Multispectral Image Segmentation in Agriculture: Evaluating Deep Learning Models with Train-Test Split and Cross-Validation Strategies

Wilgo Cardoso[1,2], Tiago Barros[1], Gil Gonçalves[2], Cristiano Premebida[1], Urbano J. Nunes[1]
[1]Institute of Systems and Robotics
Department of Electrical and Computer Engineering, University of Coimbra. Portugal
{wilgo.moreira, tiagobarros, cpremebida, urbano}@isr.uc.pt
[2]Institute for Systems Engineering and Computers at Coimbra
Dept. of Mathematics, University of Coimbra. Portugal.
gil@mat.uc.pt

***Abstract*— In agricultural robotics, the integration of multispectral image processing and deep learning (DL) has become the state-of-the-art (SOTA) in crop monitoring, yield estimation, and efficient land management. This work addresses the impact of different DL-segmentation models and evaluation protocols on multispectral imagery datasets collected by a UAV over vineyards. In terms of evaluation protocols, we have considered train-test split, standard k-fold cross-validation, and group k-fold cross-validation. While the first two assume that the training and test data are drawn from the same underlying distribution, the group k-fold cross-validation protocol assumes that each fold represents distinct distributions. Most works either adopt train-test split or k-fold cross-validation under the assumption that both the training and test sets are drawn from the same distribution. However, this assumption is rarely met in real-world applications. Therefore, the objective of this study is to evaluate and compare different evaluation protocols within the context of a real-world agricultural task, highlighting their limitations and weaknesses. Two SOTA DL-based segmentation models, SegNet and DeepLabV3, are employed to perform semantic segmentation on datasets of three Vineyards. The models have been trained and tested considering single-modality representations. In addition to the RGB modality, models trained on NDVI, GNDVI and early fusion are also evaluated. The performance of the models are evaluated using the IoU metric across different dataset configurations. The results indicate that the early fusion representation achieves the highest performance across the various splitting protocols, compared to the single-input representations. The results also show that the train-test and random k-fold splitting approaches report similar results. However, when employing group k-fold the performance drops consistently across both models and the modalities. This indicates that the models lack strong generalization capabilities to new data and, on the other hand, that the train-test and random k-fold splitting protocols are appropriate to evaluate model within the same distribution but are less adequate for out-of-distribution assessment.**

I. INTRODUCTION

Image segmentation plays a crucial role in modern precision agriculture due to facilitating the precise classification of remote sensing imagery (obtained from satellites and/or UAV drones). This, in turn, indirectly contributes to more efficient and accurate monitoring of crop health, yield estimation, and land management [4]. In this domain, multispectral imagery has emerged as a key source of information due to its sensitivity to variations in chlorophyll content, leaf moisture, and stress levels in plants [20]. Advancements in image segmentation, particularly in machine learning-based approaches, have been largely driven by deep learning (DL) techniques, which enable pixel-wise classification using end-to-end learning [15]. Among the various DL models, SegNet [3] and DeepLabV3 [8] have emerged as state-of-the-art architectures for semantic segmentation (shown in Fig. 1), each offering unique advantages in terms of performance and complexity. However, many existing studies rely on dataset splitting strategies, such as random train-test splits or k-fold cross-validation, which may not be sufficient for generalizing the model effectively across different conditions. These approaches can result in overfitting, as the evaluation is constrained to specific splits of the data, limiting the robustness of the model when faced with unseen data.

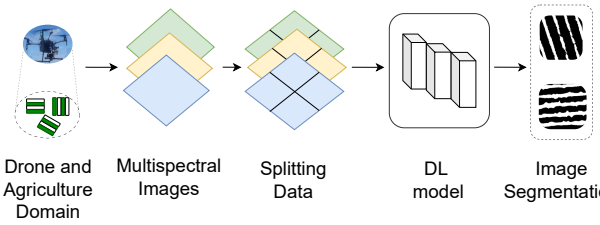

Fig. 1: Illustration of the use of deep learning models applied to precision agriculture based on mulstispectral imagery collected by an UAV (drone).

This work focuses on evaluating DL-based segmentation models in vineyard datasets by employing different dataset split strategies, including train-test splits and cross-validation techniques. The segmentation models were tested on multispectral images incorporating RGB (Red, Green, Blue), NDVI (Normalized Difference Vegetation Index), and GNDVI (Green Normalized Difference Vegetation Index) modalities. Additionally, an early fusion strategy was developed to combine these modalities to enhance segmentation performance. Complementary descriptions and the framework's stages are shown in Fig. 2.

