Approach for Automated 3D Vegetation Extraction in Urban Environment using Multispectral Image and Point Cloud Data

ARPITA SINHA[1]

Abstract: Accurate three-dimensional (3D) vegetation extraction, including low-lying plants such as grasses, shrubs, and ground cover, is critical for effective monitoring and management of urban ecological systems. However, precise classification and mapping of 3D vegetation in urban environments, particularly within private and densely constructed areas remains a complex challenge. This study explores the potential of Object-Based Image Analysis (OBIA) for classifying different types of low vegetation using very high spatial resolution imagery. The performance of OBIA was compared across several classifiers, including Support Vector Machine (SVM), Nearest Neighborhood (NN), Maximum Likelihood (ML), and Random Forest (RF). The classification process involved segmenting the image into homogeneous objects using the Segment Mean Shift tool in ArcGIS, followed by supervised classification based on labeled training samples for various vegetation classes, such as grasses, shrubs, trees, hedges, and green roofs. Key parameters, including spectral detail, spatial detail, and minimum segment size, were optimized to improve segmentation results. The combination of spectral detail = 18, spatial detail = 5, and minimum segment size = 5 pixels yielded the best segmentation, balancing spectral and spatial coherence. A total of 1,000 ground truth samples were collected to validate the classification, with results evaluated through a confusion matrix. Grass was classified with the highest accuracy (over 93% user accuracy), while Hedges and Shrubs had the lowest accuracy (often below 60%). SVM performed best for trees (86% producer accuracy), RF for Hedges (75% producer accuracy), and ML for Shrubs (56.5% producer accuracy). Overall, Grass was the easiest to classify, while Hedges and Shrubs were the most challenging. To further enhance accuracy, point cloud data will be integrated to introduce 3D structural information into the classification process. This data provides precise height and surface measurements, offering geometric insights that complement the spectral information. By combining these two data sources, the methodology enables more accurate differentiation of low vegetation types based on both spectral and structural characteristics. The final output will support a Digital Twin platform, offering a detailed urban vegetation database for urban planning, environmental monitoring, and smart city applications. This study demonstrates that OBIA with high-resolution imagery can effectively classify urban greenery, though some challenges remain for certain vegetation types like hedges and shrubs and 3D information from point clouds can most likely assist in a higher classification rate.

1 Introduction

In the era of rapidly expanding cities dominated by concrete, urban greenery acts as a living thread that brings balance and vitality to urban life. It offers numerous benefits, including improved air quality, enhanced public health, reduced carbon emissions, and increased property values (RAUPP et al. 2006; SÆBØ et al. 2012). Effective and sustainable urban planning requires a comprehensive inventory of urban vegetation to evaluate the ecological services it provides. Many European municipalities conduct detailed inventories of public green spaces,

[1] Stuttgart University of Applied Sciences, Schellingstr. 24, D-70174 Stuttgart,
E-Mail: arpita.sinha@hft-stuttgart.de

often focusing on individual trees and recording attributes like species, height, and trunk diameter (MA et al. 2021). However, these efforts are time-consuming, costly, and frequently overlook non-tree vegetation such as shrubs and grasses, limiting a full understanding of urban ecosystems (NEYNS & CANTERS 2022). Moreover, traditional surveys focus largely on public land, leaving private green spaces unmonitored. Though individually small, these areas collectively cover large portions of urban landscapes and play a critical role in providing ecosystem services in dense cities (BAKER et al. 2018; NEYNS & CANTERS 2022). Advances in remote sensing technologies, such as high-resolution satellite imagery and point cloud data, have improved the ability to map vegetation with greater detail and frequency (GUO et al. 2021). Despite progress in modeling man-made features, capturing the complexity of natural elements like trees, shrubs, and lawns remains challenging (MAN et al. 2020). The integration of Light Detection and Ranging (LiDAR) technology marks a major step forward, enabling detailed three-dimensional analyses of vegetation structure and its environmental impact (ZHANG et al. 2023). LiDAR-based point cloud data allows for more frequent, cost-effective, and inclusive vegetation inventories, incorporating features like shrubs and hedges that are often neglected in traditional surveys despite their ecological importance (BURMEISTER et al. 2023). These technological and methodological advances offer immense potential to close existing research and monitoring gaps, enabling more efficient, inclusive, and sustainable urban vegetation management.

