IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 June 2025, accepted 22 July 2025, date of publication 31 July 2025, date of current version 6 August 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3594406

RESEARCH ARTICLE

AI-Driven Background Segmentation for High-Throughput 3D Plant Scans

SERKAN KARTAL[1], JAN MASNER[2], JANA KHOLOVÁ[2], ALEXANDER GALBA[2], THARANYA MURUGESAN[3], REKHA BADDAM[3], VOJTĚCH MIKEŠ[2], AND EVA KÁNSKÁ[2]

[1]Department of Computer Engineering, Çukurova University, 01380 Adana, Türkiye
[2]Department of Information Technologies, Czech University of Life Sciences Prague, 165 00 Prague, Czech Republic
[3]Crop Physiology and Modeling, International Crops Research Institute for the Semi-Arid Tropics (ICRI-SAT), Patancheruvu, Telangana 502324, India

Corresponding author: Jan Masner (masner@pef.czu.cz)

This work was supported in part by the European Comission's (EC's) Horizon Europe funded by the Project maximising the CO-benefits of agricultural Digitalisation through conducive digital ECoSystems (CODECS), under Grant 101060179; in part by the Institutional Grant of the Internal Grant Agency of the Faculty of Economics and Management, Czech University of Life Sciences Prague, under Grant IGA 2023B0005; and in part by the Project: Precizni zemědělstvi a digitalizace v Digitalization in Czech Republic, under Grant QK23020058.

ABSTRACT Accurate background segmentation in 3D plant phenotyping is crucial for reliable trait assessment but remains challenging. Current methods are either excessively complex, developed for a different domain, or lead to data loss (coordinate-based). This paper addresses these issues by introducing an AI-driven approach using a Multi-Layer Perceptron (MLP) model, leveraging RGB, spatial (XYZ), and near-infrared (NIR) data to enhance precision. The method was evaluated on high-throughput phenotyping data, achieving a classification accuracy of 0.9993, significantly reducing false positives and false negatives compared to coordinate-based segmentation. The proposed approach improved segmentation, particularly in early growth stages and for prostrate species, where traditional methods often fail. The model's impact on leaf area estimation was validated against destructive measurements, demonstrating substantial accuracy improvements, especially for species with small and prostrate canopies. Additionally, the model exhibited strong generalization capabilities when applied to an external 3D dataset, confirming its reusability beyond plant phenotyping tasks. Integrating this simple method into phenotyping pipelines will enhance efficiency and accuracy in high-throughput trait estimation, supporting advancements in plant science and precision agriculture.

INDEX TERMS Background segmentation, 3D imaging, point cloud processing, multi-layer perceptron, machine learning, plant phenotyping, remote sensing, precision agriculture, artificial intelligence.

I. INTRODUCTION

Phenomics, particularly plant phenotyping, is an emerging research discipline focusing on understanding plant-based systems' dynamics. It is frequently aided by imaging technologies that generate a lot of data, which, in turn, requires processing using fit-for-purpose computational methods. The phenomics tasks support a range of applications in plant-related research disciplines, particularly those requiring high throughput and non-destructive plant monitoring – e.g., crop characterization for breeding, genebanks, agronomy, or basic research of plant-based system dynamics (e.g., stress responses to biotic/abiotic stimuli). Much of such research is done with sensors that capture 2D/3D RGB/multi-/hyper-spectral reflection of the plant systems (i.e., digital twins). While 2D imaging has been intensively explored for decades [1], the utilization of information from 3D imaging methods (e.g., LiDAR) is lagging behind, particularly for the new generation of sensors that also capture color and other information (e.g., near-infrared). In this paper, we particularly focus on the 3D plant scans. Those scans usually contain information on the background

The associate editor coordinating the review of this manuscript and approving it for publication was Sunil Karamchandani.

surrounding the plant, such as soil, a growing container, irrigation tubes, etc. Therefore, this background must be identified (segmented) and removed for further analysis to assess the plant characters per se.

Popular methods used to gather 3D data reconstruction are based on multiple 2D images taken from different positions – see categorization by [2]. Those systems capture 2D images of a plant from various angles and process them to create 3D models. The photogrammetry methods usually provide high-quality models. On the other hand, these systems do not provide high throughput. This paper primarily focuses on laser-based scanner data originating from high-throughput scanners like Phenospex's (PSX) PlantEye$^{(R)}$, used for plant phenotyping. These types of scanners are usually mounted on stationary gantries with fixed positions and used for more accurate evaluation of crops.

Extensive research is being conducted in automotive and robotics, using mainly LiDAR (Light Detection and Ranging) sensors. In this field, various methods, such as filtering, have been used to preprocess the raw point cloud data [3]. Ground segmentation in 3D point clouds is also addressed throughout the literature. Most of the methods were developed for the automotive LiDAR datasets, e.g., [4], [5], [6], [7], [8], [9], [10]. Other authors address the area of robotics, e.g., [11], [12]. In the mentioned studies, LiDAR mainly captures information about moving scenes where the scanner and/or objects are not static. Thus, the analytics suitable for LiDAR image analysis might not fit all 3D data types. The abovementioned methods consider the ground as a flat surface that is narrow or tilted. In the plant-systems analysis (e.g., phenotyping), non-flat soil is often surrounded by additional background data. In plant-system analysis, LiDAR is frequently mounted on aerial moving vehicles like UAVs, tractors, or planes. It is used to infer plant canopy traits like plant height or canopy cover over large areas.

Various methods for background segmentation are suitable for different types of data – in general, these can be methods based on thresholding (i.e., height, the color of the individual points) or methods based on the image features that might be evolving/considering combinations of point-cloud features and their hidden patterns. In plant-systems analyses, authors mostly use direct methods to segment/detect whole plants, their organs, or roots, e.g., [13], [14], [15], [16], [17]. In our previous studies [18], [19], we utilized Region Growing Segmentation (RGS) [20] and Random Sample Consensus (RANSAC) [21], which are widely recognized and frequently applied algorithms for background segmentation in the literature. However, to achieve satisfactory results with these methods, the structure of our dataset required the separation of growing containers based on their mathematical coordinate information prior to segmenting background and plant data. Additionally, it was observed that these methods are not well-suited for scenarios involving small plants or inclined/wavy soil surfaces.

A typical simplistic segmentation strategy, which is supported by most preprocessing pipelines, such as Phenospex's Phena, is to divide plant and soil by employing a specified cutoff height (z-coordinate). The upper portion is considered plant data, and the lower portion is background. The cut height is usually set above the rim of the tray to get rid of the tray points fully. However, as shown in Figure 1, in certain situations (e.g., early growth stages, prostrate type of plant canopy architectures), part of the plant is below the rim of the tray. This cut can lead to data loss. Therefore, this approach often leads to inaccurate data analyses and trait estimation. Additionally, the optimal separation height must be determined for each setup, complicating standardization and requiring precise repetition across experiments.

FIGURE 1. A scan of a small plant that grows below the rim of the tray. Traditional coordinate-based soil segmentation would cut the scan above the rim and lead to significant plant data loss.

