SEMANTIC SEGMENTATION FOR MAPPING AGRICULTURAL WASTE SOURCES: A VINEYARD CASE STUDY FOR ENERGY VALORIZATION VIA BIOGAS PRODUCTION

Emina PETROVIĆ [1], Ana MOMČILOVIĆ [2], Dragana DIMITRIJEVIĆ JOVANOVIĆ[3*], Gordana STEFANOVIĆ[1], Miloš SIMONOVIĆ[1], Maša MILOŠEVIĆ[1], Vlastimir NIKOLIĆ[1]

[1]Faculty of Mechanical Engineering, University of Niš, Serbia

[2]Academy of Applied Sciences Polytechnic, Belgrade, Serbia

[3*]Faculty of Civil Engineering and Architecture, University of Niš, Serbia

[*]Corresponding author; E-mail: dragana.dimitrijevic@gaf.ni.ac.rs

Given the growing trend of increasing waste and diminishing resources, considerable efforts are being directed toward developing innovative methods for utilizing various types of waste as potential energy and material resources. Agriculture generates large quantities of waste, and inadequate management of this waste can cause severe environmental challenges. Transforming agricultural waste (AW) into biogas presents an excellent opportunity for its effective use; however, commercializing this process requires a comprehensive understanding of potential AW sources, primarily the types and quantities of waste generated. Consequently, this paper proposes a deep learning-based image segmentation approach for identifying potential AW sources using remote sensing images. The research examines the effectiveness of the DeepLabV3+ with various backbone networks for semantic segmentation with an emphasis on detecting vineyards as potential contributors to agricultural waste for biogas generation.

Key words*: agricultural residues, satellite imaging, vineyard mapping, semantic segmentation*

1. Introduction

With the increasing generation of waste and depletion of natural resources, innovative strategies have emerged to repurpose waste as valuable resources. Agricultural residues are particularly important for their dual role in providing chemical energy and essential nutrients. However, continuous extraction of agricultural waste (AW) depletes soil nutrients, requiring fertilization with synthetic nutrients like phosphorus, a critical resource in the EU [1]. Studies indicate that some landfills contain higher phosphorus levels than agricultural soils [2]. If left unmanaged, AW decomposes, releasing significant greenhouse gases and exacerbating environmental issues. On the other hand, AW can be converted into biogas through anaerobic digestion, reducing emissions by 13.4 kg CO_2-equivalent per cubic meter [3]. Europe generates 700 million tons of AW annually, underscoring its global potential for energy valorization [4]. Despite various waste management initiatives, such as composting and anaerobic digestion, AW remains underutilized, especially in regions like the Western Balkans [5, 6]. Anaerobic digestion represents a scientifically validated approach for the valorization of AW, enabling the conversion of organic matter into biogas a

renewable energy carrier, and digestate, a nutrient-rich byproduct suitable for soil enrichment. This process not only mitigates environmental impacts associated with untreated waste but also contributes to the transition towards a circular economy by transforming waste streams into valuable energy and material resources. Efficient AW management is hindered by a lack of comprehensive data on the types and quantities of AW produced. Variability in crop locations, yield fluctuations due to climate conditions, and other factors exacerbate this challenge. To bridge this gap, robust methodologies are needed to accurately estimate AW quantities. This study addresses the first step in this process: the identification of crop types and their spatial distribution using satellite imagery. Future work will focus on predictive modeling to estimate AW quantities based on crop conversion rates and statistical yield data.

The proposed methodology leverages deep learning-based image segmentation techniques on satellite imagery, captured during the vegetative period of crops, to identify and classify potential AW sources. By automating the detection process, this approach overcomes the limitations of traditional manual or drone-based methods, which are labor-intensive and time-consuming. Furthermore, the study focuses on vineyards as a case study, emphasizing the environmental and energy potential of grape pomace, a byproduct of wine production that poses environmental challenges due to its high acidity and organic content. Anaerobic digestion of grape pomace not only addresses its environmental impact but also converts it into biogas and digestate, creating a sustainable pathway for energy production and nutrient recovery. The objective of this study is to examine whether information can be effectively extracted from satellite imagery, with a particular focus on identifying potential sources of AW. In subsequent stages of the research, images will be acquired during the crop vegetation period, which will be a critical factor for accurate analysis.

