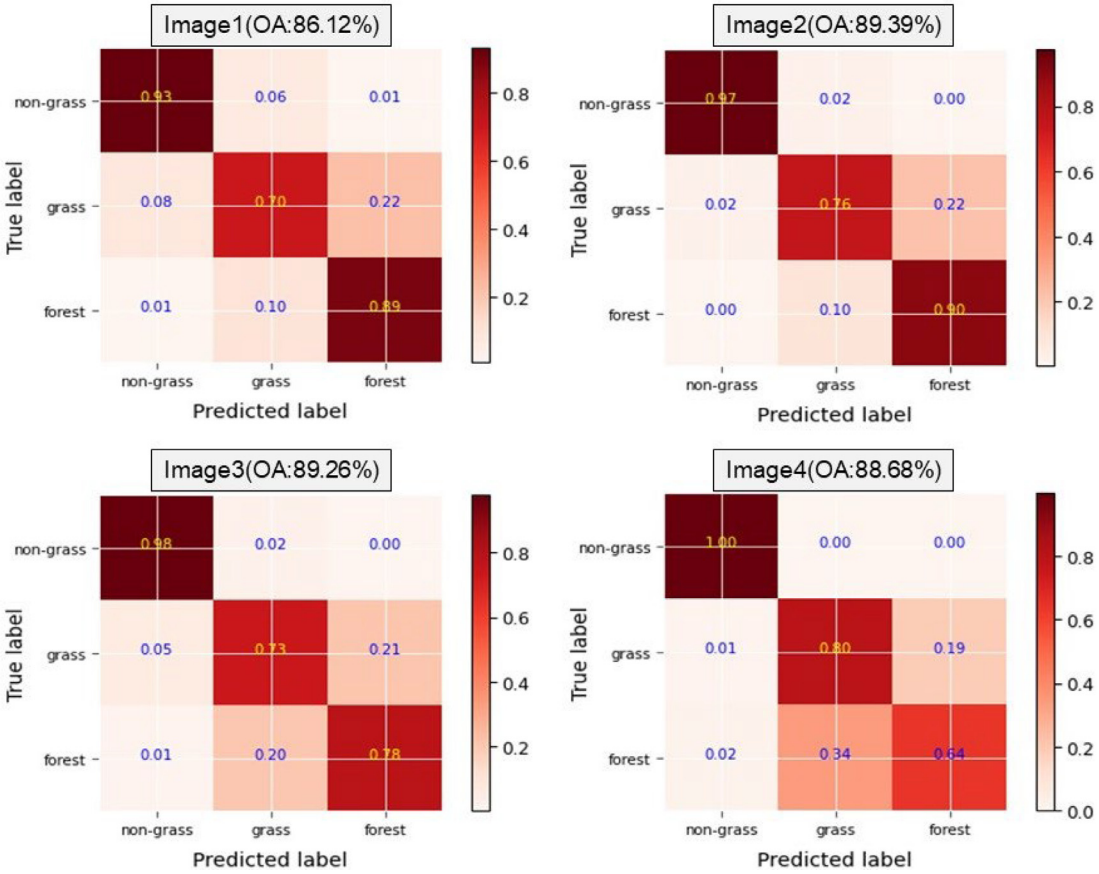

Fig. 9. Confusion matrix for the test dataset used in the proposed LeNet model's training process for mono-temporal data vegetation classification. Images 1, 2, 3 and 4 present the normalized confusion matrix for the four original Sentinel-2 images, respectively.

reached 92.74% for the multi-temporal data. In other words, by adding short time-series images and EVI information to the base date, the accuracy of the classification maps is increased, and the method could improve the accuracy to 4.38% compared to the results achieved for the mono-temporal data. In fact, the suggested technique could identify grass pixels more accurately even under complex backgrounds and vast areas of occlusion.