2 Materials and Methods

2.1 Data

The dataset used comprises high-resolution multispectral images captured from nadir aerial views at a spatial resolution of 8 cm, acquired in 2023. These images, containing four spectral bands (Red, Green, Blue, and Near-Infrared), were utilized to extract and classify vegetation, particularly low-lying plants and ground cover by leveraging their detailed spectral information. The aerial images set was cordially provided to this project by the City Surveying Office ("Stadtmessungsamt") of the state capital Stuttgart ("Landeshauptstadt Stuttgart") of the state Baden-Württemberg.

2.2 Study Area

The study site is Asemwald, situated on the southern outskirts of the state capital. Asemwald, located in Stuttgart, Germany, has approximate geographic coordinates of 48.7260° N latitude and 9.1932° E longitude This area is characterized by a distinctive combination of high-rise residential buildings and a diverse array of vegetation types, including landscaped green spaces, scattered trees, and patches of natural vegetation. The heterogeneous urban fabric of Asemwald presents a more complex and dynamic environment compared to uniform urban settings. This variability makes it particularly suitable for evaluating and validating models intended for urban applications.

2.3 Process Overview

The main aim of this project is to accurately classify and map different types of vegetation using object-based image analysis (OBIA) applied to high-resolution multispectral aerial imagery captured in nadir view. The object-based approach utilizes spectral analysis in combination with machine learning algorithms, particularly supervised classification, to interpret and categorize the spectral properties of the landscape. In this phase, the focus is on identifying low vegetation. While these methods provide preliminary results, they are limited in terms of the complexity of the data they can process.

Moving forward, the next phase will integrate deep learning techniques, specifically convolutional neural networks (CNNs) to enhance classification accuracy and automate feature extraction. Deep learning models are capable of learning more complex patterns in the data, allowing them to detect subtle variations in the spectral signature of vegetation that traditional methods might miss. CNNs, for example, are particularly good at recognizing spatial hierarchies in imagery, improving the classification of low vegetation with higher precision. Semantic segmentation models further improve the process by classifying each pixel in the image into predefined categories, enabling a more detailed analysis of the vegetation.

Finally, point cloud data will be integrated to add 3D structural information to the classification process. Point cloud data provides precise measurements of the terrain's surface and vegetation height, offering a geometric perspective that complements the spectral information from the multispectral imagery. By combining the spectral data from the aerial imagery with the geometric data from point cloud, the classification process benefits from improved spatial precision. This integration allows for more accurate differentiation between types of low vegetation and better identification of their structural characteristics, such as height and density. Terrestrial multispectral images and derived points clouds can increase the reliability of object detection in areas identified as critical in the aerial image/point cloud approach. Ultimately, combining spectral and geometric properties, results in a more detailed and accurate classification of low vegetation.

2.4 Methodology for the preliminary analysis based on aerial image data only

For the preliminary analysis, five vegetation classes- grasses, shrubs, hedges, trees, and green roofs were selected for classification within an urban environment. Initially, the entire image was classified into various vegetation categories. However, this approach did not provide satisfactory results due to the inference of non-vegetative areas e.g., buildings, roads etc. To improve accuracy, a pre-processing step was introduced where non-vegetative areas were first removed from the image. This was achieved by separating the image into vegetation and non-vegetation areas, effectively masking out non-vegetated regions.

In order to achieve this, the image was initially classified into 36 classes using an unsupervised classification approach. Subsequently, the classification result was reclassified into two broad categories: vegetation and non-vegetation. To isolate the vegetation class, the Con tool in ArcGIS was used, extracting only the pixels corresponding to the vegetation category. These extracted vegetation areas were then converted into polygons using the Raster to Polygon tool. To simplify the dataset, the Dissolve tool was applied to merge all individual polygons into a single unified polygon representing the vegetated areas. Finally, the Clip Raster tool was used to clip the original image, effectively removing all non-vegetated regions and retaining only the areas identified as vegetation. By focusing solely on the vegetated areas, the subsequent classification process became more targeted and efficient. This refinement significantly improved the quality and accuracy of the classification results, as it reduced confusion between vegetated and non-vegetated surfaces and allowed the classifiers to better distinguish between different vegetation types.

The classification process employed an object-based approach using a supervised classification method using the support vector machine classifier, nearest neighborhood classifier, maximum likelihood classifier and random vector classifier. This supervised classification is to categorize all image pixels into distinct vegetation classes, including the low vegetation. As part of the Object-Based Image Analysis (OBIA) approach, the segments were categorized into groups of pixels with similar characteristics, using them as training samples.