2. Related work

Research on the use of remote satellite images for object detection, segmentation, and classification has evolved significantly, leveraging advancements in both traditional and deep learning methods. Early studies utilized texture statistics [7,8] for sea ice classification using Sentinel-1 (S1) synthetic aperture radar (SAR) data. Other approaches employed morphological characteristics to enhance image segmentation [9,10]. For object detection in high-resolution remote sensing images, advanced models, such as the rotation-invariant parts-based model [11] and rotation-invariant CNN (RICNN) [12] were introduced. The RICNN model attained rotation invariance by embedding a novel rotation-invariant layer within existing CNN architectures.

Further advancements included integrating sparse representations for local-feature detection with generalized Hough transforms to identify object classes or specific instances in high-spatial-resolution optical images [13]. The spatial sparse coding bag-of-words (SSCBOW) framework was also proposed to address the challenge of detecting targets with complex shapes in high-resolution remote sensing imagery [14]. These methods laid the foundation for applying deep neural networks (DNNs), whose adoption has grown due to their superior accuracy compared to conventional techniques [15]. Convolutional Neural Networks (CNNs) have emerged as a standout DNN architecture, excelling at image-driven pattern recognition tasks using images as direct input [16,17].

Recent innovations in deep learning, such as deep semantic segmentation networks, have significantly advanced the field of semantic segmentation [18]. This approach partitions an image into individual pixels, predicting the category of each pixel using large neural networks. Pixel-wise

semantic segmentation has proven efficient for identifying contextual features and improving classification accuracy [19], with applications spanning geographic information systems (GIS), autonomous vehicles, medical image analysis, and robotics.

Shelhamer et al. [20] transformed conventional classification networks such as AlexNet, VGG Net, and GoogLeNet into Fully Convolutional Networks (FCNs) specifically adapted for semantic segmentation tasks. Unlike traditional CNNs, FCNs omit fully connected layers and incorporate upsampling techniques to produce output images matching the input size, enabling precise pixel classification. SegNet [21], a deep convolutional encoder-decoder architecture, further advanced semantic pixel-wise segmentation. Its encoder network mirrors the convolutional layers in VGG16 [22], delivering improved performance in segmentation tasks by maintaining high-resolution feature maps. These innovations collectively highlight the progression from foundational methodologies to sophisticated deep learning techniques, offering unprecedented accuracy and efficiency in remote sensing and related applications. The development and refinement of semantic segmentation networks have been crucial for advancing the precise classification of objects in remote sensing imagery. Among these, SegNet stands out for its compact architecture, which was achieved by removing fully connected layers from the VGG16 encoder, making it smaller and easier to train. The decoder network in SegNet, its core component, uses a hierarchical structure where each decoder corresponds to an encoder, performing non-linear upsampling using max-pooling indices received from the encoder.

Further innovations include the Semantic Channel Upsampling Network (SCU-Net) proposed by Wang et al. [23], which integrates a channel attention mechanism and an upsampling convolution-deconvolution module (CDeConv). The CDeConv module aligns the feature map channels with segmentation task categories via point convolution and then adjusts the feature map size to match the original image dimensions through deconvolution. The SCU-Net also incorporates the Channel Feature Weight Extraction (CFWE) module to enhance feature extraction capabilities [24].

Another notable approach, U-Net [25], employs a U-shaped architecture featuring a contracting path for effective context capture and a symmetric expanding path to ensure precise localization. U-Net's efficiency relies heavily on data augmentation, enabling end-to-end training with limited annotated samples. Google's DeepLab series has also significantly advanced semantic segmentation. DeepLabV1 [26,27], based on VGG16, combines responses from Deep Convolutional Neural Networks (DCNNs) with a fully connected Conditional Random Field (CRF) to improve segmentation accuracy by modeling pixel-label agreement and contextual relationships. Subsequent improvements include DeepLabV2, introducing Atrous Spatial Pyramid Pooling (ASPP) [28,29] to capture contextual information at multiple scales by applying Atrous convolutions with varying sampling rates. DeepLabV3 [30] enhanced boundary sharpness and computational efficiency through an encoder-decoder structure using Atrous separable convolution. Unlike its predecessors, DeepLabV3 removed the CRF module and integrated batch normalization for improved training. DeepLabV3+ refined segmentation further with a decoder module designed to enhance results along object boundaries.