The accuracy of the proposed LeNet network for vegetation cover classification was assessed using the confusion matrix of the test dataset, which is depicted in Figs. 9 and 10 for both mono-temporal and multi-temporal data, respectively. As it is clear from both figures, the presented technique obtained more accurate results for the confusion matrix on multi-temporal data than for mono-temporal data. In other words, the method could identify the pixels for each class label more effectively when we combined the short time-series images to the base date data, which resulted in obtaining higher OA compared to the mono-temporal images. For instance, for image 1 in Fig. 9, the model could identify only 70% of the grass pixels while predicting 82% of grass pixels for the same image in Fig. 10. This confirms that the addition of short temporal data has a significant effect on increasing the model accuracy for the classification of vegetation maps and the production of high-resolution grassland maps.

5. Discussion

Table 5 illustrates the quantitative outcomes (F1_score) for the grassland maps creation from Sentinel-2, DLCD, and ALUM. The F1_score achieved for the DLCD grassland map is ranked the lowest with 31.1%, while the F1_score for the ALUM grassland map slightly increased to 43.04%. In contrast, the F1_score for the Sentinel-2 grassland map produced by SVM, RF and LeNet is much higher than the DLCD and ALUM. The proposed LeNet model achieved an accuracy of 86% for generating grassland maps from Sentinel-2 data, which could improve the quantitative results to 54.9% and 42.96% compared to the DLCD and ALUM grassland maps, respectively. The conventional machine learning models, such as SVM and RF achieved 80.67% and 82.9% for F1_score, respectively, improving the quantitative results compared to DLCD and ALUM results. However, our proposed LeNet model could obtain better results and increased the F1_score to 5.33% and 3.1% compared to the SVM and RF, respectively, which proved the model's effectiveness for grassland mapping.

Fig. 11 depicts the visual grassland map products derived from DLCD, ALUM, and Sentinel-2 data to assess the efficacy of mixing deep learning-based techniques with Sentinel-2 data for producing high-resolution grassland maps. In addition, Fig. 12 shows the results of the proposed LeNet model with other SVM and RF methods for grassland mapping derived from Sentinel-2 data. It is clear that the proposed LeNet technique could classify grass pixels accurately and generate high-quality grassland maps from Sentinel-2 imagery compared to the DLCD and ALUM maps and the results of SVM and RF. In contrast, the grassland maps derived from DLCD, ALUM. SVM and RF are shown in low-resolution types. This is because most of the grass pixels are not identified precisely, especially for regions like complex urban areas, where the grass pixels are covered by buildings, shadows, and other occlusions. Also, in these maps, the other pixels are classified as grass pixels, and more FPs and less FNs grass pixels are detected. Therefore, very low-resolution grassland maps are produced using those methods com-

