SEMANTIC SEGMENTATION USING A UNET ARCHITECTURE ON SENTINEL-2 DATA

I. Kotaridis [1] *, M. Lazaridou [1]

[1] Aristotle University of Thessaloniki, Faculty of Engineering, School of Civil Engineering,
Lab. of Photogrammetry - Remote Sensing, 54124 Thessaloniki, Greece - (iskotarid, lazamari)@civil.auth.gr

KEY WORDS: CNNs, UNET, superpixel segmentation, Python, Sentinel-2.

ABSTRACT:

This paper presents the development of a methodological framework, based on deep learning, for the efficient mapping of main land cover classes (built-up, vegetation, barren land, water body) on different urban and suburban landscapes. In particular, the proposed framework integrates the superpixel segmentation (an essential procedure) with deep learning. A combination of spectral bands and indices is introduced to produce optimal results, ensuring adequate discrimination between built-up and barren land classes. A UNET architecture is implemented, which can learn the characteristics of main land cover classes from the input data that can be deployed from a Colab notebook without excessive computational needs. The resulted classifications depict promising accuracy values (above 90%).

1. INTRODUCTION

Global population has increased rapidly over the last century, which triggered transformations of the earth surface increasing the rate of land cover (LC) changes, particularly in urban areas, where more than a half of the global population lives (Addae and Oppelt 2019; Talukdar et al. 2020). Land is a limited resource and cities are continuously expanding. Urban growth can be described as the expansion of built-up areas that implies alterations in land cover of the natural landscape. Urban land cover and land use mapping plays an important role in urban planning and management. Quantifying urban growth is essential to perform an evaluation of its environmental, economic, and social impacts (Bhat et al. 2017; Zhang et al. 2018; Sapena and Ruiz 2019; Addae and Oppelt 2019). The above can be achieved, among others, through thematic information extraction techniques from satellite data.

Several methods have been developed for thematic information extraction from satellite observations that offer a cost-effective, spatially extensive, multi-temporal, and time-saving solution in comparison with traditional field surveys (Talukdar et al. 2020). Pixel-based methods categorize individual pixels mainly based on the spectral information and this may cause "salt-and-pepper" issues due to the fact that spectral responses of individual pixels do not represent the characteristics of the surface object. Over the last two decades, a framework that employs objects as a basis for the analysis has emerged that is called Object-Based Image Analysis (OBIA). Image segmentation is the preliminary and critical step process to produce the fundamental elements of OBIA, that includes the partition of an imagery into spatially adjoining and relatively homogenous regions (segments). These elements form the foundation for further analysis as classification units (Blaschke et al. 2004; Baatz et al. 2008; Nussbaum and Menz 2008; Thenkabail 2015; Cheng and Han 2016; Kotaridis and Lazaridou 2021). However, in a complex urban environment, the selected features cannot be representative of all land cover types. Thus, instead of using manually selected features, automatic feature learning from remote sensing data is valuable (Zhang et al. 2018).

In recent years, more advanced methods for pattern recognition i.e., Deep Learning (DL) architectures contributed to a breakthrough in semantic segmentation of remote sensing imagery (Kotaridis and Lazaridou 2021), assigning every pixel a class label of its corresponding image object (Mi and Chen 2020). The rise of Convolutional Neural Network (CNN) has played a major role towards this direction, emphasizing on automatic feature learning. Satellite image semantic segmentation, including the extraction of roads, buildings, and identification of land cover types, is essential for sustainable development, urban planning, and climate change research. on are based on semantic segmentation task (Wu et al. 2019).

A few researchers have employed satellite data and implemented CNNs and specifically UNETs to carry out image land cover classification tasks (Zhang et al. 2018; McGlinchy et al. 2019; Yi et al. 2019; Soni et al. 2020; Han et al. 2020).

The major objectives of this paper are:

1. The combination of spectral bands and indices to produce optimal results, ensuring adequate discrimination between built-up and barren land classes.
2. The development of a methodological framework, based on deep learning, for the efficient mapping of main land cover classes (built-up, vegetation, barren land, water body) on different urban and suburban landscapes. In particular, the proposed framework integrates the superpixel segmentation (an essential procedure) with deep learning.
3. The implementation of a UNET architecture, which can learn the characteristics of main land cover classes from the input data that can be deployed from a Colab notebook without excessive computational needs.

