Land Decoding: A Comparative Study on Image Recognition Using U-Net for Urban Parks

Elif Serdar Yakut[1], Meltem Erdem Kaya[2]

[1]Istanbul Technical University, Istanbul/Turkey · serdarsa@itu.edu.tr
[2]Istanbul Technical University, Istanbul/Turkey · erdemmel@itu.edu.tr

Abstract: Extracting and analyzing landscape features from aerial imagery has great potential for understanding characteristics and change. The level of spatial detail required for design and planning problems can vary depending on the methods and data resolutions. Although broad-scale data may suffice for landscape characterization in ecologically oriented planning, site-specific design and analysis for urban landscapes require fine-scale methods. Given the complexity and loss of data resulting from the translation of 3D spatial data into 2D representations such as land classification maps, the research aimed to formulate a feature extraction approach to recognize the implicit characteristics of urban parks. The U-net algorithm, a detailed semantic segmentation, was applied and compared with two other commonly referenced alternatives with their relevant data types – Random Forest (RF) and Object-Based Image Analyses (OBIA) – in the land classification literature. The selection of the algorithms and data sources was grounded in considering differentiation between landscape character classification methods, which are highly emphasized and easily applicable, and the spatial feature extraction of fine-scale landscapes as an overlooked field. A new urbanized image segmentation approach was adapted to complex landscapes by exploring the possibilities and drawbacks of methods and medium-resolution data. The study showed that the U-Net algorithm can predict in-between areas of urban parks and give more consistent recognition than OBIA for very high-resolution aerial images. In conclusion, using the U-net algorithm for site-specific and theme-based tacit features, such as detailed spatial compositions of underwood textures regarding urban parks, can be extracted.

Keywords: Image recognition, U-Net, random forest, OBIA, urban parks

1 Introduction

Landscape, by definition, is associated with human interactions with land as perceived at eye level (STILGOE 2018). However, since the mid-19th century, this vision began to ascend from the ground and reveal the land on which we live. Today's remote sensing has evolved from early attempts on land observation and photographs from balloons to planes. Due to the fundamental connections within analyzing spatial dynamics and classification of landscape elements for ecological research, image recognition has been used as a land-based decoding medium. With the nature of aerial photography and ecologically oriented design incentives, urban ecology has been restructured since the 1960s. This transition has also profoundly shifted the possibilities and notion of the primary means of landscape planning and design. It can be said that, with the advent of aerial photography and its broad accessibility, the landscape became an increasingly important subject of modern urbanism (WALDHEIM 2005).

The interdependent relationship between ecology and remote sensing has led to the development of pattern mapping as a new method for understanding extensive landscapes. Aerial imagery has made it possible to delineate landscape types and usages, resulting in land classifications (ULBRICHT & HECKENDORFF 1998). Remotely sensed data, such as aerial photography, satellite radar, or infrared images from various sensors, are used for these maps instead of field data collection. Land classifications often contain excessive data for large scales,

Journal of Digital Landscape Architecture, 9-2024, pp. 632-639. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-752-6, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537752058.

such as regional or national, rather than detailed and refined spatial characteristics (DESIMINI & WALDHEIM 2017). Due to objective and time-framed data, land cover classification plays a fundamental role in landscape metrics and statistics regarding spatial analysis. Therefore, understanding the ecological dynamics and actual value of the land is primarily related to accurately recognizing patterns. However, mapping land cover data can be challenging for several reasons, starting with the heterogeneity of landscapes and the resolution of the data source. Given the dispersed land use types varying from natural habitats to built-up areas and spatial complexity due to the detailed level of urbanized settings, landscape feature extraction entails challenging methods and detailed data types. In addition, the resolution of data sources and public accessibility can be listed as another impediment. With the availability of multispectral satellite data since the 70s with Landsat MSS, cost-efficient and time-saving methods have been advantageous for larger areas (SCHNEIDER 2012). Conventional remote sensing and analyzing methods mostly use multispectral bands regarding surface reflectance values to determine land characteristics. In fact, land cover classification research primarily focuses on the basic categorization scheme of vegetation, impervious surfaces, soils, and water bodies with limited subcategories based mostly on 30m-resolution multispectral images' pixel values (RIDD 1995).