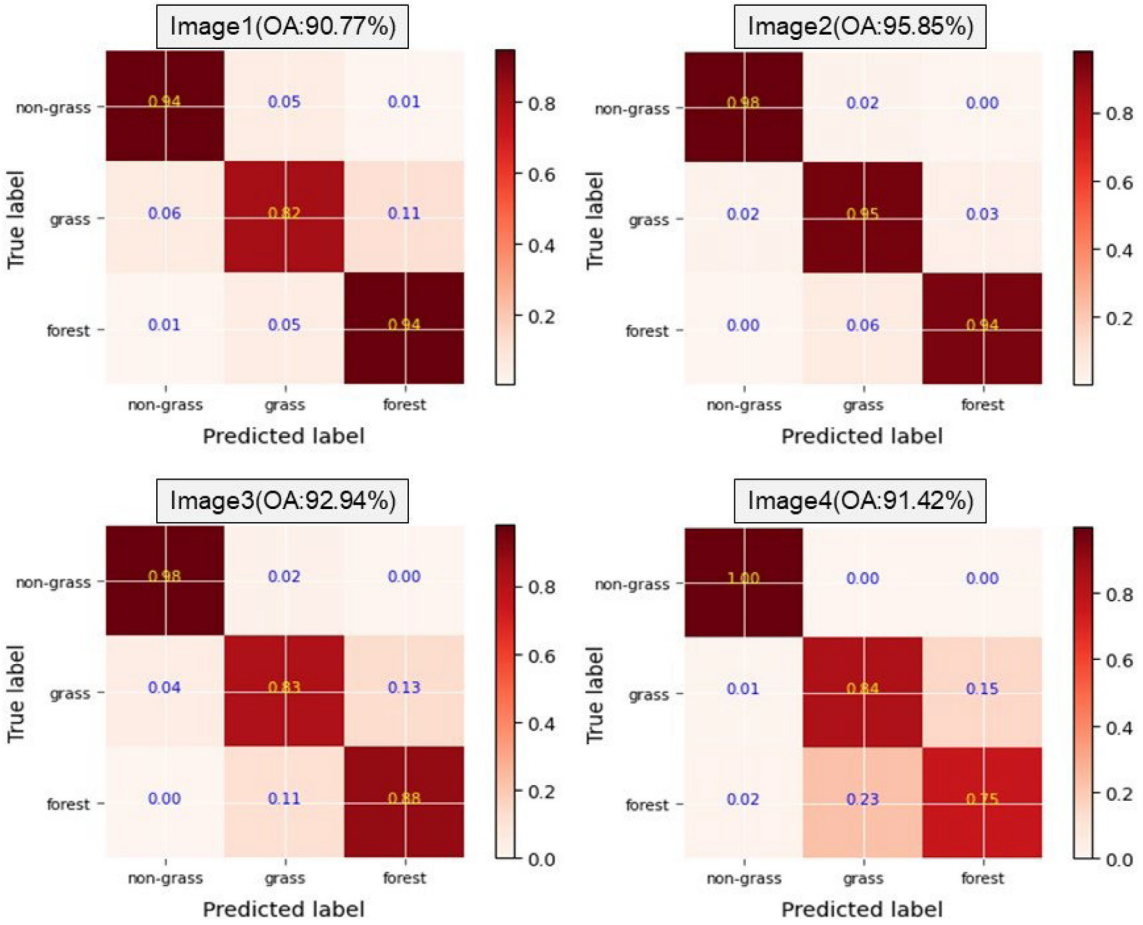

Fig. 10. Confusion matrix for the test dataset used in the proposed LeNet model's training process for short time series (multi-temporal) data vegetation classification. Images 1, 2, 3 and 4 present the normalized confusion matrix for the four original Sentinel-2 images, respectively.

Table 5
Percentage of OA obtained for the grassland maps over testing data sets.

	DLCD	ALUM	SVM	RF	LeNet
F1 score	31.1	43.04	80.67	82.9	86

pared to the proposed LeNet model. By adding more time-series Sentinel-2 data, more information and features about the grass pixels were deduced, which helped the proposed model to classify the vegetation cover better. Also, we used a deep learning model that can encode both spatial and spectral information into a classification scheme, the ability to set parameters and learn the features from data simultaneously and produce high-quality classification maps with satisfactory quantitative results.

5.1. Landsat 8 OLI

We also applied our proposed LeNet model on Landsat 8 OLI data to demonstrate the model's efficiency in vegetation classification and grassland mapping from different remote sensing images. To fairly compare the results, we captured the data for the same region (Sydney region), and the date 07/05/2020 that is the same season as Sentinel-2 based date (04/05/2020) with high vegetation growth. We only used the bands with 30 m spatial resolution (e.g., coastal/aerosol, blue, green, red, Near-IR, SWIR-1, SWIR-2, and Cirrus) that show detailed features of the Earth's surface. We applied a bicubic interpolation (Stoian et al., 2019) to resample 30 m resolution bands to 10 m and achieved a similar spatial resolution like Sentinel-2 data. We applied the proposed model using the same parameters on one image similar to image 3 (Fig. 4) and achieved both quantitative and qualitative results for the Landsat 8 OLI data. In general, the proposed method could attain satisfactory results for classifying the Landsat 8 image and mapping the vegetation covers. However, it obtained an OA of 84.93%, which decreased by 4.33% compared to the same Sentinel-2 image with an accuracy of 89.26%. In addition, the model classified other pixels as grass and forest pixels and detected more FPs and fewer FNs for forest and grass pixels compared with the Sentinel-2 data (Fig. 13). The reason could be due to the resampling of the Landsat data, which results in over-smoothing and blurring or loss of image resolution