* Corresponding author

4. The effective integration of the presented methodology in relevant thematic information extraction tasks.

2. MATERIALS AND METHODS

2.1 Study area and satellite data

In this paper, three Sentinel-2 level-2A (Bottom-Of-Atmosphere) corrected reflectance images were obtained. The first imagery concerns the train area (Thessaloniki city). The second and third images include the test areas (two Italian cities, Bari and Genoa). The criteria for the selection of scenes were the high quality of data and the limited cloud coverage. A subset was extracted from each scene for analysis in order to include an urban area of various density values. They comprise of several diverse land cover types including concrete, asphalt, water, vegetation and soil, as presented in natural color composites in Figure 1. The detailed features of these images are provided in Table 1.

Product	Acquisition date	Spatial resolution (m)	Tile
S2B_MSIL2A (Thessaloniki)	2021-08-01	10	T34TFL
S2B_MSIL2A (Bari)	2021-08-17	10	T33TXF
S2B_MSIL2A (Genoa)	2021-08-10	10	T33SVV

Table 1. Optical satellite data used.

2.2 Methodological framework

2.2.1 Tools: For the purpose of this study, Python code was developed and executed on Google Colab (Colaboratory). In specific, the following libraries were used:

- Numpy, a core library for scientific computing, for basic array operations.
- Pyrsgis to read and export GeoTIFFs.
- patchify library to split images into small overlapping patches by given patch and step size, and merge patches into original image during the prediction step.