The dataset used in this study was collected from three distinct vineyards in Portugal: Valdoeiro, Quinta-de-Baixo,

and ESAC. A detailed description is presented in [5]. These datasets encompass various environmental conditions ensuring a diverse representation for model evaluation. The performance of the segmentation models was assessed using the Intersection over Union (IoU) metric across different dataset configurations.

This paper aims to provide valuable insights into the application of SOTA deep-segmentation models for multispectral image segmentation in precision agriculture using real-world datasets. By exploring the impact of various dataset split strategies on model performance and modalities, this study concentrates on the advantages and drawbacks of single vs multi-representations and their performance on cross-domains (distribution shift) test-sets.

One of the key findings of this work is the significant improvement in segmentation performance achieved by early fusion when combining multispectral modalities. Additionally, although group-based cross-validation exhibited the lowest performance, it provides a more realistic - and challenging - measure of model generalization. This methodology highlights the importance of considering dataset size and diversity, as performance can substantially degrade when the model is applied to more heterogeneous data.

The remainder of this paper is organized as follows, Section II describes the related work on multispectral semantic segmentation applied to precision agriculture. Section III presents the proposed approach. In Sect. IV will be describe the experimental part and discuss the reported results. Finally, Sect. V concludes the paper with remarks and future directions.

II. Related Work

The literature on image segmentation is vast and diverse, encompassing a wide range of applications and methodologies. However, given the scope of this work, our focus is specifically on those techniques related to the context of precision agriculture with an emphasis on advanced image processing methods, including deep learning (DL) approaches [13] [17].

In the area of precision agriculture, the effectiveness of deep learning models in distinguishing between crops and weeds has been of great importance, generating the potential for automated weed control and modern agricultural robots thus, increasing production and reducing the use of chemicals [18] [11]. The use of convolutional neural networks (CNN) to analyse hyperspectral images has proved essential for verifying soil conditions and properties, which optimises crop growth. These advances indicate the increasing use of computational techniques to solve complex problems [14] [1]. The development of SOTA models such as U-Net [15], SegNet [3], and DeepLabV3plus [7] have grown and brought even more improvements to semantic segmentation in various areas [19].

The recent adoption of deep learning methods in agriculture related tasks for UAV or Satellite imagery segmentation can be successfully exemplified by deep neural networks (DNNs) such as Mask R-CNN [12], U-Net [15], and FracTAL ResUNet [10]. These networks have shown great results for satellite images. Additionally, the innovative architecture of new networks, such as the U-Net++ model [21], has also shown great results, which shows the power of deep learning approaches for solving complex problems [16] [11] [5].

Some works have highlighted the importance of variance in cross-validation and introduced the concept of variance estimation in model evaluation. For instance, Bengio and Grandvalet [6] focused on theoretical aspects with limited practical implementation guidelines. Additionally, Arlot and Celisse [2] provided a comprehensive survey of cross-validation methods, discussing their theoretical properties and practical applications. However, their survey is more focused on theoretical comparisons rather than empirical evaluations. In contrast, our work emphasizes practical implementation and application, specifically focusing on the use of cross-validation techniques in precision agriculture.

III. Methodology

This section details the methodology adopted for enhancing multispectral image segmentation on vineyard datasets using deep learning models. The approach encompasses the selection of models, the dataset domains, and the implementation of various dataset split strategies to evaluate the performance of the segmentation techniques.

Two state-of-the-art DL architectures are utilized for semantic segmentation, SegNet and DeepLabV3. The SegNet model is a deep convolutional encoder-decoder architecture designed for pixel-wise segmentation. It is known for its simplicity and effectiveness in various segmentation tasks [3]. Regarding DeepLabV3, it is a more complex segmentation model that utilizes atrous convolutions and spatial pyramid pooling to capture multi-scale context [7].

The dataset comprises multispectral images from three vineyards in Portugal: Valdoeiro (VAL), Quinta-de-Baixo (QDB), and ESAC [5]. These datasets include images captured by a drone (UAV) equipped with a multispectral camera sensitive to the bands R, G, B, NIR and RE.