The workflow consisted of several key steps, as illustrated in Fig. 1. Firstly, the segmentation process was performed to partition the entire image into multiple homogeneous segments, facilitating object-based analysis using the Segment Mean Shift tool in ArcGIS, incorporating various spectral and spatial details, along with a specified minimum segment size in pixels. Different parameters were tested to achieve the optimal segmentation results. During the segmentation process, a series of parameter combinations were systematically tested to determine the most effective configuration for object-based image analysis. Specifically, different values for spectral detail, spatial detail, and minimum segment size were evaluated to optimize segment delineation based on both spectral and spatial characteristics of the imagery.

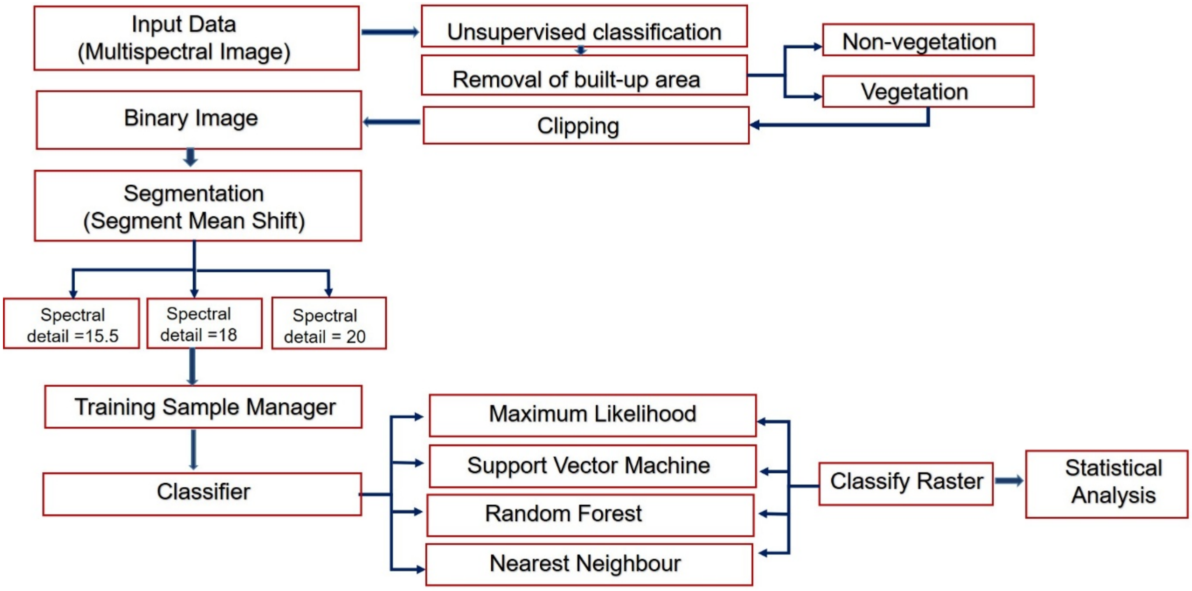

Fig. 1: Methodology for the preliminary analysis based on aerial image data only

For spectral detail, values of 20, 15, and 18 were tested. A spectral detail value of 18 was found to produce the most accurate segmentation results, as it effectively captured variations in spectral reflectance among different vegetation types without causing excessive over-segmentation. Higher values, such as 20, tended to create fragmented segments, while lower values, such as 15, failed to sufficiently differentiate between spectrally similar classes. In terms of spatial detail, values of 15, 10, and 5 were examined. The best performance was achieved with a spatial detail value of 5, which allowed the algorithm to extract fine spatial structures and preserve the shape and boundary of vegetation objects. Higher spatial detail values led to overly generalized segments that failed to represent spatial variability in the landscape. For the minimum segment size, values of 20, 10, and 5 pixels were compared. A minimum segment size of 5 pixels produced the most suitable segmentation outcome by allowing small yet meaningful vegetation features to be captured, while still minimizing noise. Larger minimum segment sizes tended to merge distinct vegetation patches, leading to a loss of detail in the classification. Overall, the selected combination spectral detail = 18, spatial detail = 5, and minimum segment size = 5 pixels provided the most effective segmentation results, offering a balanced representation of spectral heterogeneity and spatial coherence, which is critical for achieving high classification accuracy in object-based approaches.