In summary, existing algorithms are either very complex, developed for naturally different data, data-lossy, or too simple that they need a custom preprocessing setup for each different experiment type, i.e., setup of different coordinates for different growing containers, positioning, etc. The presented paper focuses on plant scans in 3D point clouds obtained by high-throughput phenotyping platforms. In particular, we use data from Phenospex's PlanEye F600 scanning technology [22] and the installation in Hyderabad, India [23]. While the platform was validated and is being deployed for a range of end-uses [23], [24], [25], some of the uses are currently constrained by the accuracy of the plant feature inference algorithms. As illustrated in FIGURE 1, this static method causes substantial data loss, leading to inaccuracies in the plant trait assessment.

The presented research aims to address two current gaps – an accuracy for a wide range of data while maintaining the necessary performance for high-throughput systems. The main objective of this paper is to provide an AI-driven method for segmenting the background from the plant data points for a wide range of species. This method can be used during data preprocessing before other analytical algorithms provided by phenotyping platforms. Moreover, the proposed method addresses the limitations of traditional lossy threshold-based segmentation methods. Additionally, we evaluate the efficiency in several ways, i.e., AI model evaluation using a test set, point counts comparison to the coordinate-based method, canopy trait inference using destructive measurement, and finally, a generalization capacity using an external, different-domain data set.

II. MATERIALS AND METHODS

This section describes the methodology used for background segmentation in 3D plant point clouds. It covers data

acquisition using a high-throughput scanning platform, preprocessing steps to rotate the raw scans, manual annotation for ground truth generation, dataset partitioning for model training, and model performance and accuracy evaluation. The entire workflow is summarized in FIGURE 2.

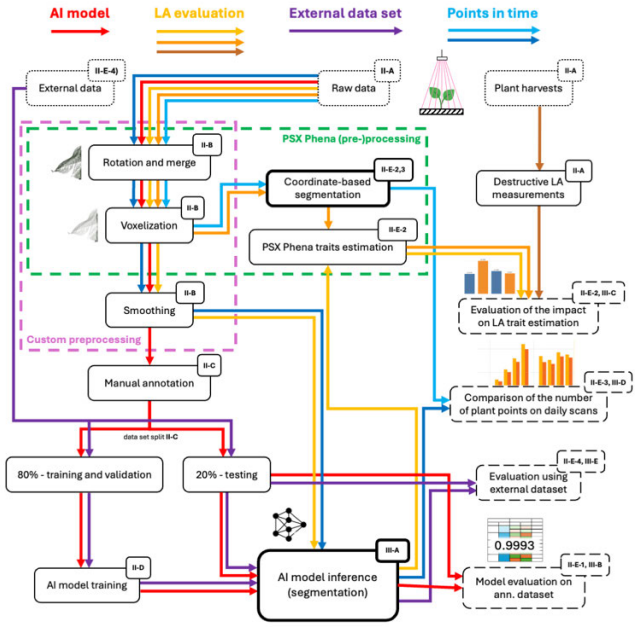

FIGURE 2. Visualization of the whole methodology. We used the data from Phenospex's PlantEye scanner, ran preprocessing, developed a classification model, and evaluated it in different manners. The figure shows individual steps (referenced by section numbers) and paths for different data visualized by colors. The colors, together with the source data, are referred to in TABLE 1.

A. DATA COLLECTION

The data used in this study were generated using a LeasyScan platform (ICRISAT, Hyderabad, India; Facilities & Services – GEMS, details in [23]), using PlantEye F600, a 3D scanner with multispectral imaging (PlantEye F600 multispectral 3D scanner for plants - PHENOSPEX). LeasyScan uses a dual scanning system (2 partially overlapping scanners capture the same area, mounted on the construction shown in FIGURE 3) to capture the 3D reflection of the target area and the capacity to scan an area of $\sim$2500m^2 in 1h30min. As shown in FIGURE 3, the scanned area was equipped with the containers in which the crops were raised. In our study, we used a setup of microplots, each consisting of 40 $\times$ 60$\times$60 cm blue plastic trays (blue containers in FIGURE 3) with $\sim$ 70 kg of vertisol equipped with drip irrigation tubes. The scanning area was configured to capture the area delineated by individual microplots up to 1.5m from the ground via HortControl software. The area is divided into sectors that contain a certain number of microplots (24 divided into two rows for this study). The sectors are marked by physical barcodes that are recognized by the scanners. Finally, each scan consists of two separate *.ply files angled towards each other that cover one sector.

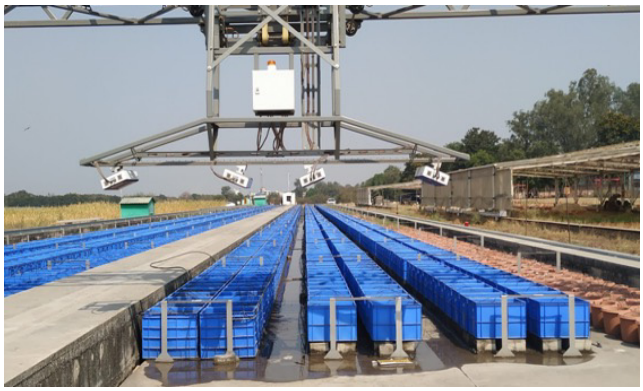

FIGURE 3. The LeasyScan high-throughput phenotyping platform was used to gather the data. The picture shows the dual position (twice two complementary, partially overlapping scanners capture the same area). The mounted scanners are moving over the field to capture the data ($\sim$2,500m^2 area in 90 minutes).

In this study, we used scans from 4 crop species with different growth habits and canopy architectures (two cereals – pearl millet and sorghum, and two legumes – chickpea and mungbean). The crops were sown throughout March-April 2022 and raised with the standard fertilizer inputs under irrigated conditions (every 2-3 days). Each crop was sown in multiple microplots. Altogether, we used scans from 6 different experiments to train the background segmentation algorithms and evaluate plant canopies inference from 3D point clouds, where the canopies' total leaf area (LA) from multiple plants in each microplot was measured destructively (Table 1). Each sector's total leaf area variability was created for the latter by i) different crop densities (2-6 plants/microplot) and ii) sequential crop harvests at four different dates. At these dates, the total crop in each microplot was harvested, and the leaf area was measured destructively by Li-Cor 4100. Altogether, we gathered 275 destructive LA measurements.

B. DATA PREPROCESSING

We used the raw scans obtained from the scanners. The whole preprocessing is visualized in FIGURE 4. Here, the first step involves rotating the scans to align them flat on the x-plane. Both scans are merged into a single file in the second step. This merging process increases the point cloud density in the overlapping areas scanned by both scanners. Therefore, the third step involves a voxelization process (dividing the 3D space into small cubes called voxels, each producing a point cloud value representing all points within it) to rearrange the points in space uniformly. The efficiency of the voxelization has been proven by, e.g., [26], [27], [28]. In this study, for instance, voxelization reduced the number of points in a single scan from 16,735,700 to 2,677,885. This step is performed by a bespoke version of Phenospex's Phena v2 pipeline, which the vendor customized for us to ensure backward compatibility with their systems.

TABLE 1. Overview of the unique experiments, sectors (identified by their barcode numbers), and a number of scanned microplots used throughout the paper for model training and evaluations. The colored arrows correspond to the arrows in FIGURE 2.