The Fig. 3 illustrates the process of splitting diagram. Initially, the conventional Train-Test Split method is employed. The entire dataset, which comprises data from three vineyards, is divided into two subsets: a training dataset and a test dataset. This methodology offers a straightforward approach for evaluating the model's performance on unseen data.

Secondly, k-fold cross-validation is utilized (illustrated in Fig. 3). The dataset is divided into k-folds where each fold is used once as a test set while the remaining folds are used for training. This process is repeated k times (the number of folds), and the performance metrics are averaged to provide an overall evaluation of the model´s performance. This strategy helps to mitigate the overfitting by ensuring that each data point is used for both training and testing.

Finally, group-based cross-validation was applied, a more challenging and less explored strategy. The dataset is divided into groups based on vineyard locations. Cross-validation is

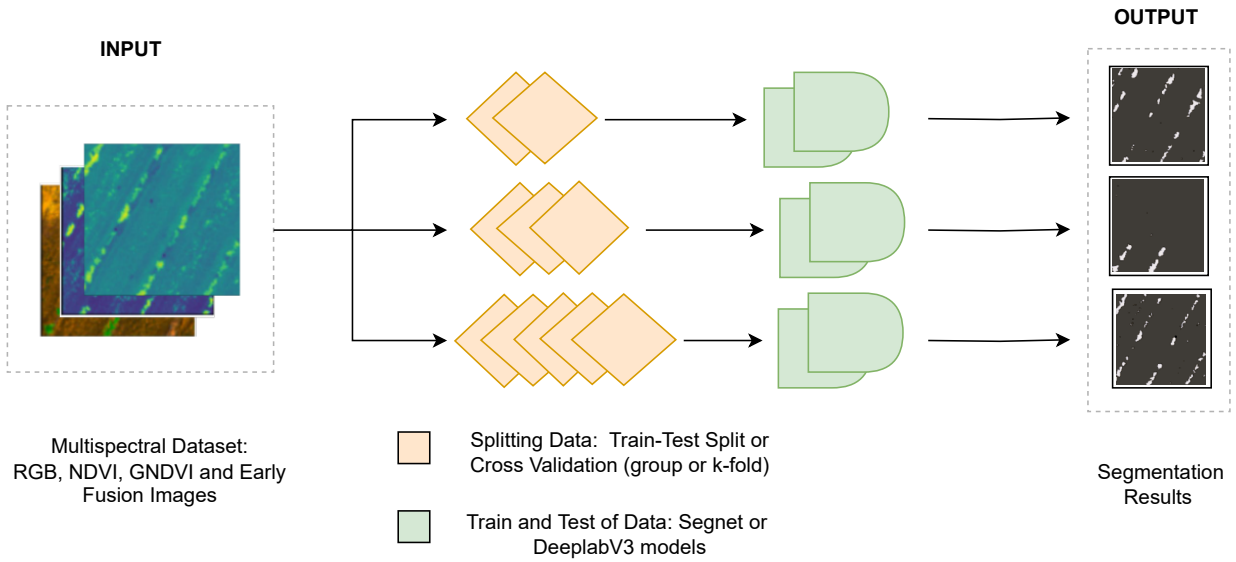

Fig. 2: Workflow of the Multispectral Dataset Processing. The diagram illustrates the steps involved in processing multispectral datasets, including the input of RGB, NDVI, and GNDVI images, data splitting for train-test or cross-validation, and the application of Segnet or DeeplabV3 models for segmentation. The workflow culminates in the output of segmentation results.

then performed by training the model on data from two vineyards and testing it on the remaining vineyard. This process is repeated for all three datasets, and the performance metrics are averaged to provide an overall performance assessment. These strategies collectively ensure a comprehensive evaluation of the segmentation models' performance and generalization capabilities.

IV. Experimental Evaluation

This section describes the experiments conducted with the proposed approach, utilizing SegNet and DeepLabV3 as baselines. The experiments are conducted within an agricultural context with the objective of segmenting vine plants. This segmentation task is crucial for precision agriculture, as it helps in monitoring the health and growth of the plants, leading to better yield and efficient resource management using robotic systems.

Specifically, RGB, NDVI, GNDVI images, and a combination of all three (Early Fusion) was employed. These modalities capture different aspects of the vine plants, providing a comprehensive and representative case-study that exposes the models and approaches to real-world conditions.