Alongside this, various machine learning algorithms and methods have been developed for more detailed classification results for land use and land cover change (LULCC) (WANG et al. 2022). Research has been conducted to tackle the insufficient detail level of satellite images, especially when unfolding the spatial heterogeneity of landscape (M'CLOSKEY & VANDERSYS 2022). Nevertheless, the urban landscapes, mostly including parks, which are the potential focus of today's microclimatic problems, require detailed examination. Built-up areas attain spatially more diverse and scale-dependent characteristics regarding other land cover typologies. The classification tasks can lead to the problem of misclassification due to assumptions of algorithms on spectral homogeneity at pixel scale. SMALL (2003) expresses this situation as "mixed pixel" problematic. Herein, to overcome this, deciphering complex urban features out of aerial images can be addressed by very-high-resolution(<5m) (VHR) data sources (BAN et al. 2010). As much as several VHR data can be outsourced from commercial satellite services like IKONOS and Quickbird, laser scanning can be another alternative way to get detailed land characteristics. Thus, not only these solutions have constraints and data limitations, but they also lack of public availability. Therefore, the article proposes an image recognition method, using publicly available data and tools, for extracting urban park features of analyses that require detailed spatial data in urban landscape research and design. Using the U-net algorithm, the goal is to develop a new approach for decoding the urban parks' tacit and in-between spatial characteristics from aerial images. Correspondingly, Random Forest (RF) and Object-Based Image Analysis (OBIA), widely used as land classification methods in the literature, were employed along with frequently associated tools and data sources for both methods. Central Park and its surrounding areas were analyzed and compared as the sample area for all approaches. These experiments compared to two different machine learning algorithms, pixel-based and object-based image classification, with a deep-learning semantic segmentation algorithm. As an outcome, the feature extraction and spatial unfoldment potential of the U-net algorithm were elaborated upon, along with limitations on publicly accessible data types, common methods and tools.

2 Decoding Methods

Urban parks constitute socio-ecological systems that are vibrant, dynamic, and ever-changing. The social and biophysical elements of these systems interact continuously, regulating populations and ecological flows. Considering these spatially dependent dynamics, the microclimatic conditions created by landscape elements, such as tree canopy quality and location, gain importance in terms of urban ecology and sustainability (EGERER et al. 2020). As Forman describes, these spatial characteristics listed as land uses, built structures, anthropogenic flows, and human activities are considered the key attributes of urban ecology (FORMAN 2016). Therewith, urban landscapes generate various unique and unpredictable patchworks of socio-ecological patterns. Spatially disintegrating and extracting these socially constructed and, in a sense, indigenous patches compounds the computability of the ecological value of urban parks. In other respects, detailed spatial examinations and decoding are essential for mitigating the growing effects of urbanization on the climate crisis. Living and built-up environment features are the constituents of microclimatic conditions in urbanized settings. These can occur in various scales from the neighborhood and urban canyon to the shade of a tree (YANG et al. 2023). In fact, as physical factors, landscape elements such as impervious surfaces and vegetation compositions and configurations have great effects on urban heat island accordingly outdoor space comfort conditions of daily life. Herewith, the delineation of urban park features is prominent. In this regard, innovative technologies in spatial decoding and mapping methods play crucial roles.

With the advent of cartographic novelties accompanied by remote sensing, possibilities have escalated for closing the gaps between real and representational grounds. However, decoding the three-dimensional information of the land from a two-dimensional plane is still the main challenge. Therefore, to overcome the mismatched representations and get the accurate spatial delineations, various classification and segmentation methodologies have been in use (MOSTAFAVI 2017). There are various image classification and segmentation methods, which are mostly based on large-scale landscapes, the RF and OBIA stand forward among them. Additionally, it can be seen that the U-Net algorithm is evident as an emerging potential in multidisciplinary literature, regarding semantic segmentation (CLARK et al. 2023, SIDDIQUE et al. 2021). As within the scope of this research, these leading and innovative methods were tested in the iconic Central Park, which is a crucial precedent of urban park literature in landscape design. It was chosen as a pilot area due to its diverse natural and structural elements, ranging from ponds and vast grasslands to dense settlement patterns around transportation routes. The location allowed for the comparison of individual landscape elements within the dense urban settlement, such as free-standing trees, and structural elements within the natural landscape features, such as canopied roads. From this point of view, it was aimed to extract the fundamental features such as tree coverage, undergrowth vegetation, soil and sand, grassland, pavements and roads, built structures, and water surfaces. These elements were defined in various labeling strategies. Consequently, a new segmentation approach was structured via tests on RF and OBIA, regarding deciphering the latent spatial characteristics of the urban parks. The aim is to unfold the dense spatial characteristics that are compacted and implicit in a two-dimensional representation. Given assumptions, they were defined as in-between spaces that cannot be recognized from aerial imagery, like tree-covered roads or wooded lands with dense undergrowth vegetation. Subsequently, with new features, an upgraded labeling system was trained and tested using the U-net algorithm.