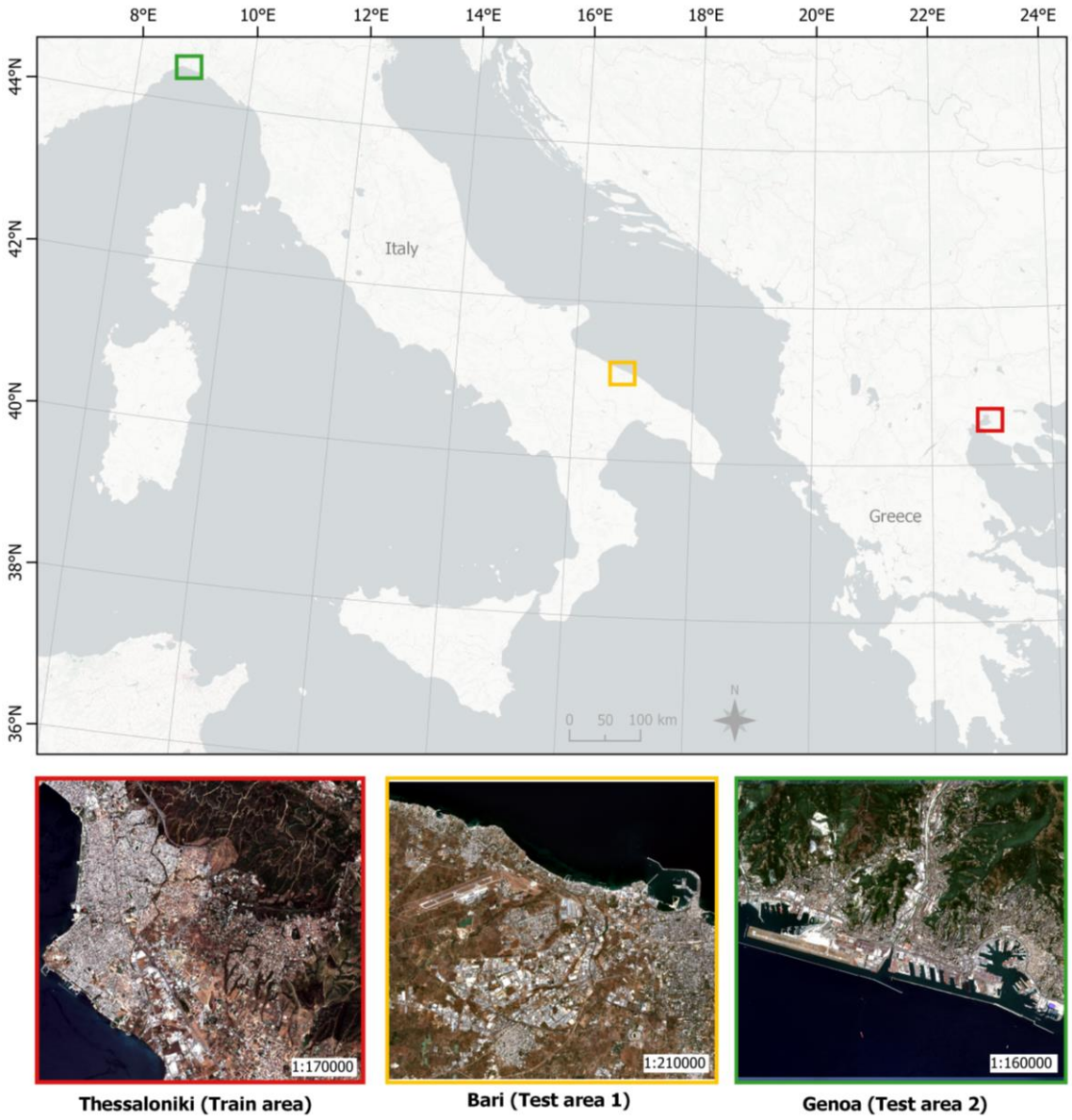

Figure 1. Location map of the train area and test regions and the corresponding Sentinel-2 natural color composite subset images.

- Scikit-learn, machine learning package that offers functionality supporting supervised and unsupervised learning, for data pre-processing and accuracy checks.
- Keras with Tensorflow backend, a machine learning and artificial intelligence framework, for building and deploying the CNN model.
- matplotlib for creating visualizations (plots).

In addition, QGIS and Orfeo Toolbox (OTB) were used to preprocess the data. QGIS is a free and open-source Geographic Information System that supports the creation, editing, visualization, and publication of geospatial data[1]. In specific, it was used for digital processing of the Sentinel-2 images. OTB, an open-source software library that supports processing of remote sensing data[2], was employed for input data normalization. Finally, the validation of the results and the visualization (maps) were also carried out in QGIS.

2.2.2 Training and testing dataset preparation: Dataset preparation produces ready-to-use samples for the UNET model. It can be distinguished into the following discrete phases: initial processing of Sentinel-2 imagery, superpixel segmentation, and data sampling (Figure 3).

Initial processing includes spectral indices calculation, normalization of spectral bands (R, G, B, NIR), stacking of the normalized bands and spectral indices to produce a single image product and clipping this product to the boundaries of the area of interest (AOI).

In this study, the combined use of four common spectral bands (R, G, B, NIR) and three suitable spectral indices (MNDWI, NDVIre, NDTI) is proposed to extract the main land cover classes in a complex and heterogeneous environment, that is, a city and its surrounding areas. Distinguishing barren land from built-up environment is often a difficult task (Osgouei et al. 2019). We deduced from several experiments that the aforementioned input produces the optimal results. Table 2 summarizes the spectral indices and their corresponding equations. Figure 2 clearly illustrates the contribution of spectral indices in terms of spectral separability between the different land cover classes.

Spectral Index	Equation	Sentinel-2 bands
MNDWI	$\frac{Green - SWIR1}{Green + SWIR1}$	$\frac{B3 - B11}{B3 + B11}$
NDVIre	$\frac{Red\ edge\ 1 - Red}{Red\ edge\ 1 + Red}$	$\frac{B5 - B4}{B5 + B4}$
NDTI	$\frac{SWIR\ 1 - SWIR\ 2}{SWIR\ 1 + SWIR\ 2}$	$\frac{B11 - B12}{B11 + B12}$

Table 2. Spectral indices calculated in this study.