Inspired by these advancements, this research employs the DeepLabV3+ architecture for semantic segmentation of satellite remote sensing images, focusing on mapping vineyards as potential sources of AW. Semantic segmentation, a pixel-level classification task, clusters image parts into object classes by associating each pixel with a specific label, making it ideal for identifying spatial and spectral features in remote sensing data [31].

3. Study Area and Dataset

In training the deep learning network, it is quite easy to cause memory to overflow, especially when large remote-sensing images are used. With an increase in the resolution of satellite images, problems like loss of spatial information and imbalance of class distribution due to many small objects visible in the image will appear. Labeling such large images is a complicated and time-consuming task, making it difficult to produce large datasets [23]. Additionally, for training a deep learning neural network, the training data set should not be too small to avoid overfitting.

The data collection was conducted to recognize the vineyard as a source of AW. For this research, the dataset was gathered in two study areas. The first area was the vineyards at the foot of the mountain Fruška Gora and the second was the vineyards in the vicinity of the town of Vlasotince, Serbia. The satellite images were obtained from the publicly available Republic Geodetic Bureau [32]. Republic Geodetic Bureau uses digital orthophoto images (DOF) with a 30 cm resolution, enabling precise georeferencing of objects on the ground. These images were produced based on data collected during 2020 and 2021, using optical satellite systems such as WorldView-2, WorldView-3, and GeoEye-1. All images were taken with a ratio of 1:625. The images were preprocessed by performing random window sampling with a window size of 300 × 300 pixels, thus creating the uniform size of the experimental dataset (Fig.1). MATLAB's application Image Labeler was used to label images for semantic segmentation. The ground truth annotation for 414 images was generated, of which 372 images were used for training, while 42 images were used for testing to assess the performance of the DeepLabV3+ with the following backbone networks: ResNet18 [33], ResNet50, MobileNet-v2 [34], Xception [35], and Inception-ResNet-v2 [36].

Five pixel-classes were defined according to the landform characteristics of study areas and based on the objective of this study, shown in Fig. 2 as follows: 1) "Vineyard", 2) "Field", 3) "Forest", 4)" Road" and 5)" Others ". The "Other" class is used to label objects that do not belong to any of the four previously defined categories. These objects can include warehouses, various industrial structures, vehicles present on the road at a given time, or any other objects.

Data augmentation was applied to transform the original data randomly during training, hence more variety to the training data was added, and network accuracy was improved. The same random transformation was applied to both image and pixel label data using the datastore. The performed image transformations include rotation, vertical flipping, and horizontal flipping.

Figure 1. Collected examples for further processing (300 × 300 pixels)

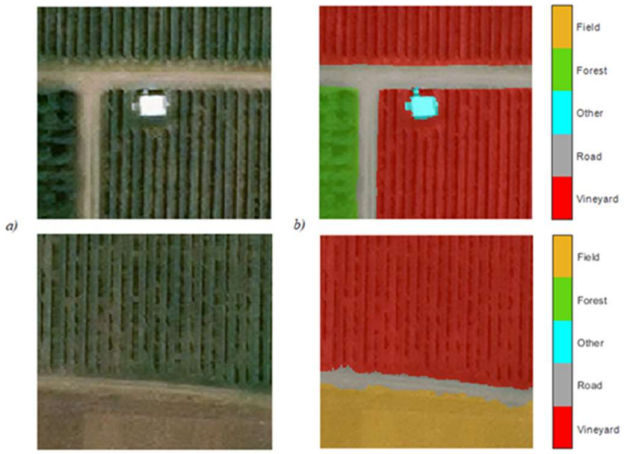

Figure 2. (a) Example images and (b) their segmentation masks

3.1. Class Imbalance

As previously stated, five-pixel classes were defined based on the study's objective therefore it could be considered a multi-class classification problem. Multi-class classification expects balanced data i.e., all classes in the training dataset should be equally distributed. This ensures the model is equally informed about all classes. In our dataset, 63% of pixels over all images were labeled as class "Vineyard", 3.80 % as "Road", 24.64% as "Forest", 8.34% as "Field", and 0.06% as "Others" as illustrated in fig.3. To improve training, class weighting was used to balance the classes. Different weights were assigned during the fine-tuning of the model with values of 0.41, 4.26, 0.86, 1.00, and 40.59 for each class, in the order specified above.