Training samples were collected for the various vegetation classes to effectively train the classification algorithms. These samples were carefully selected from representative areas of each class, ensuring they accurately reflected the spectral and spatial characteristics of the vegetation types present in the study area. By using these labeled training samples, the classifiers were able to learn the distinguishing features of each class-such as grasses, shrubs, trees, hedges, and green roofs, based on their spectral signatures and object-based properties. The training data served as the foundation for supervised classification, guiding the algorithms in assigning the correct class labels to the segmented image objects. The quality and representativeness of the training samples played a critical role in the performance and accuracy of the final classification, as they directly influenced the classifier's ability to differentiate between similar vegetation types.

Following this, different supervised classification classifiers such as SVM, NN, ML, and RF available in ArcGIS was applied to categorize the segmented image into the predefined five vegetation classes, ensuring a detailed and accurate classification of urban greenery.

Validation data

A total of 980 ground truth samples were collected to validate the classification results. These samples were generated using the "Create Accuracy Assessment Points" tool in ArcGIS, which randomly distributes points across the classified map. Each of these 980 points was then manually labeled by visually interpreting high-resolution reference imagery to determine the true land cover class at each location. Once the reference labels were assigned, a confusion matrix was generated using the "Compute Confusion Matrix" tool in ArcGIS. This matrix compares the manually interpreted ground truth data with the classified results to evaluate the model's performance by calculating accuracy metrics such as user accuracy, producer accuracy, and the kappa coefficient.

3 Results

A visual interpretation is shown in Fig 2, representing the original aerial image and the classification results produced by various classifiers which provides a comparative overview, allowing for an intuitive understanding of each model's performance and how they differ in predicting outcomes.