Used for	Specie	Experiment IDs	Barcode IDs	Number of microplots
AI model training →	Pearl millet	50,51	28,96	96
	Sorghum	48	84	48
	Chickpea	57	83, 84, 86	216
	Mungbean	59	79 - 92	336
Leaf Area trait evaluation (destructive measurement) → →	Pearl millet	56	178, 179, 180	60
	Sorghum	56	154, 176, 177	72
	Chickpea	56	125, 126, 127	72
	Mungbean	56	151, 152, 153	72
Point counts on daily scans → →	Chickpea	57	80	24

After the scanning process, specific points may be considered outlier values, where the color values differ significantly from the others. Since the developed AI model uses color values, plant, soil, and tray color values must be consistent within themselves. For this purpose, a smoothing process (4^{th} step) was applied to eliminate outlier color values in some points. In this process, each point takes the average color value of the n nearest points. In the cropped voxelized leaf section shown in FIGURE 4, there are approximately 1000 points. Based on this density, the value of N has been set to 250 in this study. In the zoomed-in view of the cropped leaf area in FIGURE 4, the color values of the points before and after the smoothing process are given, demonstrating how the smoothing process ensures that the points on the same leaf have similar color values. This approach is consistent with previous studies, where smoothing techniques have been utilized to reduce noise and irregularities in point cloud color values, improving both visual consistency and feature reliability [29], [30].

C. DATA ANNOTATION AND SPLIT

The plant and background regions were manually marked and separated using Cloud Compare (version 2.10). The annotated sector is depicted in FIGURE 5. This process involved precisely defining the plant's boundaries and background data. Regions identified as plant data were labeled and stored separately from the background data. In instances where

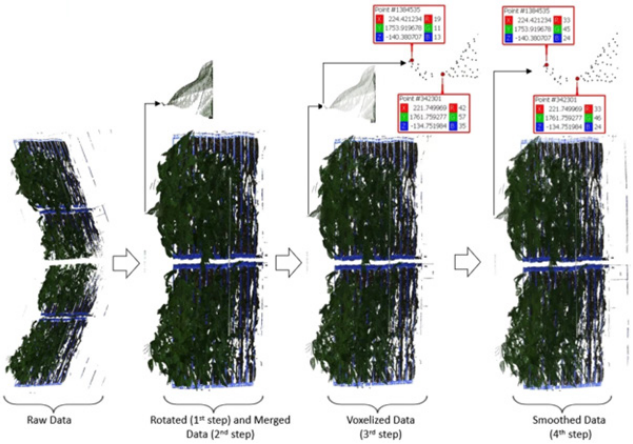

FIGURE 4. Preprocessing steps of the raw scan files to extract only plant data for individual microplots.

certain regions could not be distinguished, these ambiguous areas were excluded from the dataset to avoid model confusion. This process was performed for each plant species and all sectors. In total, we used 696 microplots, as depicted in TABLE 1.

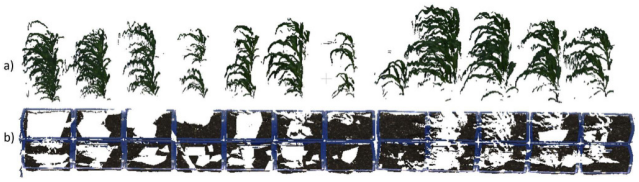

FIGURE 5. The top image (a) shows the manually separated plant data, while the bottom image (b) shows the background data.

Following the annotation process, the dataset was randomly split into three subsets for use in the AI model: 60% for training, 20% for validation, and 20% for testing. These scans, detailed in Table 1, were selected based on the following criteria to ensure a balanced and representative dataset. Since there was an unequal number of available scans for each plant species, fewer files were included for some species, such as sorghum and pearl millet. Additionally, given the differing growth rates of each plant species, scans were typically chosen a few days before harvest, as this period often provides the most representative plant structure for segmentation tasks. For species with fewer available scans, additional files were selected from earlier time points within the same experiment to enhance diversity and representation. These selections were made by considering factors such as plant species, the number of available scans, and growth rates, ensuring a dataset that reflects the variability of plant development.

D. AI MODEL DEVELOPMENT

The experiments were conducted using artificial neural networks, specifically a Multi-Layer Perceptron (MLP) model. Keras hyperparameter tuning was configured using

Bayesian Optimization, with a maximum of 500 trials to search the hyperparameter space efficiently. The hyperparameters considered are presented in TABLE 2. Due to the binary classification task, the output layer consisted of a single neuron responsible for the final prediction (plant/background). The maximum number of epochs was set to 50, and an early stopping method was employed to prevent overfitting and improve generalization. The early stopping was based on validation loss, with a patience period of 2 epochs. The model training was performed separately for three input point data: RGB, RGB+XYZ, and RGB+XYZ+NIR (where RGB represents color channels, XYZ corresponds to 3D positional coordinates, and NIR indicates near-infrared value).

TABLE 2. Range of the hyperparameters used by Keras tuner during the model development.

Hyperparameter	Range of Possible Values (well known)
Number of hidden layers	1 to 3
Neurons per hidden layer	[10, 20, 50]
Activation function	Relu, Sigmoid, Tanh
Optimizer	SGD, RMSProp, ADAM
Input data	RGB, RGB XYZ, RGB NIR XYZ

E. MODEL EVALUATION

The performance of the developed model was evaluated from three perspectives. First, the performance of the MLP model using the annotated point clouds as ground truth; second, the impact of the segmentation on Leaf Area trait estimation using ground truth gathered by the destructive measurement method; third, we provide a comparison using an external data set.

1) MODEL EVALUATION ON ANNOTATED DATASET

For the first case – the performance of the MLP model, we used mainly two well-known metrics: Accuracy (number of correct predictions divided by the total number of predictions) and Confusion matrix (a table displaying the ground truth versus predicted classifications). We also provide Precision and Recall calculated from the latter one. The evaluation used the test set defined in II-C Data annotation and split. We also evaluate the impact of the smoothing process by comparing it to the model trained without this preprocessing step.

2) EVALUATION OF THE IMPACT ON LEAF AREA TRAIT ESTIMATION

For the second case, we used the data that were scanned and harvested for the destructive measurements (ground truth). The data origins are visualized using orange arrows in TABLE 1 and FIGURE 2. In FIGURE 2 there are three paths. The darker one is for the ground truth. Light one as follows: 1) we ran the preprocessing; 2) we ran the model inference for each datapoint to classify it to separate plant and background data (shown at the top of FIGURE 4); 3) the plant data were then sliced up based on the fixed coordinates to get individual microplot data, as illustrated on the left side of FIGURE 6; 4) we predicted the Leaf Area trait using the customized version of Phena provided by Phenospex. For the middle orange, the whole Phena pipeline was used, utilizing the coordinate-based segmentation method. However, as we used the version of the Phena pipeline that is not yet in production, we preprocessed the scans using voxel resolutions from 0.55 to 0.65 in both paths.

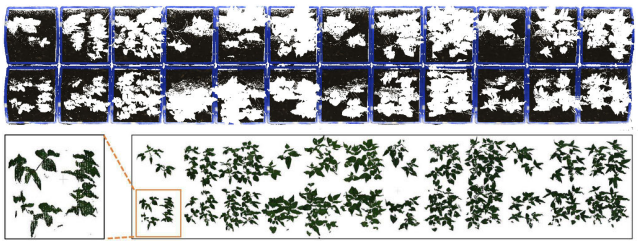

FIGURE 6. Separated background and plant data by the model. Background data are shown in the top half. On the left bottom side, the zoomed-in points show the plant data belonging to a single microplot separated from the full sector scan on the right side. Similarly, the data of all microplots are separated from each other and saved as a separate file for evaluation.