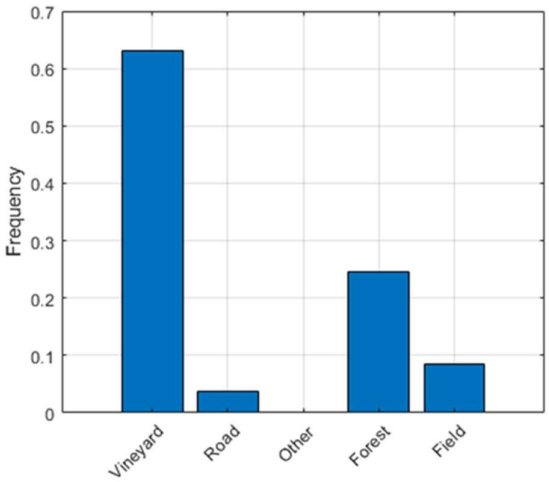

Figure 3. Distribution of pixels in different classes in the full dataset

By applying weights, we guided a model to prioritize learning based on the importance assigned to a certain class. Weights scaled the loss function. During the training process, the error was amplified by the weight assigned to each point as the model trained on each point. The main goal of the model was to decrease error on the more heavily weighted classes, since they have a larger influence on error, and consequently convey a stronger signal. This approach prevented the model from predicting the more common class more frequently.

4. Semantic segmentation based on DeepLabV3+ and backbone networks

4.1. Transfer learning

Generally, for the task of semantic segmentation, two approaches can be used. We can either create custom-built deep-learning networks or use pre-trained networks [37]. Creating a custom-built network implies that millions of labeled images are needed to develop a new model from scratch. Alternatively, we can use pre-trained deep learning networks that have already been trained on other datasets (transfer learning). Transfer learning is a technique in machine learning that allows a pre-trained deep learning model to be adapted and used as the foundation for a different but related task. Transfer learning is quite popular among researchers because, with small modifications to the network, high accuracy can be achieved in studies that lack sufficient training images and labels [38]. Following this procedure, in this paper for the task of semantic segmentation, we used the next pre-trained convolutional neural network architecture for the task of semantic segmentation: DeepLabV3+

with the following backbone networks (Fig.4): ResNet18, ResNet50, MobileNet-v2, Xception, and Inception-ResNet-v2. All backbone networks were pre-trained on more than a million images from the ImageNet database [39] and then adapted to our dataset.

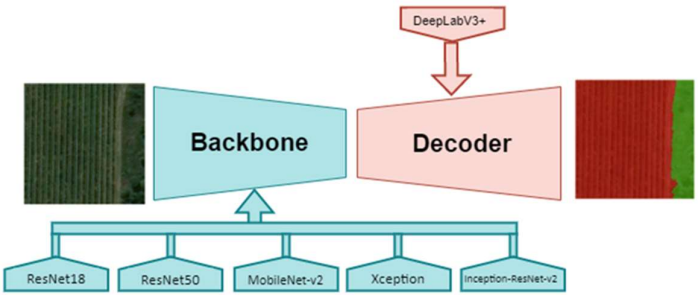

Figure 4. Proposed backbones and decoder

4.2. DeepLabV3+ and backbone

The DeepLabV3+ neural network utilizes an encoder-decoder architecture, where the encoder subnetworks (the backbone) are responsible for extracting complex features from the input images. The decoder network then processes these features to generate a segmentation map and accurately reconstruct the boundaries of objects at the original resolution of the image. Atrous spatial pyramid pooling module (ASPP), which follows the backbone network, is used to classify each pixel corresponding to their classes. The encoder incorporates an ASPP module that uses varying dilation rates (r = 6, 12, 18) to capture information across multiple scales. The output of the ASPP is then processed through a 1x1 convolution, upsampled, and merged with feature maps from the backbone network to effectively decode the encoder's outputs and refine the segmentation results. There is also the final upsampling at the end of the decoder which gives us the final segmented mask for the image.