The RGB images provide standard color information, which is useful for general visual features. NDVI is a widely used vegetation index that highlights the presence and condition of vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs). GNDVI, on the other hand, is similar to NDVI but uses the green band instead of the red band, providing an alternative measure of plant health that can be more sensitive to chlorophyll content.

TABLE I: Dataset information for three sequences: Quinta de Baixo (QTA), ESAC, and Valdoeiro (VAL). Each image has dimensions of 240 x 240 pixels. The bands used are Red (R), Green (G), Blue (B), and Near-Infrared (NIR).

Sequence	QTA	ESAC	VAL
Samples	120	189	150

To complement and extend the baseline experimental part, an early-fusion technique (E.FUS) has been developed. Basically, E.FUS concatenates the RGB, NDVI, and GNDVI images into a single multi-channel input. This approach leverages the strengths of each modality, allowing the segmentation model to utilize a richer set of features. A comprehensive study was conducted to assess the segmentation performance of SegNet and DeepLabV3 on these various modalities.

The experiments involve training both models on the different input representations and evaluating their performance using the Intersection over Union (IoU). By comparing the results across the different modalities and models this paper aims to identify the most effective approach for segmenting vine plants in realistic agricultural setting.

A. Dataset & Evaluation

The evaluation of the approaches has been conducted on a vineyard dataset comprising areal multispectral imagery from three vineyards located at distinct locations in the central region of Portugal [5]. Thus, giving representativeness and diversity conditions - which are essential in machine learning and sensory perception. The collected data were recorded

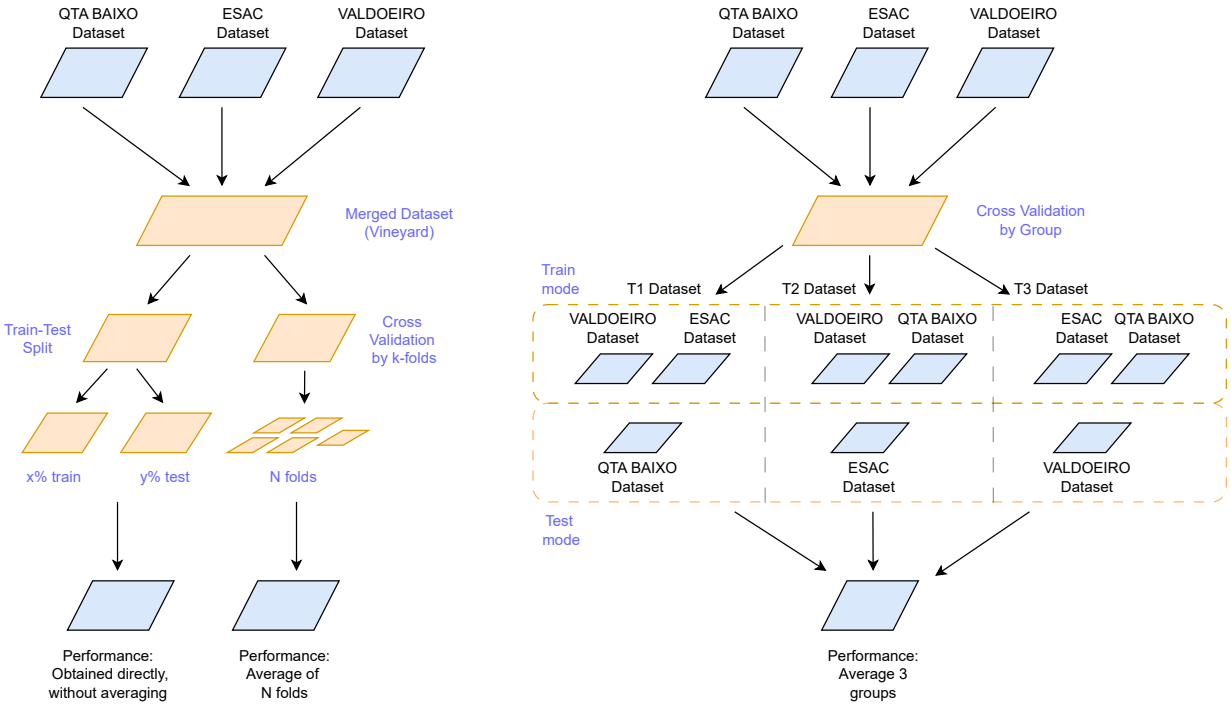

Fig. 3: Overview of Datasets and Cross-Validation Process. The diagram depicts the datasets used in the study, namely QTA Baixo, ESAC, and Valdoeiro datasets. It also illustrates the cross-validation process by group, ensuring a comprehensive evaluation of the model's performance across different subsets.

during distinct seasons, using an Unmanned Aerial Vehicle (UAV) equipped with X7 RGB camera and a multispectral Micasense Altum sensor. The dataset provides ground-truth annotations for the vine plants [5].