Finally, we compared the leaf area trait measured by the destructive method on each microplot to the estimations computed by the Phena pipeline (using coordinate-based segmentation) and our proposed method (AI-based segmentation). The comparison was evaluated using standard, well-known metrics, particularly R^2 and root mean square error (RMSE). These metrics were calculated for all the voxel resolutions (0.55 to 0.65). Because of the inherent differences between the canopy structures of different crop species and related bias in the destructive measurement of the ground truth, we provide evaluation separately for cereals (pearl millet, sorghum), broad-leaf legumes (mungbean), and narrow-leaf legumes (chickpea).

3) COMPARISON OF THE NUMBER OF PLANT POINTS ON DAILY SCANS

The trained AI model was additionally tested on everyday scans throughout the growing period using a previously unseen chickpea setup (Experiment 57, barcode No. 80). The segmentation was performed using both methods. The scan files were segmented using both methods from the initial sowing day to harvest. The point cloud counts of the separated plant data were recorded for each day. For the coordinate-based method, the segmentation threshold was manually tuned and set at the most precise z-coordinate of 118, corresponding to the microplot's optimal segmentation height.

4) EVALUATION USING AN EXTERNAL DATASET

To show the generalization capacity and effectiveness of the proposed model, we provide an additional evaluation using a dataset from a different domain (adaptability to domain shift). The model was retrained using the 3D Paris-Lille dataset collected by [31] from the streets of Paris and Lille. This dataset

encompasses three point cloud files: Lille1.ply, Lille2.ply, Paris.ply, and 50 distinct classes. The original dataset publication did not provide a baseline classification result; however, in subsequent studies, the results obtained with this dataset have been assessed in different ways. For instance, in the study by Diaz et al. (2021), they trained their model on a dataset they created themselves and used the 3D Paris-Lille dataset solely to evaluate the model's performance. While Diaz et al., in the original study, specify that they used 7 data files for training and 3D Paris-Lille data files for testing, these training datasets were not available in their shared GitHub repository. Therefore, we divided the dataset into 80% for training and validation and 20% for testing to align with our study's needs. In our case, we trained our model from scratch using this training data.

This approach allowed us to proceed with model evaluation, but complicated a direct comparison with the accuracy values reported in their study. While applying the pipeline to this dataset, we only modified the input data to align with the dataset's characteristics. Then, the dataset was processed. Considering these limitations, we evaluated the model's performance by examining the average accuracy instead of directly comparing our results with those of [32].

III. RESULTS

A. FINAL MODEL

The optimal model configuration consists of three hidden layers: 10 units in the first layer and 50 units in both the second and third layers, all using the ReLU activation function. The output layer used the sigmoid activation function, which is suitable for binary classification. The model was trained using the Adam optimizer with binary cross-entropy as the loss function. This configuration, derived from the RGB+XYZ+NIR dataset, demonstrated the best performance. The model's architecture is visualized in FIGURE 7. The final trained model and the source code (preprocessing and model) can be found in a public GitHub repository (https://github.com/serkankartal/MLP-3D-PlantSeg). The repository also contains instructions on how to train the model for custom data.

Optimizing the model with Keras Tuner (for 500 trials) on an NVIDIA A4500 GPU took approximately 10 hours for each dataset. For the soil segmentation task, the trained model processes one point at a time to classify it as plant or background data. As the model is relatively simple, the execution time for one-point prediction is usually in milliseconds, based on the hardware used. Inference of a single scan file (whole sector), including preprocessing, takes about 2 minutes on average, with most time spent on preprocessing rather than the AI model.

B. MODEL EVALUATION ON ANNOTATED DATASET

Two widely recognized metrics were used to evaluate the model's performance on the annotated dataset: the Confusion Matrix and Accuracy. The assessment used a test set

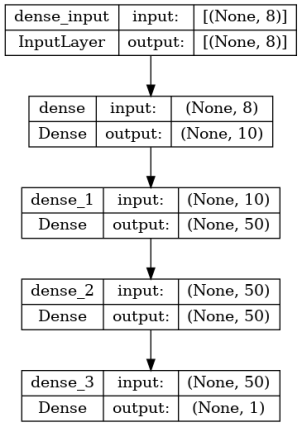

FIGURE 7. The architecture of the MLP (Multi-Layer Perceptron) model used for background segmentation. The input consists of eight neurons representing three RGB channels, one NIR channel, and spatial information (XYZ). After three hidden layers, one output layer represents the final category – background/plant.

comprising 20% of the manually annotated scans. The best model configurations identified during the hyperparameter tuning process were evaluated. TABLE 3 shows a comparison of the models using different input point data.

TABLE 3. Confusion matrices and accuracies to compare various variants of models – input data. The color shades graphically show the differences – the lighter color represents smaller values.

M		Ground Truth		
Input point data	Accuracy	0	1	Model
Unsmoothed RGB + XYZ + NIR	0.9899	18,202,033	169 835	0
		98,471	8,072,451	1
RGB	0.9958	18,233,465	44,276	0
		67,039	8,198,010	1
RGB + NIR	0.9970	18,247,405	41,614	0
		50,646	8,203,125	1
RGB + XYZ	0.9989	18,284,867	14,990	0
		13,828	8,229,105	1
RGB + XYZ + NIR	0.9993	18,296,710	11,159	0
		7,005	8,227,916	1

All the models consistently demonstrated high classification accuracy, with performance exceeding 99% in each case. Although the differences in accuracy between the different inputs were minimal, a closer look at the confusion matrices reveals that the spatial point data (XYZ) and near-infrared spectra (NIR) provided significant information to the model. For example, the false positives decreased significantly from 67,039 in the RGB model to 7,005 in the RGB+XYZ+NIR model, while false negatives dropped from 44,276 to 11,159. This additional input point data improved accuracy and enhanced precision by reducing misclassifications. Therefore, the model trained on the RGB+XYZ+NIR dataset was selected for further use in this study.

TABLE 3 also shows the specific impact of the smoothing preprocessing step. We evaluated the best model trained on unsmoothed RGB + XYZ + NIR input point data. This model produced 98,471 false positives and 169,835 false negatives. The accuracy dropped to 0.9899. This sharp contrast in performance indicates that the lack of smoothing introduced substantial noise into the classification process, leading to a much higher rate of misclassifications. Thus, the smoothing process reduced outlier effects by averaging the color values of neighboring points.

C. EVALUATION OF THE IMPACT ON LEAF AREA TRAIT ESTIMATION

For the second evaluation, we used the destructively measured dataset, as defined in TABLE 3, to compare the estimation of the Leaf Area trait. FIGURE 8 presents the resulting metrics (MAPE and R^2) of selected voxel resolution values. We picked 0.60 as the preprocessing value for each species for the model development, the best value for our AI-based method, and the best value for the coordinate-based method. Results for all species and resolutions can be seen in the supplementary file LA_results.xlsx.

When the results for Mungbean, Pearl millet, and sorghum are examined, it is observed that the error rates in the leaf area indices obtained using the AI-based method we proposed and the traditional method are close to each other (difference to coordinate-based method in R^2 ~0.1-0.2). The main reason is that these three species grow relatively quickly above the tray rim. This fact significantly affects the results. Nevertheless, the predictions made with the AI-based method still provide better results (except for mungbean).