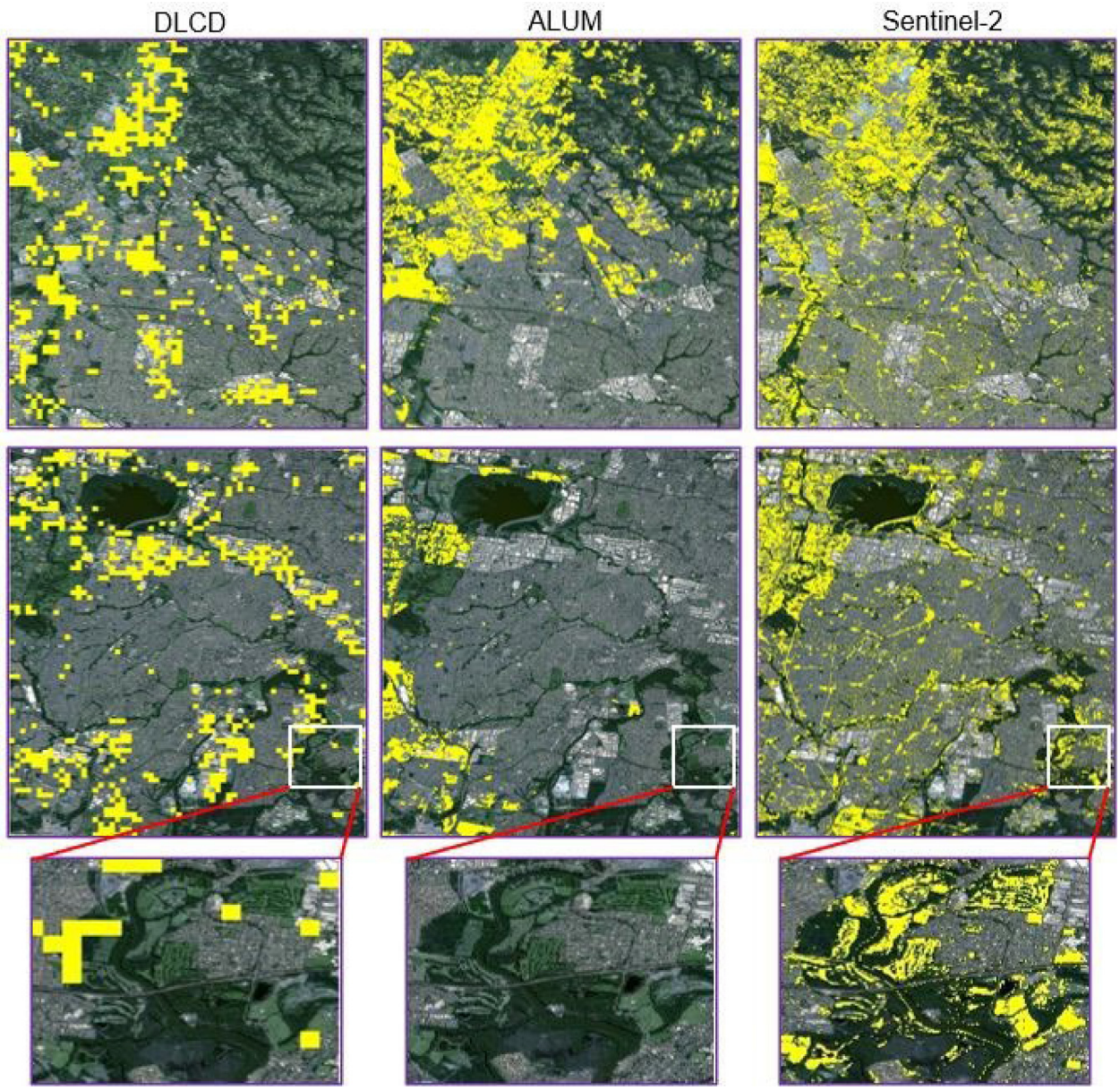

Fig. 11. A comparison of the grassland maps derived from Sentinel-2 imagery against DLCD and ALUM. The zoomed images are shown in the last row to present the differences better.