4.3. Experiment Setup

In our experiments, we evaluated the performance of various networks for the segmentation of our dataset. For training and evaluating the neural networks, we followed a standard procedure. We initiated training with pretrained weights for each network and continued the process using our selected training and validation datasets until the validation loss had stabilized. The training began with a learning rate of 0.001, and we employed a piecewise learning rate schedule, reducing the learning rate by a factor of 0.3 every 10 epochs. To predict the class distribution, the Softmax function was applied to the network's output feature map. The Softmax loss was then calculated and backpropagated, and the network parameters were subsequently updated using Stochastic Gradient Descent (SGD) with a momentum of 0.9. During the training procedure, we fed the samples into the network in batches, and each batch contained 8 images. The training strategy and parameters were the same for all networks. The proposed networks were trained using 60% of the images from the dataset, while the remaining 40% were split equally into 20% for validation and 20% for testing. Before training, the whole dataset was always randomly shuffled. To ensure the reliability of the performance comparison of different network models, the experiments in this paper were conducted under the same platform and hardware environment. Our hardware specifications include an Intel Core i7 processor (i7-8750H CPU 2.20GHz), 20 GB of RAM memory, and a Nvidia GeForce GTX 1050 Ti graphics processing unit (GPU). On the software side, Windows 10 and MATLAB (R2021a) were used.

5. Experimental Results

5.1. Evaluation metrics

There are several known metrics [40] that can used to evaluate semantic segmentation tasks. The most used metrics for semantic segmentation are the intersection over union metric (IoU) [41] also known as the Jaccard Index, and the mean boundary F1 contour matching score (meanBF score) [42] which are the metrics we also used for our evaluation. For each class, the IoU metric indicates the ratio of correctly classified pixels to the total number of pixels in both the ground truth and predicted outputs for that class. The mean IoU, on the other hand, is computed as the average IoU score across all classes and all images within the dataset. To reduce the impact of errors in datasets where images have disproportionately sized classes, we also used weighted-IoU. This metric calculates the average IoU for each class, where each class's score is weighted by the number of pixels it contains. The BF score measures how accurately the predicted boundaries of each class match the true boundaries, assessing the alignment between the predicted and actual contours. The meanBF score represents the average BF score for the class across all images. We also calculated the average BF score of all classes in all images. In terms of the confusion matrix, these metrics can be defined as:

$$\boldsymbol{IoU} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP+FN}} \tag{1}$$

$$F_1 = \frac{2TP}{2TP+FP+FN} \tag{2}$$

Where TP, FN, FP, and TN represent the number of true positives, false negatives, and false positives, in pixel predictions for the class.

5.2. Results

Based on the training strategy, parameters, and dataset described above, DeepLabV3+ with the following backbone networks: ResNet18, ResNet50, MobileNet-v2, Xception, and Inception-ResNet-v2 was trained and validated using 372 images and evaluated on 42 images. The experimental results are displayed in Tab. 1, Tab. 2, and Fig. 5.

Table 1. Comparisons of DeepLabv3+ and backbone networks for pixel-wise classification (semantic segmentation)

Decoder	Backbone	meanIoU (%)	weighted-IoU (%)	meanBF score (%)
DeepLabV3+	ResNet18	0.7774	0.9366	0.7599
	ResNet50	0.7679	0.9347	0.7558
	MobileNet-v2	0.7274	0.9184	0.7006
	Xception	0.7358	0.9160	0.6924
	Inception-ResNet-v2	0.7672	0.9340	0.7743

According to the results presented in Table 1, DeepLabv3+ with ResNet18 achieved the best performance compared to the others, with a meanIoU of 0.7774% and a meanBF score of 0.7599 %. Contrary to expectations, a larger backbone network did not improve the results of our experiments. A mean IoU of 0.7774% in semantic segmentation indicates strong performance, approaching the standards expected in high-performance applications. However, a meanBF score of 0.7599% suggests that the model is slightly under-segmenting the classes, as the predicted areas are significantly smaller

than the actual areas in the ground truth. The reason for the slightly lower meanBF score can be found when we examine Table 2.

Table 2. The IoU and BF score for each class on the test dataset

Backbone network	ResNet18		ResNet50		MobileNet-v2		Xception		Inception-ResNet-v2	
Metrics Class	**IoU (%)**	**Mean BF score (%)**	**IoU (%)**	**Mean BF score (%)**	**IoU (%)**	**Mean BF score (%)**	**IoU (%)**	**Mean BF score (%)**	**IoU (%)**	**Mean BF score (%)**
Vineyard	0.9592	0.8025	0.9589	0.7944	0.950	0.7587	0.9315	0.705	0.9576	0.799
Road	0.6519	0.5980	0.6267	0.5477	0.575	0.5211	0.5816	0.5350	0.6418	0.606
Other	0.4593	0.2397	0.4341	0.2159	0.359	0.2169	0.4820	0.3896	0.4340	0.388
Forest	0.9376	0.7641	0.9325	0.7759	0.908	0.6716	0.9047	0.7020	0.9363	0.786
Field	0.8788	0.6388	0.8873	0.6839	0.843	0.5819	0.7791	0.5692	0.8662	0.695