However, when the metric values obtained for the chickpea are examined, it is seen that the AI-based method shows higher differences. For example, at a resolution of 0.62, the R2 value for the AI-based method is 11.35% higher than the best value for the Coordinate-based method at 0.65. This difference demonstrates an improvement in prediction accuracy for the AI-based method, especially when dealing with smaller plants like chickpeas.

A visual comparison of the segmented points is also provided in FIGURE 8. As seen in the figure, the points highlighted in red at the bottom of the image represent the additional plant data selected by the AI-based method, i.e., the points that could be segmented from within the tray. This visualization clearly demonstrates the source of the extra plant data obtained by our proposed AI-based method.

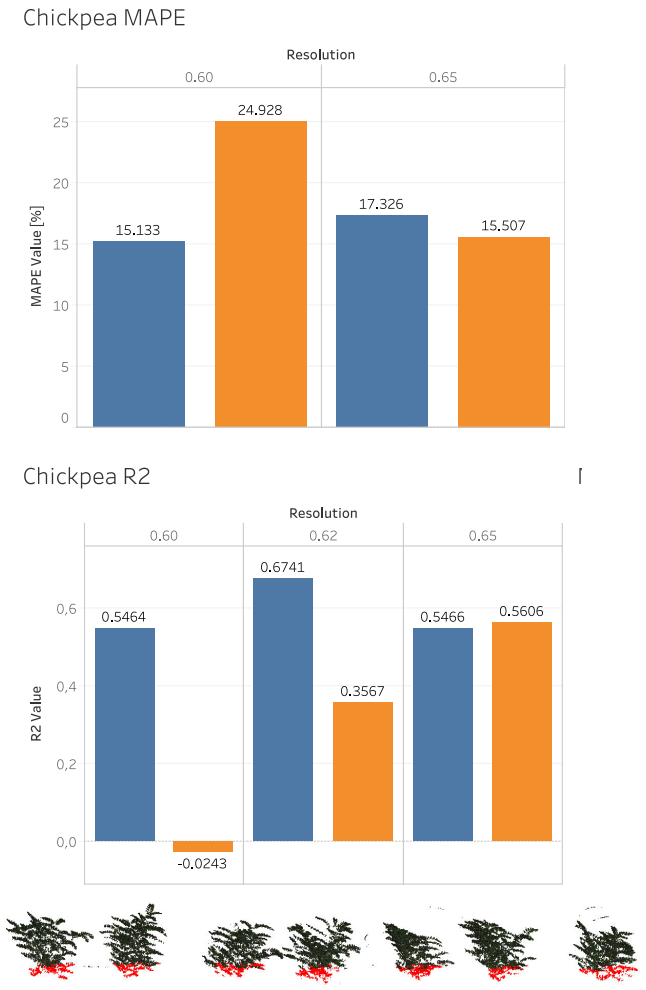

FIGURE 8. Results of Leaf Area trait evaluation using mean absolute percentage error (MAPE) and R2. We show only specific resolution values, particularly 0.60, used in AI-based method development, the best value for model inference, and the best value for the coordinate-based method. Below, there is a visualization of the difference using both methods for segmenting plants. The plants were scanned on the 22nd day from sowing. The red part represents points below the tray rim that are not segmented by the coordinate-based method.

D. COMPARISON OF THE NUMBER OF PLANT POINTS ON DAILY SCANS

The comparison of the point counts is detailed in FIGURE 9. In the early growth stages, when chickpea plants emerged and remained confined within the microplot, the coordinate-based method inaccurately classified barcode and noise data above the microplot as plant data, leading to a count of 6,228 points instead of 0. On Day 9, the AI-based method identified 1,100,423 plant points, while the coordinate-based method identified 994,136 points, resulting in a 10.7% increase in plant data segmentation. On Day 11, this difference was even more pronounced, with the AI-based method identifying 846,792 points compared to 719,871 points segmented by the coordinate-based method, a 17.6% increase. Similarly, on Day 20, the AI-based method segmented 2,463,856 points compared to 2,367,033 points by the coordinate-based method, representing a 4.1% improvement. A decrease in the number of segmented points was caused by the thinning process performed on Days 10 and 11, which removed excess plants.

The coordinate-based method in the early growth stages, when plants have not yet grown above the tray rim, captures noisy data, such as that of irrigation pipes. Thus, the number of points does not differ from the AI-based method or is even higher for the coordinate-based method.

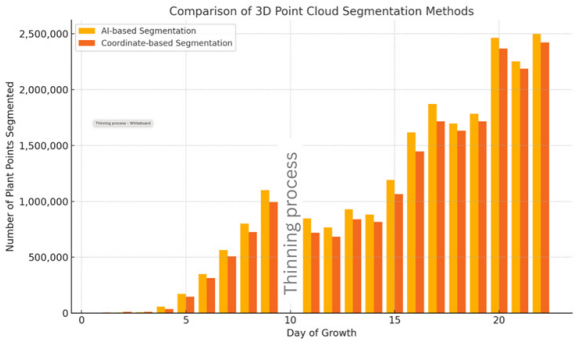

FIGURE 9. Comparison of both segmentation methods in daily plant point cloud counts using an independent chickpea experiment. On day 10, the decrease in count is caused by the manual thinning process, which removes excess plants. The reason for the difference is illustrated in FIGURE.

The AI-based method, on the other side, accurately segmented only the plant data inside the microplot. This highlights the AI-based method's capability to distinguish plant data from background and noise without relying on thresholds or manual adjustments. Such precision is particularly crucial for early-stage plant monitoring, where the traditional methods often fail to detect significant portions of plant data within the microplot.

E. EVALUATION USING THE EXTERNAL DATASET

The results obtained and reported on our pipeline are derived exclusively from the test dataset (as defined in section II-E-II)). The proposed pipeline achieved an average accuracy of 0.98, which is similar to the 0.97 accuracy reported for the entire dataset. Additionally, the result of the separation performed by the model on the Paris Street data is illustrated in FIGURE 10. In this figure, the blue regions represent background data, while the red areas denote other data (trees, cars, streetlights, etc.). The visual analysis effectively showcases the model's success in separating the ground from non-ground elements, highlighting the robustness and versatility of the AI-based pipeline across various datasets.

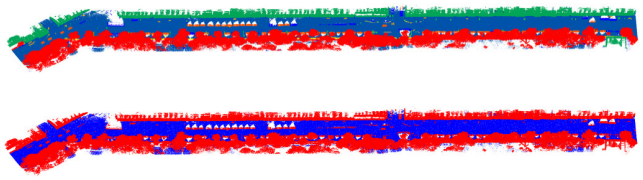

FIGURE 10. The visualization at the top shows the classified data of the original dataset. In contrast, the image at the bottom shows the dataset processed by our model into two categories (ground and non-ground) according to the format suitable for our study.

IV. DISCUSSION

A. SEGMENTATION FOR PLANT PHENOMIC TASKS

The findings of this study highlight key challenges in 3D plant phenotyping related to background segmentation in point cloud data obtained from high-throughput laser-based scanners. Traditional segmentation methods rely on height coordinates. Those methods often result in significant data loss, particularly for early-stage plants or prostrate canopy architectures. Additionally, commonly used algorithms like Region Growing Segmentation (RGS) and Random Sample Consensus (RANSAC) may not generalize well across different experimental setups.