(Giri and Muhlhausen, 2008). Thus, the model could not classify the image and map the vegetation cover as accurately.

6. Conclusion

In this work, we used Sentinel-2 imagery to classify vegetation and generate grassland maps using a deep learning-based network called the LeNet network. We performed the suggested technique on the base date (non-seasonal) Sentinel-2 images and short time-series (seasonal) data. The results obtained by the presented LeNet network for both mono-temporal and multi-temporal datasets were firstly compared. The proposed model obtained an overall accuracy (OA) of 88.36% for the mono-temporal data and 92.74% for the multi-temporal data. The visualization and quantitative products confirmed that the proposed architecture could improve the qualitative results and the OA (4.38% higher) for vegetation classification maps after adding the short time-series data to the base date data. Besides, we compared the quantitative and visualization products obtained by the LeNet model with traditional SVM and RF methods and pre-existing grassland maps such as ALUM and DLCD maps to demonstrate the efficiency of the proposed model in producing high-resolution grassland maps. All of these findings indicate that the model is an excellent learner for time-series data with high spatial and spectral resolution. As a result, given the availability of Sentinel-2 and Landsat satellite imagery for the region of interest (ROI), this methodology can also be used to determine vegetation cover and produce grassland maps from various remote sensing images at any different locations to confirm its reliability. However, the existence of salt and pepper noise also suggests that, in addition to the spectral and temporal dimensions, the textural dimension of satellite image time-series should be considered. Including the textural dimension is a good avenue for future studies.