In this work, we followed the evaluation strategy proposed in [9], where musltispectral sensory data is used, namely the RGB and NIR spectral bands. The evaluation is conducted on RGB images, and on two distinct vegetation indices, namely the NDVI and GNDVI. These indices are calculated from the spectral reflectance measurements captured in the RED, GREEN, and NIR (Near-Infrared) bands, which are computed as follows:

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{1}$$

$$GNDVI = \frac{NIR - GREEN}{NIR + GREEN}. \tag{2}$$

Besides the aforementioned modalities, an additionally representation is evaluated, combining RGB, NDVI and GNDVI in a single input modality, stacking the tree representations channel-wise. This representation is referred to as Early Fusion (E.FUS). All modalities have been re-sized to a 240 $\times$ 240 shape. Detailed information of each sequence is provided in Table I. The results are reported using IoU, which is obtained as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{3}$$

where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

In this work, we chose to use IoU instead of the Dice metric, as IoU is a widely adopted standard metric in segmentation tasks, particularly among state-of-the-art methods. IoU provides a more stringent evaluation by penalizing smaller differences between the predicted and ground truth segmentations, which allows for a more accurate assessment of the model's performance. Additionally, IoU has become the *de facto* metric in several benchmarks and segmentation challenges, making it more suitable for comparative analysis. While Dice and IoU are similar in many ways, IoU's stricter penalty for discrepancies makes it particularly appropriate for our task where precise boundary matching is crucial.

B. Training and Implementation Details

All models were trained and evaluated under uniform conditions, namely executed on a machine equipped with an AMD Ryzen 9 5900X 12-Core CPU with 64 GB RAM and a NVIDIA GeForce RTX3090 GPU. The code is implemented in Python 3.9, utilizing PyTorch with CUDA 11.7.

The DL-based segmentation models, SegNet and DeepLabV3 were implemented based on their original code. The training was made with a learning rate of 10^{-3}, a batch size of 30 images, while testing was performed with a batch size of 3. As for the DeepLabV3 model, the ResNet-50 CNN as backbone was used.

TABLE II: Performance metrics are reported using **IoU** in percentage [%]. The performance of each approach is presented across different evaluation strategies: Cross Validation (Group), Split (70-30, 75-25, and 80-20)%, Cross Validation (3 Folds, 4 Folds, 5 Folds, and 6 Folds). Values in bold highlight the best performance.

		Cross Val: 3 Folds	Cross Val: 4 Folds	Cross Val: 5 Folds	Cross Val: 6 Folds	Split: 70%-30%	Split: 75%-25%	Split: 80%-20%	Cross Val: Group
SegNet	RGB	68.33	72.71	72.64	72.54	70.47	**75.35**	74.11	36.83
	NDVI	68.91	69.04	69.54	69.00	67.24	62.22	68.02	38.10
	GNDVI	67.58	67.90	67.50	68.21	68.39	67.77	68.81	37.15
	E.FUS.	**73.33**	**73.58**	**74.03**	**73.69**	**72.86**	73.31	**74.17**	**46.43**
DeeplabV3	RGB	72.37	**73.25**	73.23	73.45	71.37	69.78	70.92	27.78
	NDVI	68.78	69.22	69.53	69.64	66.86	67.22	67.39	23.69
	GNDVI	67.27	67.63	68.01	68.32	66.78	67.74	66.90	12.66
	E.FUS.	**72.72**	72.88	**73.54**	**73.65**	**72.76**	**72.23**	**71.17**	**32.23**

C. Results & Discussion

The evaluation of the approaches is conducted on multispectral imagery collected from three vineyards in Portugal: Valdoeiro (VAL), Quinta-de-Baixo (QDB), and ESAC. Each vineyard dataset includes RGB, NDVI, and GNDVI images, providing a comprehensive basis for segmentation tasks. Table II reports the segmentation performance of SegNet and DeepLabV3 across different dataset split strategies, train-test split and two cross-validation techniques. The results indicate that early fusion (E.FUS.) of RGB, NDVI, and GNDVI images consistently outperforms individual modalities, demonstrating significant improvements in IoU metrics. For instance, SegNet achieved an IoU better with early fusion under the cross-validation by k-folds strategy, compared with the train-test split.