Existing solutions in automotive and robotics applications [4], [5], [6], [7], [8], [9], [10], [11], [12] focus on flat surfaces and dynamic environments. This makes them less suitable for plant phenotyping. In such contexts, soil surfaces are often uneven, and additional background noise is present..

Computer vision-based methods, e.g., [13], [14], [15], [16], are very complex and rely on labor-intensive annotation. Training those models requires more and wider input data and computational power. Simple models, as the presented MLP-based, usually better adapt to concept drifts and data domain shifts, as shown in 3-E. Moreover, the model classifies one point at a time. Adapting it to different scanner types (e.g., UAV-based, LiDAR-based, photogrammetry-based, NERF-based) should be smoother.

B. SIGNIFICANCE OF THE RESULTS

This study presents an AI-driven segmentation method designed to improve the accuracy of traditional coordinate-based segmentation methods while addressing the above-mentioned issues. The proposed method has been rigorously evaluated through AI model performance metrics, comparison with coordinate-based segmentation, canopy trait inference, and external dataset validation. The results demonstrate its potential to enhance plant feature extraction in phenotyping workflows.

We have shown the results of a simple MLP-based model. The RGB+XYZ+NIR input point data configuration achieved a classification accuracy of 0.9993, with only 7,005 false positives (points) and 11,159 false negatives, significantly outperforming more straightforward configurations. This result highlights the importance of integrating spatial and spectral data to enhance model precision and reliability. Besides the segmentation model, we also implemented a preprocessing step, smoothing to reduce noise and irregularities in point cloud color values. The smoothing step improved the model's accuracy by 0.0094%, reducing 87,312 false positives and 158,676 false negatives.

The method's impact on the accuracy of crops' leaf area estimation was evident. The accuracy of the algorithms presented for leaf area estimation in crops such as pearl millet and mungbean was comparable to that of the traditional coordinate-based method. This is because these crop seedlings grow rapidly in a vertical direction above the tray rim into the space where coordinate-based algorithms capture the canopy. In smaller crop species with prostrate growth habits, such as chickpeas, there were significant improvements in accuracy in capturing the canopy area. For chickpeas, the hereby proposed method achieved an R^2 value 0.1135 higher than the coordinate-based approach. The model segmented 7% more plant data that were below

the tray rim. Such a magnitude of differences in canopy size for small and prostrate crops like chickpeas is important for crops' capacity to adapt to abiotic stresses like drought [25], [33]. This indicates the hereby presented method significantly improves crop evaluation on PSX's platforms, such as LeasyScan, particularly the ones with small and prostrate canopy structures. This was demonstrated in the case of chickpeas. Finally, we expect that the accuracy gains can be even higher for crops like peanuts or wild crop relatives.

C. SOURCE CODE AND HOW TO USE THE MODEL

The hereby reported algorithm is available on the GitHub repository (https://github.com/serkankartal/MLP-3D-PlantSeg). The model is simple to reuse, retrain, or fine-tune for similar types of data. Instructions are in the repository README file, including the specific information for users with data from Phenospex's PlantEye scanner. We aim for a dynamic, continuous, feedback-based future development. Any feedback is welcome by e-mail to the corresponding author or via Issues on the GitHub platform, where everyone can see answers.

D. GENERALIZATION CAPACITY

The model also demonstrated its generalization capacity when tested on an external source of 3D point cloud data – the Paris-Lille 3D dataset by [31]. It achieved an accuracy of 0.98, which is similar to the 0.97 reported by [32]. This evaluation needed model retraining as the external dataset represents a different and broader range of data (trees, streets, sidewalks, traffic signs, lightning, etc.). Despite the retraining need, the results confirm the robustness and adaptability of the proposed model to diverse datasets. This makes it suitable for applications beyond agriculture/phenomics. The model, in general, can be used for any 3D point cloud data.

E. LIMITATIONS AND FURTHER IMPROVEMENTS

While the proposed method shows substantial accuracy (see section C) and reasonable performance, certain limitations remain. Although we selected species to cover a wide range of crop canopy types (grain legumes and cereals), the model was trained and evaluated on a limited number of species. Additional improvements to robustify and generalize the model, especially for more species, canopy types, and different types of trays, can be performed to transfer the method into the production-ready phase. Specifically, further resolution (voxel sizes) optimization for specific (small-sized, prostrate) crop types or growth stages could enhance accuracy. Using additional training data with more diverse crop types (e.g., peanut or wild crop progenitors) might improve the model's generalization capacity.

V. CONCLUSION

This study presents a novel AI-driven (MLP-based) method for background segmentation in 3D plant phenotyping that addresses the limitations of traditional coordinate-based methods. Using a Multi-Layer Perceptron with integrated RGB, spatial (XYZ), and near-infrared (NIR) inputs, our approach achieves a classification accuracy of 0.9993.

We innovatively used the smoothing process in the preprocessing stage, effectively reducing noise by averaging color values across neighboring points (accuracy increased by 0.0094). This enhancement, combined with integrating spatial and spectral data, leads to a more reliable separation of plant and background regions, particularly for small or prostrate species where a particular part of the plant is below the growing tray.

The improved segmentation accuracy directly benefits trait estimation; for example, our method increased the R^2 value by 0.114% in chickpea crops and captured an additional 7.01% of plant data compared to the traditional coordinate-based method. We also showed the model's ability to adapt to domain shift by retraining using the external dataset, confirming its generalization capacity.

Overall, this paper offers an efficient novel method usable for automating plant trait analysis in high-throughput phenotyping. The complete model and code are available in our public GitHub repository for possible feedback and future development.

ACKNOWLEDGMENT

The authors thank Phenospex (the manufacturer of the PlantEye 3D scanners), mainly Thorsten Karrer, Christiaan Vonk, and András Tóth, for providing a customized version of the Phena pipeline that allowed the Leaf area trait evaluation.

REFERENCES

[1] T. Meraj, M. I. Sharif, M. Raza, A. Alabrah, S. Kadry, and A. H. Gandomi, "Computer vision-based plants phenotyping: A comprehensive survey," *iScience*, vol. 27, no. 1, Jan. 2024, Art. no. 108709, doi: 10.1016/j.isci.2023.108709.

[2] F. Okura, "3D modeling and reconstruction of plants and trees: A cross-cutting review across computer graphics, vision, and plant phenotyping," *Breeding Sci.*, vol. 72, no. 1, pp. 31–47, 2022, doi: 10.1270/jsbbs.21074.

[3] X.-F. Han, J. S. Jin, M.-J. Wang, W. Jiang, L. Gao, and L. Xiao, "A review of algorithms for filtering the 3D point cloud," *Signal Process., Image Commun.*, vol. 57, pp. 103–112, Sep. 2017, doi: 10.1016/j.image.2017.05.009.

[4] J. Cheng, D. He, and C. Lee, "A simple ground segmentation method for LiDAR 3D point clouds," in *Proc. 2nd Int. Conf. Adv. Comput. Technol., Inf. Sci. Commun. (CTISC)*, Mar. 2020, pp. 171–175.