2.1 Random Forest

Among other classification methods, RF has been favored due to its prediction accuracy and interpretability of complex data (ZHANG & YANG 2020). As a machine learning algorithm, it comprised of ensemble decision trees that randomly select features to boost the accuracy of classification or regression (BREIMAN 2001, FAWAGREH et al. 2014). It is widely applied to land-cover classification studies (GHOSH et al. 2014, SCHNEIDER 2012) associated with especially medium resolution satellite imageries. To this basis, the RF model was applied with the Landsat 9, which is the newest and publicly available satellite imagery with medium-resolution data. The Google Earth Engine interface which is widely and easily applicable for accessing satellite data and image-based analyses, was used for data acquisition and model training for classification. Firstly, training samples were obtained from a minimum of 60 different points for each label of water, land/sand, grassland, structures and tree selected based on the spatial detail of image. In order to increase pixel-based classification prediction accuracy total sample data was above 400 spots were selected. 80% of these data sets were used as training data, and 20% was used for validation. Using training data links and random forest classifier code references of the Earth Engine, pattern-based classification was generated. Finally, confusion matrix and overall accuracy was calculated.

2.2 Object-Base Image Analysis

Even though, pixel-based methods are highly referential methods in landscape design and planning literature, object-based image segmentation methods promise more accurate results (ZAKI et al. 2022). Despite the pixel-based classifications, OBIA uses not only spectral values but also geographical objects by decomposing an image into logical homogeneous parts. Therefore, it offers a classification based on shape, texture, and other topological elements using more detailed data resolutions (BYUN et al. 2013). Based on these, very high-resolution aerial photography of Central Park was segmented using QGIS 3.32.3 and Orfeo Tool Box 8.1.2, which both are open-source software. The data resolution acquired from Google Earth Pro, dated June 2022, was approximately 50cm per pixel. By this means of visual quality, it was aimed to analyze the detailed spatial features for urban parks. A similar sampling procedure to the application of the RF model was carried out by defining point features for same label classes using QGIS, considering sample size and training and testing ratio. Then, a series of statistical calculations were conducted using the Orfeo Tool Box as in arial segmentation of the image, zonal statistics of segmentation, training model with sample data and respective areal statistics, and finally vector classifier prediction. As final step, confusion matrix and overall accuracy calculated using 20% of sampling points.

2.3 U-NET Algorithm

As a specific version of convolutional artificial networks, U-net is a deep learning architecture for image segmentation tasks. Given its highly suitable performance on medical images with its emergence in 2015, its field of applications has been expanding (SIDDIQUE et al. 2021). U-Net architecture combines encoder and decoder paths. The encoder is primarily used for the analysis of the image for pixel-based classification by down-sampling. The decoder, on the other hand, is used for the synthesis of the labeled data as features by up-sampling (RONNEBERGER et al. 2015). With these paths, as a deep learning algorithm, U-net analyzes the relationships between real image (aerial image) and its corresponding mask (land classification), unlike RF and OBIA with sample selection approach. Regardless of its origin, this model is widely applicable to contemporary design and planning research, especially for land use classification and aerial image recognition tasks (CLARK et al. 2023, ULMAS & LIIV 2020). Thus, by the scope of the research, the defined labels were adopted as two different segmentation strategies using a Python scripting language. The first attempt was to extract the same urban landscape features using the exact labels with OBIA and Random Forest in order to compare the results. Thereafter, two additional labels were added to unfold the compacted information in 2D images like tree-covered impermeable surfaces and woodlands with densely vegetated undergrowth. In this procedure, the same aerial image resolution with OBIA, which covers the Southern part of Central Park, was used as the data set with labeled masks to predict the whole park. Thanks to its time saving principles, U-net can predict whole area using limited sample data rather than RF and OBIA, and also can predict different areas based on its accuracy. By this means, the 4000x3600 pixel image and its related mask were first augmented by mirroring operations and fragmented into 256x256 pixel patches for training. Both approaches were trained using 80% percent of the 1680 pieced data set (image-mask combinations) for 100 epochs. Consequently, attained outputs were smoothly assembled (CHEVALIER 2017) and results were elaborated as both wholistic images and detailed patches. In the scope of this research, these two approaches were called plain and layered U-net predictions. Finally, Jaccard index (IoU) which is one of the most referred accuracy calculation method for image segmentation has been applied.