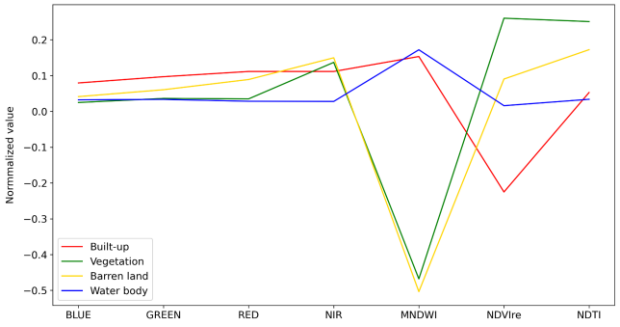

Figure 2. Spectral responses of characteristic segments represented by the mean normalized value.

Another concern during the train/test dataset preparation was to normalize the input features into similar value ranges. Data normalization is critical to ensure that all the features are treated in an equal manner, considering that neural networks are sensitive to the distribution of data (Chollet 2018). We implemented a data normalization (range from 0 to 1) procedure to obtain a land cover probability output. To achieve that, we used the maximum value from each individual spectral band (spectral indices are already normalized). For this purpose, the BandMathX application was accessed through the otbApplication Python module. The normalized spectral bands and indices were then stacked to from a single raster image. Finally, this raster image was clipped to the boundaries of the AOI.

Image segmentation was carried out in Terminus QGIS plugin[3]. Terminus was developed in order to provide an easily accessible tool that allows user to perform image segmentation tasks. It is a fast and straightforward plugin that includes four popular image segmentation algorithms: Felzenszwalb's, quickshift, SLIC and watershed. Each algorithm produces two outputs, a vector file and a raster file with the produced segments. The plugin offers user the option to compute various statistics over each segment. If this is the case, these statistics are included in the fields of the output vector file and the bands of the multiband raster file that is created. This raster file contains the statistics of the pixels within each segment as the output bands. Thus, it can be displayed as a color composite of user's choice.

For the purpose of this study, following several trial-and-error attempts, Felzenszwalb's superpixel segmentation algorithm was employed. Felzenszwalb's method is a graph-based image segmentation algorithm based on pairwise region comparison. It produces a segmentation of a multichannel image using a fast, minimum spanning tree-based clustering on the image grid. An important aspect of this algorithm is the ability to maintain detail in low-variability image areas whereas ignoring detail in high-variability areas (Felzenszwalb and Huttenlocher 2004). In addition, the mean value statistic was included in the form of the bands of the output raster file. Segmentation produced a controlled oversegmentation that is preferable to under-segmentation, since splitting segments a posteriori is a more complicated task than merging them.

[1] https://www.qgis.org/en/site/
[2] https://www.orfeo-toolbox.org/
[3] https://github.com/ikotarid/Terminus

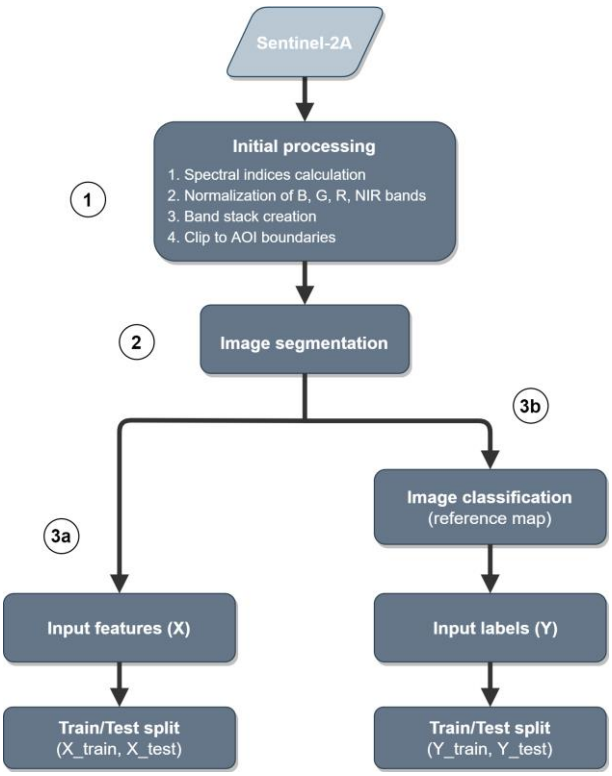

Figure 3. Schematic diagram of data preparation.