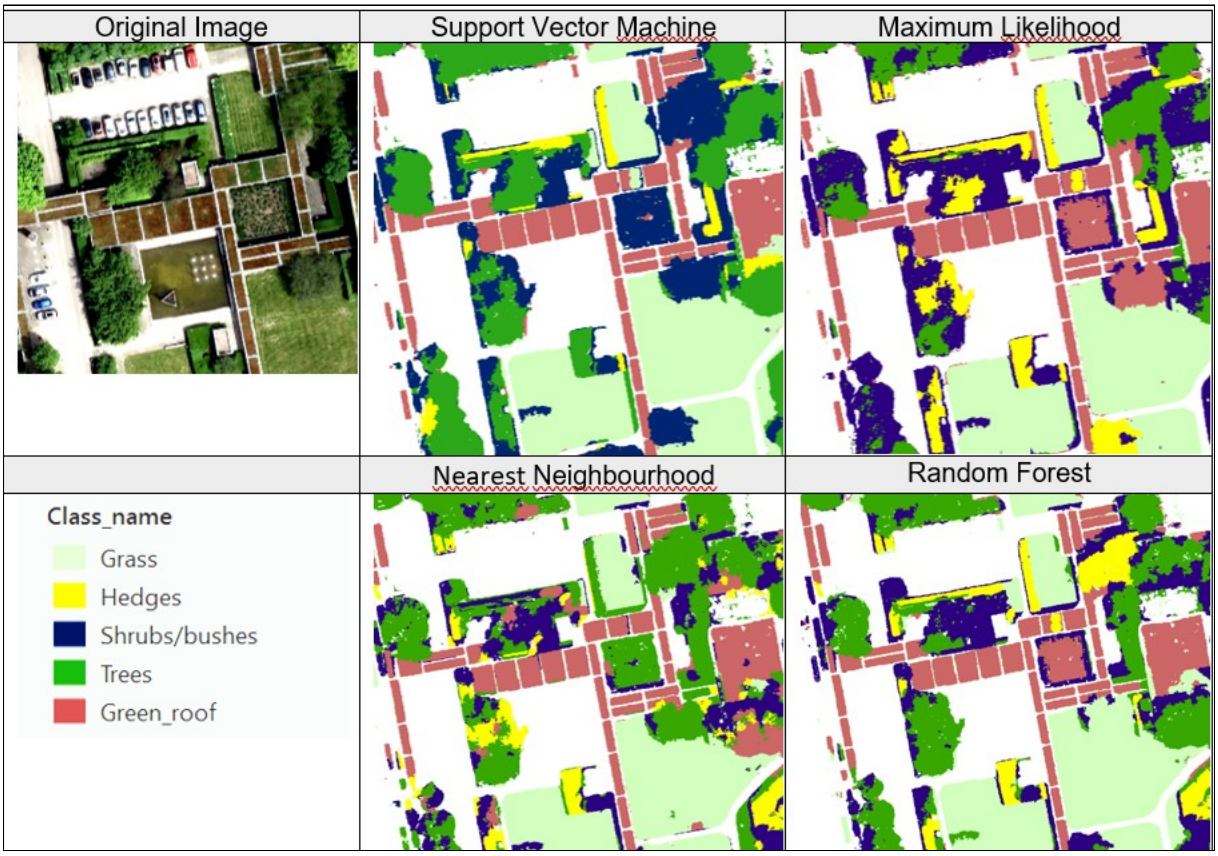

Fig. 2: Classification results of different classifiers. Aerial image: © 2024, Stadtmessungsamt Stuttgart

The classification performance varied across different vegetation classes, with distinct strengths and weaknesses observed for each classifier. Among the four classifiers - SVM, RF, NN, and ML each performed best for different types of vegetation. SVM was most effective at classifying trees, achieving the highest producer accuracy and the most correct classifications, while also performing well for grass. RF stood out in classifying hedges, with the highest user and producer accuracies, and also performed excellently for grass and green roofs. ML was best for identifying shrubs/bushes, showing the highest producer accuracy and most correct predictions in that class, and also performed well in detecting green roofs. Overall, grass was the easiest class for all models, while hedges and shrubs posed more challenges, with RF and ML handling them more effectively than others.

Green roofs are frequently misclassified as grass due to their similar spectral reflectance characteristics, particularly in the visible and near-infrared (NIR) bands. Likewise, shrubs and hedges are often confused with trees, as these vegetation types also exhibit overlapping spectral signatures. These challenges arise primarily because the classification relies solely on spectral data, which lacks information about the vertical structure of the features. Incorporating height information (e.g. 3D point clouds) such as from LiDAR or stereo photogrammetry can significantly enhance classification accuracy by capturing the three-dimensional characteristics of vegetation. This additional structural context enables better differentiation between low-lying vegetation like grass and green roofs, and taller features such as shrubs and trees.

Tab. 1: Correctly classified ground points, and Kappa values

	SVM	RF	NN	ML	Total ground points
Grass	610	545	596	604	682
Hedges	2	6	4	5	8
Shrubs/Bushes	38	48	25	56	99
Trees	147	135	91	71	169
Green Roofs	17	18	16	18	22
Kappa	0.6615	0.57289	0.501	0.5588	