3 Results and Comparison

Results showed that very high-resolution data are more promising for urban park feature extraction as anticipated, regarding results of RF. Wholistic predictions revealed that the U-net and OBIA models have more potential for the recognition of urban park features (Fig. 1). At first glance, OBIA indicated more detailed information due to spatial dispersion, even though U-net predictions revealed more accurate feature recognition based on visual valuation. Main roads, built-up structures, and vegetative mass can be classified more easily than OBIA. Considering the statistical accuracy results, which are calculated via various methods based on used tools, RF stands forward at overall 85%. However, algorithm efficiency could not be assessed due to medium-resolution data type in contrast with OBIA and U-net's very-high resolution data. OBIA came in second with 79%, which is followed by the U-Net approaches average as plain 65% and layered 53%. On the other side, U-net training IoU values went over 80% for plain and 70% for layered. Aside from these, both plain and layered U-net approaches draw attention given the pixel representations. Considering 256x256 pixel detailed patches, OBIA predictions were distorted and misleading due to over-division of solid

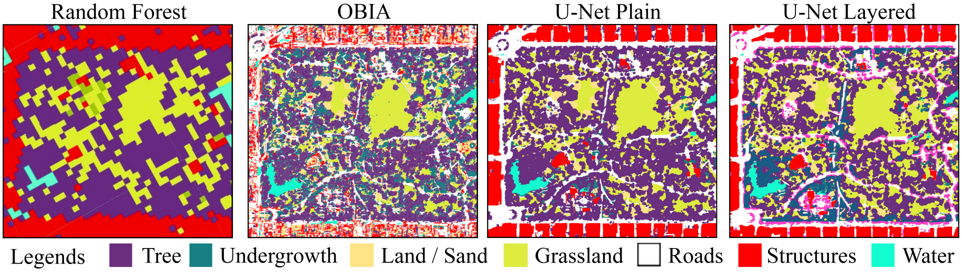

Fig. 1: Comparisons for Random Forest, OBIA, and plain U-NET outcomes for the South side of Central Park

areas such as woodlands or roads. On the other hand, both plain and layered U-net results showed more fidelity to spatial characteristics. In particular, in-between areas of roads can be traced under the canopy with layered results (Fig. 2). Accurate segmentation of rocky surfaces and partial recognition of structures in roads stand out as examples in row a (Fig. 3). Nevertheless, the layered U-net has resulted in excessive segmentation, leading to occasional erroneous conclusions, in contrast to the negligible underestimations of the plain version in row b of Figure 3. This differentiation originates from the training process with two additional land characteristics (tree covered impermeable surfaces and undergrowth) in the same time constraints.