The results presented in Table 2 show that classes with the highest number of pixels, such as “Vineyard” and “Forest,” achieved the best IoU and meanBF scores, while the class with the fewest pixels, “Other,” exhibited the lowest performance. The lower meanBF score of the model can be attributed to this class, as the model struggles with accurately recognizing and segmenting it. This is due to the class being less defined and often encompassing a variety of different objects. To further evaluate the overall performance of the semantic segmentation (pixel classification) method (DeepLabv3+ with ResNet18), a normalized confusion matrix was calculated. Figure 5 shows the classification accuracy of the true classes (ground-truth classes) versus the predicted classes. The most common misclassification of pixels is shown between the class “Road” and the class “Other”, The next misclassification is between class ‘Vineyard’ and class “Field”. The reason is that, in addition to the fields that have been planted with wheat for example, orchards are also included in the class "Field", in this research. The orchards often share similar features with vineyards, such as color and texture.

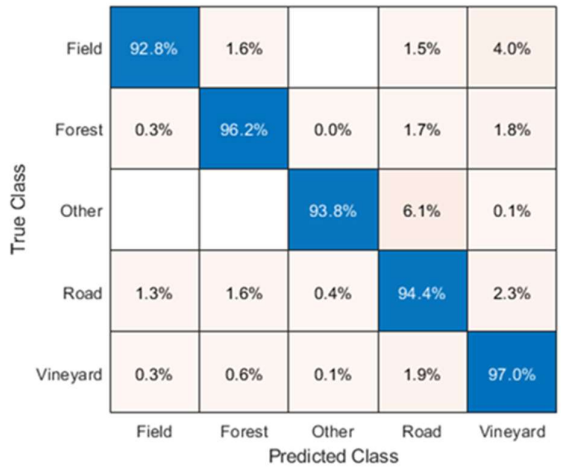

Figure 5. Normalized Confusion matrix of the pixel accuracy for all classes from the best-performing setup (DeepLabv3+ with ResNet18)

5.3. Discussion

Analyzing the outputs of the method and considering the goal of this study to identify and classify potential sources of AW, we can conclude that the proposed approach using ResNet18 is suitable for this challenging task. The results are promising and indicate that the model is quite acceptable, as most of the misclassification is due to incorrect classification between the "Road" and "Other" classes. To improve the model's performance, it would be beneficial to adapt the ground-truth images to include additional classes, such as "Orchards" and wheat "Fields", as a potential source of AW, and consider using a larger, more balanced training dataset.

The challenge arises from the fact that the boundaries in the images are not always the clearest due to the characteristics of the soil. Variations in soil texture, color, and lighting conditions can create challenges for the model, making it difficult to accurately distinguish between different regions. These factors can blur the edges of the objects being segmented, leading to a reduction in segmentation accuracy, particularly for classes like "Road" and "Other," where the boundaries are less defined. For example, the road is often shaded by tree canopies, making it more difficult to distinguish the road's boundaries in some areas.

6. Conclusion and future work

This paper focuses on recognizing and classifying AW sources using remote-sensing images. Proposed method is based on the DeepLabV3+ architecture with pre-trained deep neural networks for effective semantic segmentation. The study confirms the potential of CNN-based methods for segmenting AW sources, successfully identifying not only vineyards but also four additional classes, showcasing the method's versatility in detecting various AW sources.

Future research efforts are directed toward improving the generalizability and robustness of computational models by significantly expanding datasets. This will include incorporating additional labeled images and classes, such as "Orchards" and wheat "Fields," to capture a broader spectrum of agricultural landscapes. Moreover, research will delve into estimating the quantities of agricultural waste (AW) and analyzing seasonal dynamics, guided by regional agricultural practices and climatic variability. These investigations aim to provide a more nuanced understanding of AW availability throughout the year. Simultaneously, advancements in biogas production will emphasize the development of a holistic bioenergy chain framework. This integrated approach spans all stages, from the sourcing of biomass to its efficient conversion and delivery to end-users. By systematically assessing the physical and technical potential of biomass resources, this framework ensures a steady and optimized biogas supply, contributing to a sustainable energy system.