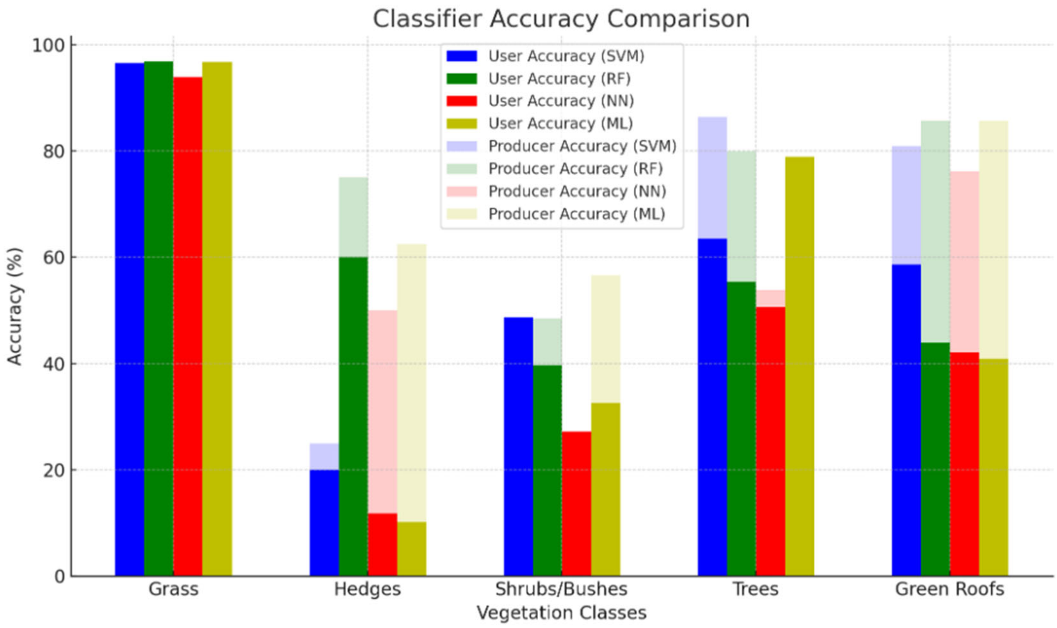

Fig. 3: Classification accuracies (%) of four classifiers- Support Vector Machine (SVM), Random Forest (RF), Nearest Neighborhood (NN), and Maximum Likelihood (ML) - showing user and producer accuracy for urban vegetation classification

4 Conclusion

In conclusion, the performance of the four classifiers SVM, RF, NN, and ML varied across different vegetation types, with SVM excelling in tree classification, RF performing best for hedges, and ML showing superior accuracy for shrubs. Grass was consistently the easiest to classify, while hedges and shrubs posed more challenges. Misclassifications, such as green roofs being confused with grass, arose due to similar spectral reflectance characteristics. Additionally, the confusion between shrubs, hedges, and trees was largely due to overlapping spectral signatures. These challenges can be addressed by incorporating height information from LiDAR or stereo photogrammetry, which would provide essential 3D structural data to enhance classification accuracy. By combining spectral and structural data, this approach can better differentiate between similar vegetation types, leading to more precise and reliable urban vegetation classification. This advancement is crucial for applications in urban planning, ecological monitoring, and smart city initiatives, offering a more comprehensive and accurate approach to mapping urban vegetation.

5 Bibliography

BAKER, F., SMITH, C. & CAVAN, G., 2018: A Combined Approach to Classifying Land Surface Cover of Urban Domestic Gardens Using Citizen Science Data and High-Resolution Image Analysis. Remote Sensing, **10**(4), 537. https://doi.org/10.3390/rs10040537.

BURMEISTER, J.-M., RICHTER, R. & DÖLLNER, J., 2023: Concepts and Techniques for Large-Scale Mapping of Urban Vegetation Using Mobile Mapping Point Clouds and Deep Learning. Wichmann Verlag. https://doi.org/10.14627/537740048.

GUO, Z., WANG, T., LIU, S., KANG, W., CHEN, X., FENG, K., ZHANG, X. & ZHI, Y., 2021: Biomass and vegetation coverage survey in the Mu Us sandy land - Based on unmanned aerial vehicle RGB images. International Journal of Applied Earth Observation and Geoinformation, **94**, 102239. https://doi.org/10.1016/j.jag.2020.102239.

MA, B., HAUER, R. J., ÖSTBERG, J., KOESER, A. K., WEI, H. & XU, C., 2021: A global basis of urban tree inventories: What comes first the inventory or the program. Urban Forestry & Urban Greening, **60**, 127087. https://doi.org/10.1016/j.ufug.2021.127087.

MAN, Q., DONG, P., YANG, X., WU, Q. & HAN, R., 2020: Automatic Extraction of Grasses and Individual Trees in Urban Areas Based on Airborne Hyperspectral and LiDAR Data. Remote Sensing, **12**(17), 2725. https://doi.org/10.3390/rs12172725.

NEYNS, R. & CANTERS, F., 2022: Mapping of Urban Vegetation with High-Resolution Remote Sensing: A Review. Remote Sensing, **14**(4), 1031. https://doi.org/10.3390/rs14041031.

RAUPP, M., CUMMING, A. & RAUPP, E., 2006: Street Tree Diversity in Eastern North America and Its Potential for Tree Loss to Exotic Borers. Arboriculture & Urban Forestry, **32**(6), 297-304. https://doi.org/10.48044/jauf.2006.038.

SÆBØ, A., POPEK, R., NAWROT, B., HANSLIN, H. M., GAWRONSKA, H. & GAWRONSKI, S. W., 2012: Plant species differences in particulate matter accumulation on leaf surfaces. Science of the Total Environment, **427-428**, 347-354. https://doi.org/10.1016/j.scitotenv.2012.03.084.

ZHANG, J., WANG, J., MA, W., DENG, Y., PAN, J. & LI, J., 2023: Vegetation Extraction from Airborne Laser Scanning Data of Urban Plots Based on Point Cloud Neighborhood Features. Forests, **14**(4), 691. https://doi.org/10.3390/f14040691.