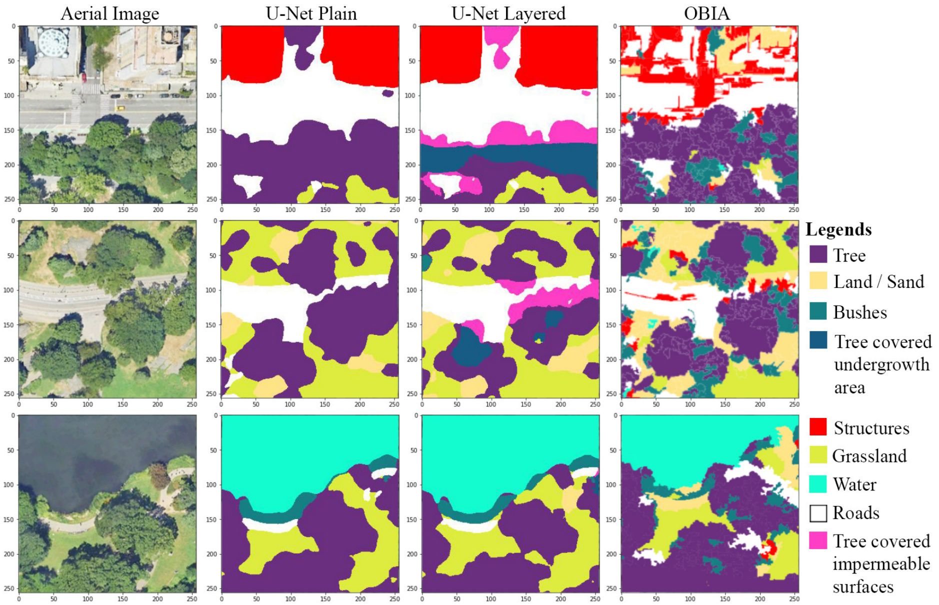

Fig. 2: Detailed prediction comparisons for OBIA with both plain and layered U-Net results on 256x256 pixel results

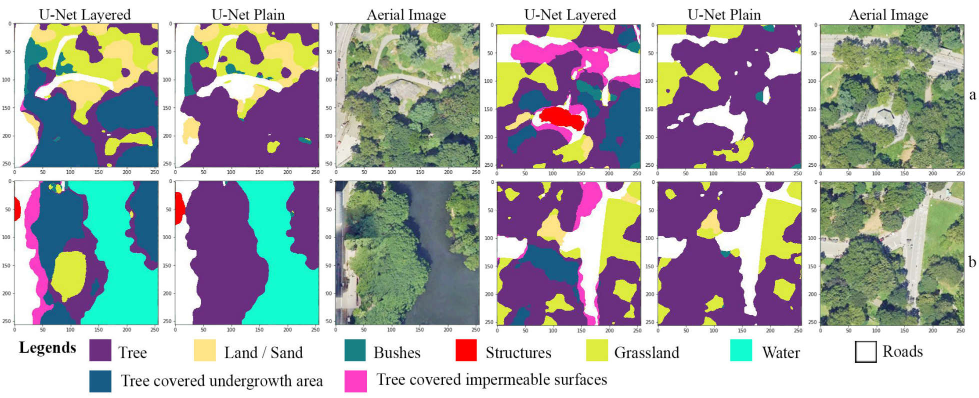

Fig. 3: Consistent (row a) and inaccurate (row b) spatial predictions of layered U-Net results in detailed comparison with plain results on 256x256 pixel results

4 Conclusion and Discussion

The research indicates that the U-net architecture shows promise for detailed land decoding for urban parks, even with very limited data and training efforts. Considering the data sources and methods, RF may serve as an easier to use method with familiar interfaces for land classification on a broad scale, although comparison with other methods may not be effective. On the other hand, significant differences were observed between OBIA and U-net. It is worth highlighting the potential of U-net for the extraction of spatial fidelity based on visual evaluation. It can be said that deep learning models, such as U-net, can perform land recognition tasks more efficiently and accurately in the long run. By increasing the labeled dataset for diverse urban park examples and the epoch time of the model, a well-trained model can be generated for specific design problems, without requiring additional training for each new site. The research results showed a precursor potential for more accurate automated land decoding for various and constantly changing elements of urban parks. For further research, conducting detailed experiments for RF, OBIA, and U-net by increasing the datasets for diverse sample areas with the same data resolution. Research on land decoding and datasets promise great potential for strategic design and analysis approaches for the socio-environmental value of urban landscapes.

Acknowledgment

This article stems from the dissertation entitled *"Form in Landscape as a Field of Probabilistic Decoding"* currently being conducted at Istanbul Technical University. We acknowledge the ongoing support and resources provided by the university for our research.

References

BAN, Y., HU, H. & RANGEL, I. M. (2010), Fusion of Quickbird MS and RADARSAT SAR data for urban land-cover mapping: Object-based and knowledge-based approach. International journal of remote sensing, 31 (6), 1391-1410.