Building on prior findings [43], future optimization of the anaerobic digestion process will target key factors influencing its efficiency. Specifically, efforts will focus on refining the carbon-to-nitrogen (C/N) ratio to achieve a balanced microbial environment conducive to biogas production. Additionally, considerations will be given to waste generation rates and the geographic distribution of AW, as spatial dispersion presents logistical challenges. Addressing these factors will involve region-specific strategies tailored to the unique characteristics of AW sources, ensuring more effective collection, transport, and utilization within the bioenergy framework.

This comprehensive approach, combining model enhancements, waste quantification, seasonal analysis, and bioenergy system integration, is critical for the sustainable scaling of biogas production.

It aims to bridge technical, spatial, and environmental gaps, ultimately contributing to a resilient and efficient bioenergy network.

Acknowledgment

This research was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia under the Agreement on the Implementation and Financing of Scientific Research Work of the NIO in 2025 - Registration number: 451-03-136/2025-03/ 200095 and under the Agreement on Financing the Scientific Research Work of Teaching Staff - Registration number: 451-03-136/2025-03/ 200109.

References

[1] European Commission, 2006. Thematic Strategy for Soil Protection (STS). COM (2006) 231

[2] Mönkäre, T.J., et. al., Characterization of fine fraction mined from two Finnish landfills, *Waste Management*, *47(Part A)* (2016), pp. 34-39, https://doi.org/10.1016/j.wasman.2015.02.034

[3] Subramaniam, V., et. al., GHG analysis of the production of crude palm oil considering the conversion of agricultural wastes to by-products, *Sustainable Production and Consumption, 28* (2021), pp.1552-1564, https://doi.org/10.1016/j.spc.2021.09.004

[4] Pavwelczyk, A., EU Policy and Legislation on recycling of organic wastes of agriculture, International Society for Animal Hygiene, (2005)

[5] Mohammed, M.A., et. al., Utilization of various agricultural waste materials in the treatment of Industrial wastewater containing Heavy metals: A Review, *International Research Journal of Environment Sciences*, *3(3)* (2014), pp.62-71

[6] Lalić, D., et. al., Analysis of the opportunities and challenges for renewable energy market in the Western Balkan countries, *Renewable and Sustainable Energy Reviews*, 15(2011), 6, pp. 3187-3195, https://doi.org/10.1016/j.rser.2011.04.011.

[7] Clausi, D.A., An analysis of co-occurrence texture statistics as a function of grey level quantization, *Canadian Journal of Remote Sensing*, *28* (2002),1, doi: 10.5589/m02-004

[8] Barber, D.G., Ledrew, E.F., SAR Sea Ice Discrimination Using Texture Statistics: A Multivariate Approach, *Photogrammetric Engineering and Remote Sensing*, *57*(1991)

[9] Pesaresi, M., Benediktsson, J.A., A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *39* (2001), 2, doi: 10.1109/36.905239.

[10] Chen, T., et. al., Object-oriented landslide mapping using ZY-3 satellite imagery, random forest and mathematical morphology, for the Three-Gorges Reservoir, China, *Remote Sens (Basel)*, *9* (2017), 4, doi: 10.3390/rs9040333

[11] Zhang, W., et. al., Object detection in high-resolution remote sensing images using rotation invariant parts-based model. *IEEE Geoscience and Remote Sensing Letters, 11*(2014), 1, doi:10.1109/LGRS.2013.2246538.

[12] Cheng, G., et. al. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 54* (2016), 12, doi: 10.1109/TGRS.2016.2601622

[13] Yokoya, N., Iwasaki, A., Object detection based on sparse representation and Hough voting for optical remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8* (2015), 5,doi: 10.1109/TGRS.2016.2601622.

[14] Sun, H., et. al., Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model, *IEEE Geoscience and Remote Sensing Letters, 9* (2012), 1,doi: 10.1109/LGRS.2011.2161569.

[15] Hamida, A., et. al., Deep learning for semantic segmentation of remote sensing images with rich spectral content, *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, , Chennai, India, 2017, pp. 2569–2572.