Figure 4 shows some image segmentation results for the various modalities and dataset split strategies, indicating qualitative differences in the segmentation outputs, showcasing the superior performance of early fusion techniques and cross validation by k-folds.

The fourth column of the image illustrates the poorest segmentation results, where the real image significantly diverges from the segmented output. This outcome is attributed to the low performance index of the model on a particular subset of the dataset, which was divided according to groups/domains. In contrast, when alternative methods of dataset division are employed, the segmentation results do not exhibit such marked discrepancies.

The quantitative and qualitative results highlight the effectiveness of using early fusion strategies for multispectral image segmentation in vineyards. By combining RGB, NDVI, and GNDVI modalities the models can leverage a richer set of features, resulting in more accurate and detailed segmentation.

This paper also reveals that cross-validation by group, while providing a more realistic measure of model generalization, exhibits the lowest performance metrics. This outcome suggests that when evaluating model generalization capabilities it is crucial to consider the size and diversity of the dataset.

In conclusion, our findings highlight the potential of advanced deep learning models combined with thoughtful dataset split strategies to enhance the performance of agricultural monitoring systems. These insights pave the way for more precise, efficient, and sustainable agricultural practices, ultimately benefiting crop management and resource utilization.

V. Conclusion

This study demonstrates the potential of DL-based multispectral image segmentation in precision agriculture, specifically within vineyard datasets, with many different splitting sets. By employing state-of-the-art segmentation models, SegNet and DeepLabV3, and evaluating them across various dataset split strategies, including train-test splits and cross-validation techniques, this research provides a comprehensive analysis of their performance.

The results highlight the significant improvements in segmentation performance achieved through the early fusion of RGB, NDVI, and GNDVI modalities. This approach leverages the unique strengths of each spectral band, leading to more detailed and precise segmentation outcomes. Additionally, the study emphasizes the importance of dataset split strategies in evaluating model performance. While cross-validation by group demonstrated the worst performance metrics, it provided a crucial insight into the generalization capabilities of the models, reflecting more realistic scenarios in diverse agricultural environments. This strategy, however, demands careful consideration of dataset size and diversity to mitigate potential performance drops.

Overall, our findings offer valuable insights for researchers and practitioners in the field of precision agriculture. The combination of advanced deep learning models with diverse dataset splitting techniques presents a powerful tool for enhancing the performance of agricultural monitoring systems. As future work, we will continue to explore the integration of additional multispectral bands and further refine dataset split strategies to maximize model generalization and performance.

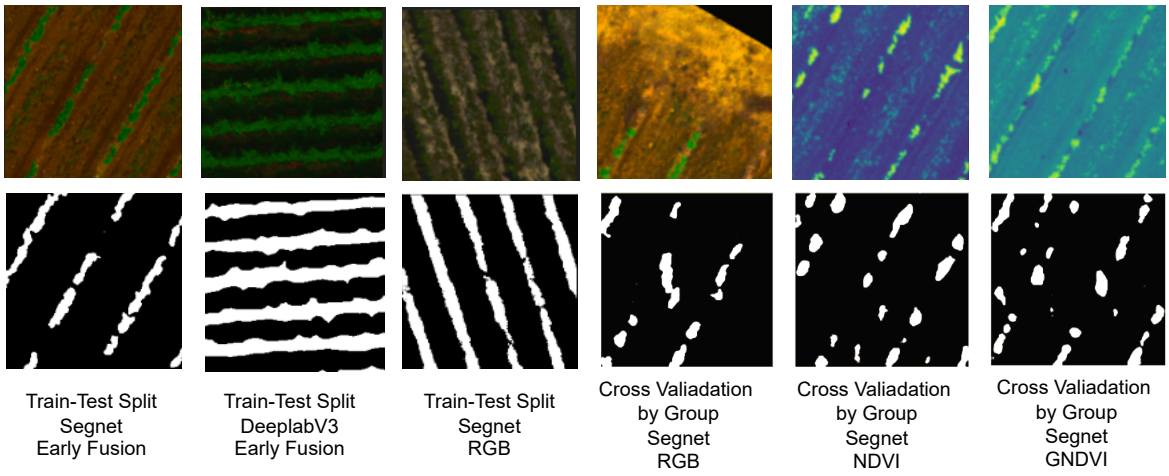

Fig. 4: Each column corresponds to the image and its mask obtained from the models and modalities described in the image.