A CNN model determines the relationship between characteristics (features) of an entity with a property (label). For this reason, several samples (features with their corresponding labels) are fed into the model and undergo a learning procedure to predict labels for new data (unlabelled data). In this paper, the samples from the segmented stacked image will be referred to as input features (X) and classified land cover data (a reference land cover map was generated) as input labels (Y).

This paper proposes a standard UNET architecture that is based on image patches to perform semantic segmentation. In such an approach, during the training phase, patches are generated from the input features. Once the UNET model is trained with these patches, it is validated on the test patches. We used a 64x64 window with a stride of 64 (window slide) that resulted in 375 non-overlapping patches that were fed to the UNET model. In order to evaluate the performance of the applied model, 70% of them (262) were used as training samples, while the rest 30% (113) were used as test samples.

2.2.3 UNET architecture: A fully convolutional neural network (FCN) replaces the fully connected layers in CNN with up-convolutional layers and concatenates with a shallow, finer layer to produce end-to-end labels (Long et al. 2015). The standard CNN operates in an "image-label" way, while the "end-to-end" labelling mode in FCN is more suitable for pixel-based image classification, i.e., assigning each pixel the label of its corresponding class (Zhang et al. 2018). UNET is an architecture for semantic segmentation. It is an improved FCN model defined by its symmetrical U-shaped architecture consisting of symmetric contracting path (follows the typical architecture of a convolutional network) and expansive path. It combines low level features with detailed spatial information with high level features with semantic information to improve segmentation accuracy (Ronneberger et al. 2015; Zhang et al. 2018). A UNET model (like other CNNs) determines the relationship between characteristics (features) of an entity with a property (label). For this reason, several samples (features with their corresponding labels) are fed into the model and undergo a learning procedure to predict labels for new data (unlabelled data).

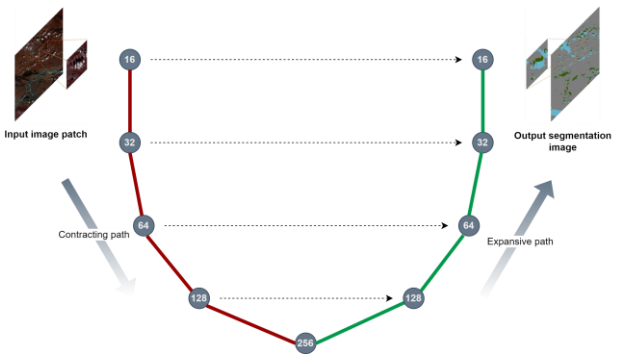

Figure 4. UNET model.

3. EXPERIMENTS

3.1 Computational resources

All source codes of the procedures previously described were seamlessly implemented in Python using the Keras framework (https://keras.io) via TensorFlow (serves as a backend engine) on GPU. Keras is a deep-learning framework for Python that provides a convenient way to define and train almost any kind of deep-learning model (Chollet 2018). The experimental results were produced on Google Colab. It allows user to write and execute Python scripts in the browser together with explanatory text in a single document (notebook). An important feature is the capability to import data from Google Drive as well as from Github.

3.2 Model performance evaluation

The proposed methodological framework introduced in section 2 was employed in the area of Thessaloniki for training, while experiments were conducted in two other areas to investigate its potential in land cover classification for urban growth monitoring. During the validation phase the performance of the trained UNET is examined on the corresponding test dataset. The model was trained for 50 epochs. A few common error metrics regarding the validation of the model are presented in Table 3 and the produced learning curves are presented in Figures 5 & 6 for both training and validation phase.

	1	2	3	4
Precision	93%	97%	86%	100%
Recall	90%	96%	90%	100%
F1-score	91%	97%	88%	100%
Accuracy	**94%**			
IoU	**88%**			

Table 3. Evaluation of the of the UNET model through a few common metrics with land cover types indicated as follows: 1 = built-up, 2 = vegetation, 3 = barren land, 4 = water body.