BREIMAN, L. (2001), Random forests. Machine learning, 45, 5-32.

BYUN, Y. G., Han, Y. K. & Chae, T. B. (2013), A multispectral image segmentation approach for object-based image classification of high resolution satellite imagery. KSCE Journal of Civil Engineering, 17, 486-497.

CHEVALIER, G. (2017), Smoothly-Blend-Image-Patches. https://github.com/Smoothly- Vooban/Blend-Image-Patches

CLARK, A., PHINN, S. & SCARTH, P. (2023), Optimised U-Net for Land Use – Land Cover Classification Using Aerial Photography. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 1/2023, 1-23.

DESIMINI, J. & WALDHEIM, C. (2017), Land Classification. In Cartographic Grounds: Projecting the Landscape Imaginary (pp. 112-135). Princeton Architectural Press.

EGERER, M., FOUCH, N., ANDERSON, E. C. & CLARKE, M. (2020), Socio-ecological connectivity differs in magnitude and direction across urban landscapes. Scientific reports, 10 (1), 1-16.

FAWAGREH, K., GABER, M. M. & ELYAN, E. (2014), Random forests: from early developments to recent advancements. Systems Science & Control Engineering: An Open Access Journal, 2 (1), 602-609.

FORMAN, R. T. (2016), Urban ecology principles: are urban ecology and natural area ecology really different? Landscape ecology, 31 (8), 1653-1662.

GHOSH, A., SHARMA, R. & JOSHI, P. (2014), Random forest classification of urban landscape using Landsat archive and ancillary data: Combining seasonal maps with decision level fusion. Applied Geography, 48, 31-41.

M'CLOSKEY, K. & VANDERSYS, K. (2022), Behind-the-Scenes: Multispectral imagery and land cover classification. Journal of Landscape Architecture, 17 (1), 22-37.

MOSTAFAVI, M. (2017), The Cartographic Imagination. In: DESIMINI, J. & WALDHEIM, C. (Eds.), Cartographic Grounds: Projecting the Landscape Imaginary (pp. 6-7). Princeton Architectural Press.

RIDD, M. K. (1995), Exploring a VIS (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. International Journal of Remote Sensing, 16 (12), 2165-2185.

RONNEBERGER, O., FISCHER, P. & BROX, T. (2015), U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234-241.

SCHNEIDER, A. (2012), Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. Remote sensing of Environment, 124, 689-704.

SIDDIQUE, N., PAHEDING, S., ELKIN, C. P. & DEVABHAKTUNI, V. (2021), U-net and its variants for medical image segmentation: A review of theory and applications. IEEE Access.

SMALL, C. (2003), High spatial resolution spectral mixture analysis of urban reflectance. Remote Sensing of Environment, 88 (1-2), 170-186.

STILGOE, J. R. (2018), What is landscape? MIT press.

ULBRICHT, K. & HECKENDORFF, W. (1998). Satellite images for recognition of landscape and landuse changes. ISPRS Journal of Photogrammetry and Remote Sensing, 53 (4), 235-243.

ULMAS, P. & LIIV, I. (2020), Segmentation of satellite imagery using u-net models for land cover classification. arXiv preprint arXiv:2003.02899.

WALDHEIM, C. (2005). Urbanism, Landscape, and the Emergent Aerial Subject. Landscape Architecture in Mutation, ed. Institute for Landscape Architecture (Zurich: ETH Zurich, gta Verlag, 2005), 117-135.

WANG, J., BRETZ, M., DEWAN, M. A. A. & DELAVAR, M. A. (2022), Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. Science of the total environment, 822, 153559.

YANG, S., WANG, L. L., STATHOPOULOS, T. & MAREY, A. M. (2023), Urban microclimate and its impact on built environment – A review. Building and Environment, 110334.

ZAKI, A., BUCHORI, I., SEJATI, A. W. & LIU, Y. (2022), An object-based image analysis in QGIS for image classification and assessment of coastal spatial planning. The Egyptian Journal of Remote Sensing and Space Science, 25 (2), 349-359.

ZHANG, F. & YANG, X. (2020), Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. Remote Sensing of Environment, 251, 112105.