ACKNOWLEDGMENT

This work has been supported by the project GreenBotics (PTDC/EEI-ROB/2459/2021), funded by the Portuguese Foundation for Science and Technology (FCT), and also by the project grants UIDP/00048/2020, UIDB/00308/2020 (with DOI 10.54499/UIDB/00308/2020), and by the PhD grant with reference 2021.06492.BD.

REFERENCES

[1] Halimatu Sadiyah Abdullahi, R Sheriff, and Fatima Mahieddine. Convolution neural network in precision agriculture for plant image recognition and classification. In *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, volume 10, pages 256–272. Ieee New York, 2017.

[2] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010.

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[4] Cristina Balaceanu, Robert Streche, Roxana Roscaneanu, Filip Osiac, et al. Advanced precision farming techniques employing wsn and uav. In *Proceedings of the 12493rd International Conference on Precision Agriculture*, pages 124930R–124930R. Vol. 12493, 2023.

[5] T. Barros, P. Conde, G. Gonçalves, C. Premebida, M. Monteiro, C. S. S. Ferreira, and U. J. Nunes. Multispectral vineyard segmentation: A deep learning comparison study. *Computers and Electronics in Agriculture*, 195:106782, 2022.

[6] Yoshua Bengio and Yves Grandvalet. No unbiased estimator of the variance of k-fold cross-validation. *Advances in Neural Information Processing Systems*, 16, 2003.

[7] 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*, 40(4):834–848, 2018.

[8] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2094–2107, 2017.

[9] Nuno Cunha, Tiago Barros, Mário Reis, Tiago Marta, Cristiano Premebida, and Urbano J Nunes. Multispectral image segmentation in agriculture: A comprehensive study on fusion approaches. *arXiv preprint arXiv:2308.00159*, 2023.

[10] Foivos I. Diakogiannis, François Waldner, and Peter Caccetta. Looking for change? roll the dice and demand attention. *Remote Sensing*, 13(18), 2021.

[11] H. Fathipoor, R. Shah-Hosseini, and H. Arefi. Crop and weed segmentation on ground-based images using deep convolutional neural network. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2023.

[12] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[13] Mihai Valentin Herbei, Cosmin Alin Popescu, Radu Bertici, Adrian Smuleac, and George Popescu. Processing and use of satellite images in order to extract useful information in precision agriculture. *Bulletin of the University of Agricultural Sciences & Veterinary Medicine Cluj-Napoca. Agriculture*, 73(2), 2016.

[14] Y. Miao, C. Silva, and F. H. Holman. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 104:103–112, 2019.

[15] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer, 2015.

[16] G. O. Tetteh, M. Schwieder, S. Erasmi, C. Conrad, and A. Gocht. Comparison of an optimised multiresolution segmentation approach with deep neural networks for delineating agricultural fields from sentinel-2 images. *Pfg – Journal Of Photogrammetry, Remote Sensing And Geoinformation Science*, 2023.

[17] Dimosthenis C Tsouros, Stamatia Bibi, and Panagiotis G Sarigiannidis. A review on uav-based applications for precision agriculture. *Information*, 10(11):349, 2019.

[18] Suneel Tummapudi et al. Deep learning based weed detection and elimination in agriculture. In *2023 International Conference on Inventive Computation Technologies (ICICT)*, pages 147–151. IEEE, 2023.

[19] Vivian Wen Hui Wong et al. Segmentation of additive manufacturing defects using u-net. *Journal of Computing and Information Science in Engineering*, 22(3):031005, 2022.

[20] Liyuan Zhang, Huihui Zhang, Yaxiao Niu, and Wenting Han. Mapping maize water stress based on uav multispectral remote sensing. *Remote Sensing*, 11(6):605, 2019.

[21] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*, volume 4, pages 3–11. Springer International Publishing, 2018.