[5] S. Choi, J. Park, J. Byun, and W. Yu, "Robust ground plane detection from 3D point clouds," in *Proc. 14th Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2014, pp. 1076–1081, doi: 10.1109/ICCAS.2014.6987936.

[6] W. Huang, H. Liang, L. Lin, Z. Wang, S. Wang, B. Yu, and R. Niu, "A fast point cloud ground segmentation approach based on coarse-to-fine Markov random field," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7841–7854, Jul. 2022, doi: 10.1109/TITS.2021.3073151.

[7] K. Liu, W. Wang, R. Tharmarasa, J. Wang, and Y. Zuo, "Ground surface filtering of 3D point clouds based on hybrid regression technique," *IEEE Access*, vol. 7, pp. 23270–23284, 2019, doi: 10.1109/ACCESS.2019.2899674.

[8] P. Narksri, E. Takeuchi, Y. Ninomiya, Y. Morales, N. Akai, and N. Kawaguchi, "A slope-robust cascaded ground segmentation in 3D point cloud for autonomous vehicles," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 497–504, doi: 10.1109/ITSC.2018.8569534.

[9] Y. Qian, X. Wang, Z. Chen, C. Wang, and M. Yang, "Hy-seg: A hybrid method for ground segmentation using point clouds," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 2, pp. 1597–1606, Feb. 2023, doi: 10.1109/TIV.2022.3187008.

[10] Z. Shen, H. Liang, L. Lin, Z. Wang, W. Huang, and J. Yu, "Fast ground segmentation for 3D LiDAR point cloud based on jump-convolution-process," *Remote Sens.*, vol. 13, no. 16, p. 3239, Aug. 2021, doi: 10.3390/rs13163239.

[11] H. Vu, H. T. Nguyen, P. M. Chu, W. Zhang, S. Cho, Y. W. Park, and K. Cho, "Adaptive ground segmentation method for real-time mobile robot control," *Int. J. Adv. Robotic Syst.*, vol. 14, no. 6, Nov. 2017, Art. no. 172988141774813, doi: 10.1177/1729881417748135.

[12] H. Vu, H. T. Nguyen, P. Chu, S. Cho, and K. Cho, "A ground segmentation method based on gradient fields for 3D point clouds," in *Lecture Notes in Electrical Engineering*. Singapore: Springer, Dec. 2017, pp. 388–393, doi: 10.1007/978-981-10-7605-3_64.

[13] K. Mirande, C. Godin, M. Tisserand, J. Charlaix, F. Besnard, and F. Hétroy-Wheeler, "A graph-based approach for simultaneous semantic and instance segmentation of plant 3D point clouds," *Frontiers Plant Sci.*, vol. 13, Nov. 2022, doi: 10.3389/fpls.2022.1012669.

[14] L. Wang, L. Zheng, and M. Wang, "3D point cloud instance segmentation of lettuce based on PartNet," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1646–1654, doi: 10.1109/cvprw56347.2022.00171.

[15] R. Zhang, Y. Wu, W. Jin, and X. Meng, "Deep-learning-based point cloud semantic segmentation: A survey," *Electronics*, vol. 12, no. 17, p. 3642, Aug. 2023, doi: 10.3390/electronics12173642.

[16] J. Zhou, X. Fu, S. Zhou, J. Zhou, H. Ye, and H. T. Nguyen, "Automated segmentation of soybean plants from 3D point cloud using machine learning," *Comput. Electron. Agricult.*, vol. 162, pp. 143–153, Jul. 2019, doi: 10.1016/j.compag.2019.04.014.

[17] K. Zarzyńska, D. Boguszewska-Mańkowska, and A. Nosalewicz, "Differences in size and architecture of the potato cultivars root system and their tolerance to drought stress," *Plant, Soil Environ.*, vol. 63, no. 4, pp. 159–164, Apr. 2017, doi: 10.17221/4/2017-pse. [Online]. Available: https://pse.agriculturejournals.cz/doi/10.17221/4/2017-PSE.html

[18] S. Kartal, S. Choudhary, J. Masner, J. Kholová, M. Stočes, P. Gattu, S. Schwartz, and E. Kissel, "Machine learning-based plant detection algorithms to automate counting tasks using 3D canopy scans," *Sensors*, vol. 21, no. 23, p. 8022, Dec. 2021, doi: 10.3390/s21238022.

[19] S. Kartal, S. Choudhary, M. Stočes, P. Šimek, T. Vokoun, and V. Novák, "Segmentation of bean-plants using clustering algorithms," *Agris On-Line Papers Econ. Informat.*, vol. 12, no. 3, pp. 36–43, Sep. 2020, doi: 10.7160/aol.2020.120304.

[20] L. Garcia Ugarriza, E. Saber, S. R. Vantaram, V. Amuso, M. Shaw, and R. Bhaskar, "Automatic image segmentation by dynamic region growth and multiresolution merging," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2275–2288, Oct. 2009, doi: 10.1109/TIP.2009.2025555.

[21] U. Weiss and P. Biber, "Plant detection and mapping for agricultural robots using a 3D LiDAR sensor," *Robot. Auto. Syst.*, vol. 59, no. 5, pp. 265–273, May 2011, doi: 10.1016/j.robot.2011.02.011.

[22] Phenospex. *PlantEye F600 Multispectral 3D Scanner for Plants-PHENOSPEX*. [Online]. Available: https://phenospex.com/products/plant-phenotyping/planteye-f600-multispectral-3d-scanner-for-plants/

[23] Phenospex. *PlantEye F600 Multispectral 3D Scanner for Plants—PHENOSPEX*. Accessed Jun. 30, 2025. [Online]. Available: https://phenospex.com/products/plant-phenotyping/planteye-f600- multispectral-3d-scanner-for-plants/

[24] M. Tharanya, J. Kholova, K. Sivasakthi, D. Seghal, C. T. Hash, B. Raj, R. K. Srivastava, R. Baddam, T. Thirunalasundari, R. Yadav, and V. Vadez, "Quantitative trait loci (QTLs) for water use and crop production traits co-locate with major QTL for tolerance to water deficit in a fine-mapping population of pearl millet (*Pennisetum glaucum* l. R.Br.)," *Theor. Appl. Genet.*, vol. 131, no. 7, pp. 1509–1529, Jul. 2018, doi: 10.1007/s00122-018-3094-6.

[25] K. Sivasakthi, M. Thudi, M. Tharanya, S. M. Kale, J. Kholová, M. H. Halime, D. Jaganathan, R. Baddam, T. Thirunalasundari, P. M. Gaur, R. K. Varshney, and V. Vadez, "Plant vigour QTLs co-map with an earlier reported QTL hotspot for drought tolerance while water saving QTLs map in other regions of the chickpea genome," *BMC Plant Biol.*, vol. 18, no. 1, Dec. 2018, doi: 10.1186/s12870-018-1245-1.

[26] H. Yang, S. Zhang, X. Han, B. Zhao, Y. Ren, Y. Sheng, and X. Zhang, "Denoising of 3D MR images using a voxel-wise hybrid residual MLP-CNN model to improve small lesion diagnostic confidence," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2022, pp. 292–302, doi: 10.1007/978-3-031-16437-8_28.