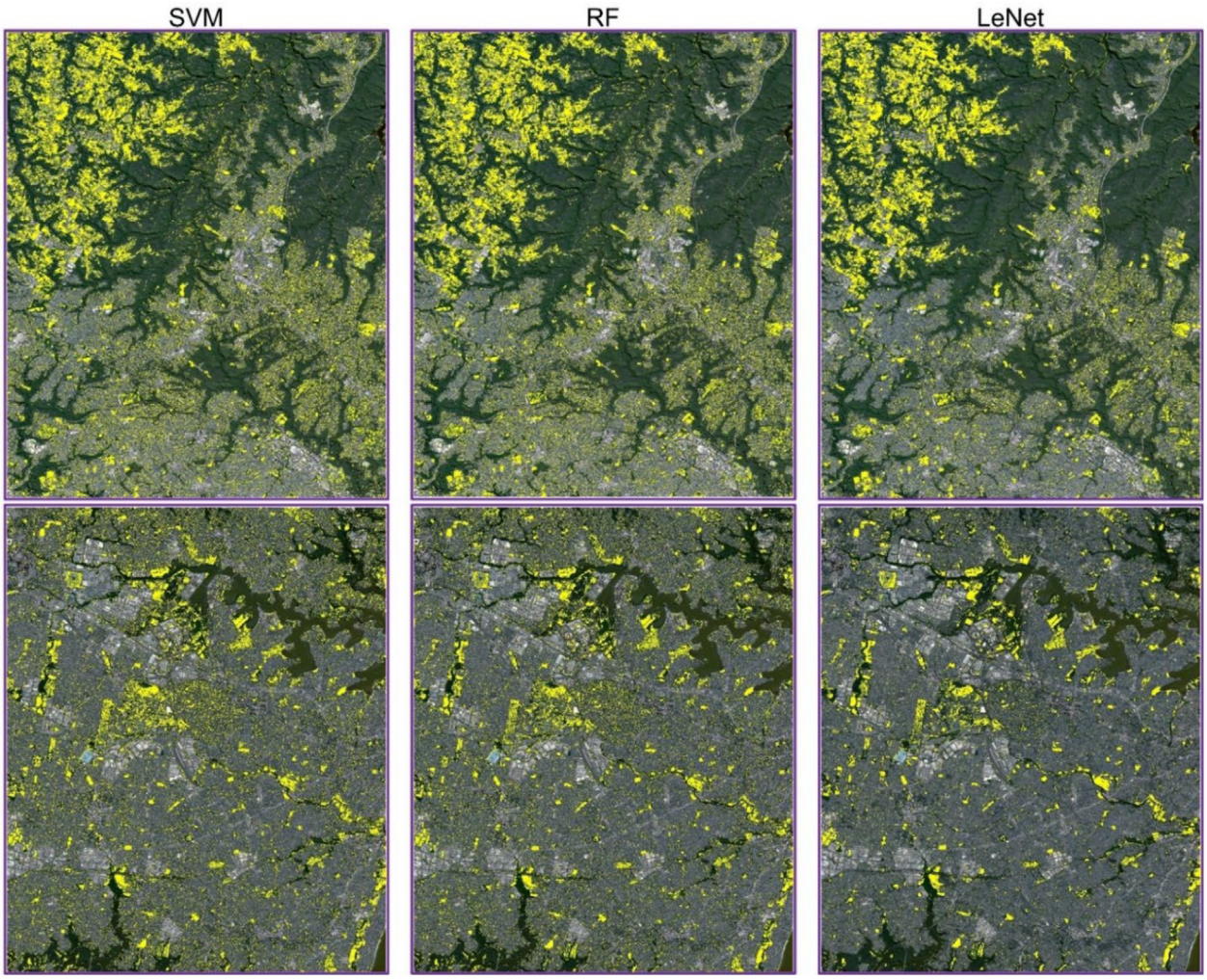

Fig. 12. A comparison of the grassland maps derived from Sentinel-2 imagery using SVM, RF, and LeNet methods.

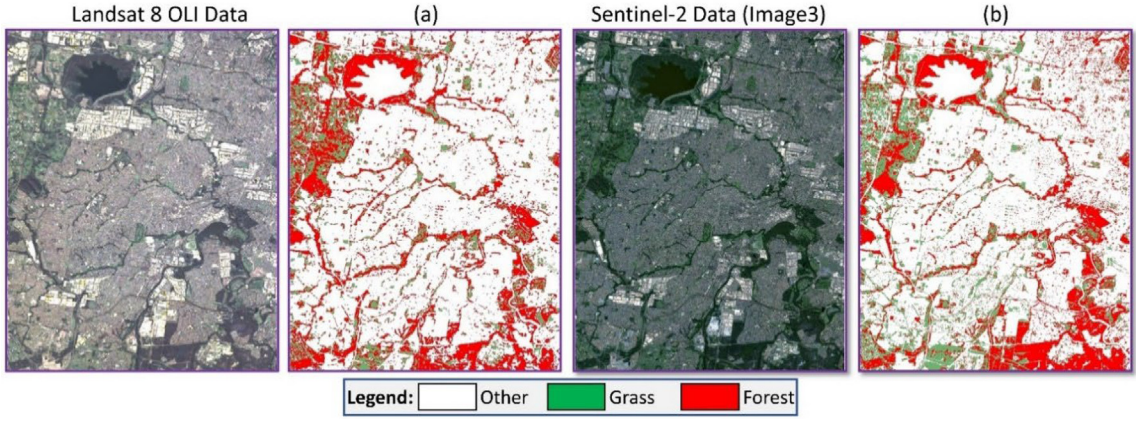

Fig. 13. Qualitative outcomes of the presented approach for vegetation mapping from Sentinel-2 and Landsat 8 OLI data. (a) shows the results of Landsat 8 OLI image and (b) presents the results of the Sentinel-2 image.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Cisneros, A., Fiorio, P., Menezes, P., Pasqualotto, N., Van Wittenberghe, S., Bayma, G., Furlan Nogueira, S., 2020. Mapping Productivity and Essential Biophysical