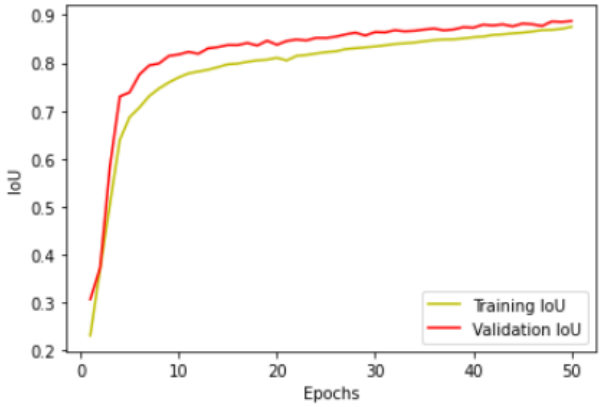

Figure 5. Training and validation learning curves of IoU.

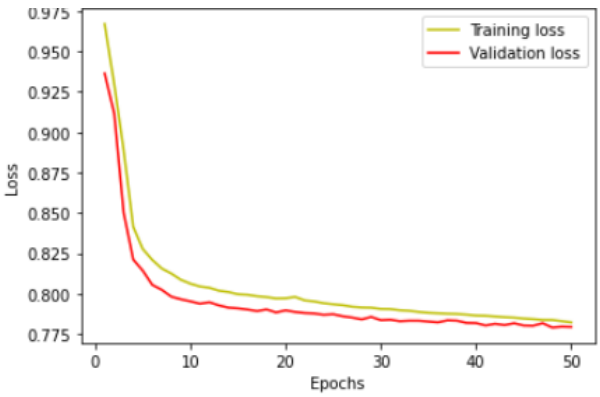

Figure 6. Training and validation learning curves of loss.

3.3 Results

Following the successful training of the proposed model, it was applied to the test areas to produce a land cover map for each of them (Figure 7 & 8). It has to be underlined that the testing material has to undergo the same preprocessing steps as the input training data (i.e., steps 1 and 2 in Figure 3).

4. DISCUSSION AND CONCLUSION

In this research, a promising land cover classification method based on deep learning is proposed for Sentinel-2 data. The method combines superpixel segmentation with deep learning on input data that include spectral indices to distinguish built-up environment from barren land and overcome traditional pixel-based limitations. The imagery is first segmented into superpixels using Felzenszwalb's algorithm, and then a UNET architecture is employed, which can extract land cover features. The proposed framework was validated in two coastal cities on the Mediterranean Sea, and performed quite well depicting high accuracy values for an improved classification of main land cover classes.

In Genoa, built-up land cover class covers an area of 21.2 km^2 (24 %), vegetation 30.1 km^2 (34 %), barren land 3.5 km^2 (4 %), and water body 34.2 km^2 (39 %) from a total of 88.9 km^2. In Bari, built-up land cover class covers an area of 47.6 km^2 (30 %), vegetation 10.6 km^2 (7 %), barren land 59.3 km^2 (38 %), and water body 40.5 km^2 (26 %) from a total of 158.1 km^2.

In the proposed land cover classification framework, superpixel segmentation was carried out to cluster the raster file from the initial processing of Sentinel-2 imagery into small homogeneous regions. We strongly support that this step is necessary, since it transforms the input data in a way that when the UNET model is applied, the final classification map does not suffer from the common issues of a pixel-based approach.

To our concern, a critical aspect is the feasibility of implementation of the proposed method. Since it has relatively minor computational needs, it can be deployed for similar purposes from Google Colab, executing the code on Google's cloud servers. The setup of the pre-requisite ML libraries is quite easy task, since most of them are already installed. In addition, a comprehensive understanding of machine learning fundamentals is essential for a smooth implementation of the aforementioned procedures.

The proposed methodology can be implemented in urban growth monitoring, ensuring an automated procedure on multitemporal imagery. This will be included in a future research work. In due time, the Google Colab notebooks will be available in the authors' GitHub repository[4].