[27] Y. Liu, Z. Yang, J. Tong, J. Yang, J. Peng, L. Zhang, and W. Cheng, "ET-PointPillars: Improved PointPillars for 3D object detection based on optimized voxel downsampling," *Mach. Vis. Appl.*, vol. 35, no. 3, pp. 1–13, May 2024, doi: 10.1007/s00138-024-01538-y.

[28] C. Lv, W. Lin, and B. Zhao, "Approximate intrinsic voxel structure for point cloud simplification," *IEEE Trans. Image Process.*, vol. 30, pp. 7241–7255, 2021, doi: 10.1109/TIP.2021.3104174.

[29] I. C. Engin and N. H. Maerz, "Investigation on the processing of LiDAR point cloud data for particle size measurement of aggregates as an alternative to image analysis," *J. Appl. Remote Sens.*, vol. 16, no. 1, Feb. 2022, doi: 10.1117/1.jrs.16.016511.

[30] H. Wang, F. Chen, W. Liu, and X. Zeng, "Unfolding gradient graph regularization for point cloud color denoising," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, 2024, pp. 565–579, doi: 10.1007/978-981-97-8508-7_39.

[31] X. Roynard, J.-E. Deschaud, and F. Goulette, "Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification," *Int. J. Robot. Res.*, vol. 37, no. 6, pp. 545–557, May 2018, doi: 10.1177/0278364918767506.

[32] N. Diaz, O. Gallo, J. Caceres, and H. Porras, "Real-time ground filtering algorithm of cloud points acquired using terrestrial laser scanner (TLS)," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102629, doi: 10.1016/j.jag.2021.102629.

[33] M. Zaman-Allah, D. M. Jenkinson, and V. Vadez, "Chickpea genotypes contrasting for seed yield under terminal drought stress in the field differ for traits related to the control of water use," *Funct. Plant Biol.*, vol. 38, no. 4, p. 270, 2011, doi: 10.1071/fp10244.

SERKAN KARTAL received the B.E. and Ph.D. degrees in computer engineering from Çukurova University, in 2010 and 2017, respectively. He is currently an Associate Professor with the Department of Computer Engineering, Çukurova University. He conducted a Postdoctoral Research with Delft University of Technology and the Czech University of Life Sciences Prague, focusing on artificial intelligence applications in remote sensing and agriculture. He is actively involved in several national and international research and development projects, including EU Horizon-funded initiatives in fisheries monitoring and sustainable agriculture. His research interests include deep learning, computer vision, and AI-powered environmental monitoring.

JAN MASNER received the Ph.D. degree in information management from Czech University of Agriculture, Prague, and the Habilitation degree in systems engineering and informatics, in 2024. He is an expert in information management and modern information technologies, particularly UX and AI. His extensive professional background, including participation in research projects funded by both Czech and international grant agencies and contractual research focused on digitization, user experience (UX), usability, and AI, while actively contributing to the development of innovative solutions in the digitization of public administration and precision agriculture, attests to his professional qualifications. Since then, he has been an Associate Professor with the Department of Information Technologies, where he not only imparts his knowledge to students but also develops research activities with an emphasis on UX, usability, and AI. His research focuses on bridging information technologies with the life sciences. He deals in detail with computer vision, the integration of time series with computer vision methods, and the detection of plants and their organs from 3D scans, with a particular focus on innovative approaches to data augmentation.

JANA KHOLOVÁ received the M.Sc. and Ph.D. degrees in plant genetics and physiology from Charles University, Prague, in 2006 and 2010, respectively. In her current position at CZU and CATRIN, she coordinates trans-disciplinary, international research teams and integrate relevant tools and technologies to help communities address global challenges related to the sustainability of agricultural crop production systems in the face of climate change. She leads teams that support crop improvement programs, genebanks, agro-tech industries, primary crop producers, and basic research within academia, both in developing countries and globally.

ALEXANDER GALBA is currently pursuing the Ph.D. degree in systems engineering and informatics with the Faculty of Mathematics and Physics, Charles University, in the field of theoretical cybernetics. At the Department of Information Technologies, Faculty of Life Sciences, Czech University of Life Sciences Prague, he is engaged in research in the field of using computer vision in plant phenotyping. Primarily, he is developing advanced algorithms for processing 3D data for use in the field of AI. His professional focus is on the area of connecting modern information technologies with practice in agriculture. He participates in projects of a data platform for storing and sharing research data, creating algorithms for modern 3D augmentation methods, and using neural networks to solve phenotyping problems. He applies his research experience in teaching, where he focuses on foreign students and teaching English in subjects on the topic to information systems and programming. In addition to academic activities, he is also involved in commercial projects in the fields of information systems design, infrastructure, and applications. He holds certifications from HPE and Microsoft. He received the Microsoft Silver Award for his web application reservation system project.

THARANYA MURUGESAN received the Ph.D. degree in biotechnology from Bharathidasan University, India, with her Ph.D. work focused on drought adaptation in pearl millet through integrated physiological, molecular, and genetic approaches. She is currently an Associate Scientist with ICRI-SAT, India. She is a Crop Physiologist with over a decade of research experience. She has also pioneered the use of computer tomography imaging for rapid, non-destructive post-harvest phenotyping in crops like groundnut and rice. She has authored 20 research articles, four book chapters, and presented them at numerous national and international platforms. In recognition of her research excellence, she has been awarded the DST-SERB National Postdoctoral Fellowship and the Doreen Margaret Mashler Award.

REKHA BADDAM received the B.Sc. degree in statistics and the M.Sc. degree in mathematics, in 2002 and 2006, respectively. She is currently a Senior Scientific Officer with ICRI-SAT, specializing in crop physiology. With a strong analytical background and over a decade of experience, she contributes to research supporting agricultural sustainability and crop improvement. She is also a Drone Pilot helping in research through drones.

VOJTĚCH MIKEŠ is currently pursuing the Ph.D. degree in systems engineering and informatics with Czech University of Life Sciences Prague. His research is centered on the application of data science in the domain of plant phenotyping and stress detection. He is actively engaged in international collaborative projects, including partnerships with the International Crops Research Institute for the Semi-Arid Tropics (ICRI-SAT), India, and Çukurova University, Türkiye, focusing on advanced analysis of 3D plant models and phenotypic traits under environmental stress. In 2024, he undertook a research stay at the University of Pisa, Italy, where he contributed to a project aimed at forecasting the occurrence of Sterility Mosaic Disease using high-throughput phenotyping data and computational modeling. He possesses strong technical proficiency in tools and platforms, such as Python, Spark, AWS, and Databricks. His interdisciplinary background combines expertise in systems engineering, informatics, and applied data analytics, positioning him at the intersection of computer science and agricultural research.

EVA KÁNSKÁ is currently an Assistant Professor with the Department of Information Technology, Faculty of Economics and Management, Czech University of Life Sciences Prague. In her research activities, she focuses primarily on using information and communication technologies in agriculture, especially in data collection and analysis, information system design, and their application in practice. She is involved in more than seventeen research and educational projects, both at the national and international levels (Horizon 2020, Horizon Europe, Digital, and Erasmus+ programs). Within these projects, she creates teaching materials, research activities, and professional practice cooperation. An essential part of her pedagogical work is teaching in English, aimed at international students. She lectures on subjects related to information systems, programming, and digital technologies. In her teaching, she uses knowledge from current research and emphasizes connecting theory with practical use in the real environment. She is an Active Member of Czech Society for Information Technology in Agriculture (CSITA).