[16] Yun, K., et. al., Deep neural networks for pattern recognition, *In Advances in Pattern Recognition Research*, (2018), pp. 49–79

[17] Wu, J., Introduction to convolutional neural networks, *Introduction to Convolutional Neural Networks*, 2017.

[18] Wang, P., et al., Understanding Convolution for Semantic Segmentation, *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision*, (2018), doi:10.1109/WACV.2018.00163.

[19] Guo, R., et al., Pixel-Wise Classification Method for High Resolution Remote Sensing Imagery Using Deep Neural Networks, *ISPRS International Journal of Geo-Information, 7*(2018), 3, pp. 110, doi: 10.3390/IJGI7030110.

[20] Shelhamer, E., et al., Fully Convolutional Networks for Semantic Segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*, (2017), 4, pp. 640-651, doi: 10.1109/TPAMI.2016.2572683.

[21] Badrinarayanan, V., SegNet: a deep convolutional encoder-decoder architecture for image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*, (2017), ,pp. 2481–2495

[22] Simonyan, K., Zisserman, A., Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556, 2014.

[23] Wang, W., Kang, Y., Liu, G., & Wang, X. (2022). SCU-Net: Semantic Segmentation Network for Learning Channel Information on Remote Sensing Images. Computational Intelligence and Neuroscience, 2022.

[24] W. Wang, H. Liu, J. Li, H. Nie, and X. Wang, "Using CFW-net deep learning models for X-ray images to detect COVID-19 patients," International Journal of Computational Intelligence Systems, vol. 14, no. 1, pp. 199–207, 2020.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, p. 3, 2015.

[26] L. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," in ICLR, 2015 .

[27] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, 2018.

[28] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking Atrous convolution for semantic image segmentation,” arXiv:1706.05587, 20

[29] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with Atrous separable convolution for semantic image Segmentation,” in Proceedings of the European Conference on Computer Vision (ECCV), pp. 833–851, Springer, Munich, Germany, September 2018.

[30] S. Das, A. A. Fime, N. Siddique and M. M. A. Hashem, "Estimation of Road Boundary for Intelligent Vehicles Based on DeepLabV3+ Architecture," in IEEE Access, vol. 9, pp. 121060-121075, 2021, doi: 10.1109/ACCESS.2021.3107353.

[31] Zhang, X., Fu, L., Karkee, M., Whiting, M. D., & Zhang, Q. (2019). Canopy segmentation using ResNet for mechanical harvesting of apples. *IFAC-PapersOnLin*e, 52(30), 300-305.

[32] Home | Statistical Office of the Republic of Serbia.” https://www.stat.gov.rs/en-US/

[33] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December. doi: 10.1109/CVPR.2016.90.

[34] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

[35] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition (*p. 1251-1258).

[36] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. *In the Thirty-first AAAI conference on artificial intelligence.*

[37] X. Zhang, L. Fu, M. Karkee, M. Whiting, and Q. Zhang, “Canopy Segmentation Using ResNet for Mechanical Harvesting of Apples,” in *IFAC-PapersOnLine*, 2019, vol. 52, no. 30. doi: 10.1016/j.ifacol.2019.12.550.

[38] B. Neupane, T. Horanont, and J. Aryal, “Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis,” *Remote Sensing*, vol. 13, no. 4. 2021. doi: 10.3390/rs13040808

[39] ImageNet. http://www.image-net.org

[40] E. Fernandez-Moral, R. Martins, D. Wolf, and P. Rives, “A New Metric for Evaluating Semantic Segmentation: Leveraging Global and Contour Accuracy,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2018, vol. 2018-June. doi: 10.1109/IVS.2018.8500497.

[41] D. Pathak, E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Multi-Class Multiple Instance Learning,” *3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings*, Dec. 2014, doi: 10.48550/arxiv.1412.7144.

[42] G. Csurka, D. Larlus, and F. Perronnin, “What is a good evaluation measure for semantic segmentation?,” *undefined*, 2013, doi: 10.5244/C.27.32.

[43] A. J. Momčilović, G. M. Stefanović, P. M. Rajković, N. V. Stojković, B. B. Milutinović, and M. P. Ivanović, "The organic waste fractions ratio optimization in the anaerobic co-digestion process for the increase of biogas yield," *Thermal Science*, vol. 22, Supplement 5, 2018.

Paper submitted: 17.12.2024

Paper revised: 31.01.2025

Paper accepted: 05.02.2025