[4] https://github.com/ikotarid?tab=repositories

Figure 7. Genoa natural color composite (a), segmented natural color composite (b), and classification result (c).

		Produced labels				Total	User's accuracy
		1	**2**	**3**	**4**		
Reference labels	**1**	25	0	0	0	25	100%
	2	1	24	0	0	25	96%
	3	5	2	18	0	25	72%
	4	0	0	0	25	25	100%
Total		31	26	18	25	100	
Producer's accuracy		81%	92%	100%	100%		
Overall accuracy		**92%**					

Table 4. The error matrix used to report accuracy results in Genoa with land cover types indicated as follows: 1 = built-up, 2 = vegetation, 3 = barren land, 4 = water body.

Figure 8. Bari natural color composite (a), segmented natural color composite (b), and classification result (c).

		Produced labels				Total	User's accuracy
		1	**2**	**3**	**4**		
Reference labels	**1**	24	0	1	0	25	96%
	2	2	23	0	0	25	92%
	3	0	0	25	0	25	100%
	4	0	0	0	25	25	100%
Total		26	23	26	25	100	
Producer's accuracy		92%	100%	96%	100%		
Overall accuracy		**97%**					

Table 5. The error matrix used to report accuracy results in Bari with land cover types indicated as follows: 1 = built-up, 2 = vegetation, 3 = barren land, 4 = water body.

REFERENCES

Addae B, Oppelt N. 2019. Land-Use/Land-Cover Change Analysis and Urban Growth Modelling in the Greater Accra Metropolitan Area (GAMA), Ghana. Urban Sci. 3(1):26. https://doi.org/10.3390/urbansci3010026

Baatz M, Hoffmann C, Willhauck G. 2008. Progressing from object-based to object-oriented image analysis. In: Blaschke T, Lang S, Hay GJ, editors. Object-Based Image Anal Spat Concepts Knowl-Driven Remote Sens Appl [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 29–42. https://doi.org/10.1007/978-3-540-77058-9_2

Bhat PA, Shafiq M ul, Mir AA, Ahmed P. 2017. Urban sprawl and its impact on landuse/land cover dynamics of Dehradun City, India. Int J Sustain Built Environ. 6(2):513–521. https://doi.org/10.1016/j.ijsbe.2017.10.003

Blaschke T, Burnett C, Pekkarinen A. 2004. Image Segmentation Methods for Object-based Analysis and Classification. In: Jong SMD, Meer FDV, editors. Remote Sens Image Anal Spat Domain. Vol. 5. Dordrecht: Springer; p. 211–236. https://doi.org/10.1007/1-4020-2560-2_12

Blaschke T, Hay GJ, Kelly M, Lang S, Hofmann P, Addink E, Queiroz Feitosa R, van der Meer F, van der Werff H, van Coillie F, Tiede D. 2014. Geographic Object-Based Image Analysis – Towards a new paradigm. ISPRS J Photogramm Remote Sens. 87:180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

Cheng G, Han J. 2016. A survey on object detection in optical remote sensing images. ISPRS J Photogramm Remote Sens. 117:11–28. https://doi.org/10.1016/j.isprsjprs.2016.03.014

Chollet F. 2018. Deep learning with Python. Shelter Island, New York: Manning Publications Co.

Del Rosso MP, Sebastianelli A, Ullo SL, editors. 2021. Artificial Intelligence Applied to Satellite-based Remote Sensing Data for Earth Observation [Internet]. [place unknown]: Institution of Engineering and Technology; [accessed 2021 Nov 4]. https://doi.org/10.1049/PBTE098E

Felzenszwalb PF, Huttenlocher DP. 2004. Efficient Graph-Based Image Segmentation. Int J Comput Vis. 59(2):167–181. https://doi.org/10.1023/B:VISI.0000022288.19776.77

Geiss C, Klotz M, Schmitt A, Taubenbock H. 2016. Object-Based Morphological Profiles for Classification of Remote Sensing Imagery. IEEE Trans Geosci Remote Sens. 54(10):5952–5963. https://doi.org/10.1109/TGRS.2016.2576978