Parameters of Cultivated Tropical Grasslands from Sentinel-2 Imagery. Agronomy 10, 711.
Esch, T., Metz, A., Marconcini, M., Keil, M., 2014. Combined use of multi-seasonal high and medium resolution satellite imagery for parcel-related mapping of cropland and grassland. Int. J. Appl. Earth Obs. Geoinf. 28, 230–237.
ESA Copernicus Open Access Hub. Available online: (accessed on 10 December 2020).
Feng, Q., Liu, J., Gong, J., 2015. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. Remote Sens. 7, 1074–1094.
Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global consequences of land use. Science 309 (5734), 570–574.
Giri, C., Muhlhausen, J., 2008. Mangrove forest distributions and dynamics in Madagascar (1975–2005). Sensors 8, 2104–2117.
Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: A review. ISPRS J. Photogramm. Remote Sens. 116, 55–72.
Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sens. Environ. 122, 66–74.
Lang, N., Schindler, K., Wegner, J.D., 2019. Country-wide high-resolution vegetation height mapping with Sentinel-2. Remote Sens. Environ. 233, 111347.
Lefebvre, A., Sannier, C., Corpetti, T., 2016. Monitoring urban areas with Sentinel-2A data: Application to the update of the Copernicus high resolution layer imperviousness degree. Remote Sens. 8, 606.
Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28 (5), 823–870.
Naidoo, L., Cho, M.A., Mathieu, R., Asner, G., 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. ISPRS J. Photogramm. Remote Sens. 69, 167–179.
Nijhawan, R., Sharma, H., Sahni, H., Batra, A., 2017. A deep learning hybrid CNN framework approach for vegetation cover mapping using deep features. In: 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, pp. 192–196.
Nogueira, K., Penatti, O.A., Dos Santos, J.A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recogn. Lett. 61, 539–556.
Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., Wilhelm, V.L., 2017. Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery. Eur. J. Remote. Sens. 50 (1), 564–576.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9 (2), 181–199.
Rapinel, S., Mony, C., Lecoq, L., Clement, B., Thomas, A., Hubert-Moy, L., 2019. Evaluation of Sentinel-2 time-series for mapping floodplain grassland plant communities. Remote Sens. Environ. 223, 115–129.
Schuster, C., Schmidt, T., Conrad, C., Kleinschmit, B., Förster, M., 2015. Grassland habitat mapping by intra-annual time series analysis–Comparison of RapidEye and TerraSAR-X satellite data. Int. J. Appl. Earth Obs. Geoinf. 34, 25–34.
Shi, Q., Liu, X., Li, X., 2018. Road detection from remote sensing images by generative adversarial networks. IEEE Access 6, 25486–25494.
Sohn, Y., Rebello, N.S., 2002. Supervised and unsupervised spectral angle classifiers. Photogramm Eng Remote Sensing. 68, 1271–1282.
Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., Mochizuki, K.-I., 2018. Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. J. Appl. Remote Sens. 12 (02), 1.
Stoian, A., Poulain, V., Inglada, J., Poughon, V., Derksen, D., 2019. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. Remote Sens 11, 1986.
Trisurat, Y., Eiumnoh, A., Murai, S., Hussain, M.Z., Shrestha, R.P., 2000. Improvement of tropical vegetation mapping using a remote sensing technique: a case of Khao Yai National Park. Thailand. Int. J. Remote Sens. 21 (10), 2031–2042.
Vasilakos, C., Kavroudakis, D., Georganta, A., 2020. Machine learning classification ensemble of multitemporal Sentinel-2 images: The case of a mixed mediterranean ecosystem. Remote Sens. 12, 2005.
Zhang, C., Xie, Z., 2013. Object-based vegetation mapping in the Kissimmee River watershed using HyMap data and machine learning techniques. Wetlands 33 (2), 233–244.