Han Z, Dian Y, Xia H, Zhou J, Jian Y, Yao C, Wang X, Li Y. 2020. Comparing Fully Deep Convolutional Neural Networks for Land Cover Classification with High-Spatial-Resolution Gaofen-2 Images. ISPRS Int J Geo-Inf. 9(8):478. https://doi.org/10.3390/ijgi9080478

Hay GJ, Castilla G. 2006. Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT). In: Lang S, Blaschke T, Schöpfer E, editors. Proc 1st Int Conf Object-Based Image Anal - Bridg Remote Sens GIS. Vol. XXXVI. Salzburg University, Austria.

Kotaridis I, Lazaridou M. 2021. Remote sensing image segmentation advances: A meta-analysis. ISPRS J Photogramm Remote Sens. 173:309–322. https://doi.org/10.1016/j.isprsjprs.2021.01.020

Long J, Shelhamer E, Darrell T. 2015. Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conf Comput Vis Pattern Recognit CVPR [Internet]. Boston, MA, USA: IEEE; [accessed 2020 Nov 12]; p. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

McGlinchy J, Johnson B, Muller B, Joseph M, Diaz J. 2019. Application of UNet Fully Convolutional Neural Network to Impervious Surface Segmentation in Urban Environment from High Resolution Satellite Imagery. In: IGARSS 2019 - 2019 IEEE Int Geosci Remote Sens Symp. [place unknown]; p. 3915–3918. https://doi.org/10.1109/IGARSS.2019.8900453

Mi L, Chen Z. 2020. Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation. ISPRS J Photogramm Remote Sens. 159:140–152. https://doi.org/10.1016/j.isprsjprs.2019.11.006

Nussbaum S, Menz G. 2008. Object-based image analysis and treaty verification: new approaches in remote sensing - applied to nuclear facilities in Iran. New York, NY: Springer.

Osgouei PE, Kaya S, Sertel E, Alganci U. 2019. Separating Built-Up Areas from Bare Land in Mediterranean Cities Using Sentinel-2A Imagery. Remote Sens. 11(3):345. https://doi.org/10.3390/rs11030345

Ronneberger O, Fischer P, Brox T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv150504597 Cs [Internet]. [accessed 2022 Mar 5]. http://arxiv.org/abs/1505.04597

Sapena M, Ruiz LÁ. 2019. Analysis of land use/land cover spatio-temporal metrics and population dynamics for urban growth characterization. Comput Environ Urban Syst. 73:27–39. https://doi.org/10.1016/j.compenvurbsys.2018.08.001

Soni A, Koner R, Villuri VGK. 2020. M-UNet: Modified U-Net Segmentation Framework with Satellite Imagery. In: Mandal JK, Mukhopadhyay S, editors. Proc Glob AI Congr 2019. Singapore: Springer; p. 47–59. https://doi.org/10.1007/978-981-15-2188-1_4

Talukdar S, Singha P, Mahato S, Shahfahad, Pal S, Liou Y-A, Rahman A. 2020. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. Remote Sens. 12(7):1135. https://doi.org/10.3390/rs12071135

Thenkabail PS, editor. 2015. Remotely Sensed Data Characterization, Classification, and Accuracies [Internet]. Boca Raton, Fl: CRC Press; [accessed 2020 Apr 4]. https://doi.org/10.1201/b19294

Wu M, Zhang C, Liu J, Zhou L, Li X. 2019. Towards Accurate High Resolution Satellite Image Semantic Segmentation. IEEE

Access. 7:55609–55619. https://doi.org/10.1109/ACCESS.2019.2913442

Yi Y, Zhang Z, Zhang W, Zhang C, Li W, Zhao T. 2019. Semantic Segmentation of Urban Buildings from VHR Remote Sensing Imagery Using a Deep Convolutional Neural Network. Remote Sens. 11(15):1774. https://doi.org/10.3390/rs11151774

Zhang P, Ke Y, Zhang Z, Wang M, Li P, Zhang S. 2018. Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery. Sensors. 18(11):3717. https://doi.org/10.3390/s18113717