MLP-Based Classification of Multispectral Point Clouds for Digital Agriculture

Isabella Subtil Norberto[1], Clodoaldo Souza Faria Junior[2], Antonio Maria Garcia Tommaselli[1], Mauricio Galo[1], Rahuan Miguel da Silva[1]

[1]Faculty of Science and Technology, São Paulo State University (UNESP) at Presidente Prudente, São Paulo 19060-900, Brazil – (isabella.subtil, a.tommaselli, mauricio.galo, rahuan.miguel)@unesp.br
[2] Aeronautics Institute of Technology (ITA), São José dos Campos at São Paulo 12228-900, Brazil – clodoaldo.junior.101406@ga.ita.br

Keywords: deep learning, 3D cloud, crop characterization, precision agriculture, terrestrial.

Abstract

High-resolution monitoring of individual plants is crucial for improving decision-making processes in precision agriculture, particularly when it comes to assessing development, nutrition, and health status. Deep Convolutional Neural Networks (DCNNs) have proven to be highly effective in classifying vegetation components from point cloud data based on geometric features. Combining radiometric information with geometric data can further improve classification accuracy. The fusion of LiDAR and spectral data has proved effectiveness for detailed plant discrimination. However, some challenges remain in fusing terrestrial LiDAR and multispectral data, with few studies focusing exclusively on ground-based sensor integration for plant-level classification. In this study, we propose using a Multi-layer Perceptron (MLP) architecture to classify terrestrial multispectral and LiDAR point cloud data collected around an apple tree. Despite using a lightweight architecture with fewer parameters compared with architectures described in the literature, our approach achieved accuracies exceeding 94%, comparable to state-of-the-art methods. Among the spectral bands evaluated, the combination of image bands near 490 and 735 nm showed the best balance between accuracy and generalisation, consistently discriminating between leaf, wood, and fruit classes with over 90% of accuracy. These results demonstrate the potential of combining terrestrial data fusion with efficient MLP models for achieving precise plant-level classification in precision agriculture.

1. Introduction

Monitoring vegetative development, nutritional status and plant health at an individual plant level is essential for advancing precision agriculture practices (Jayakumari et al., 2021). Substantial progress in crop classification has been achieved with data collected by optical sensors, including multispectral and hyperspectral systems. However, they are mostly effective for coarse-resolution applications, such as identifying crop types or performing plot-level mapping (Avola et al., 2019). While well suited for large-area monitoring, these systems often lack the structural detail required for fine-scale analysis at plant-level (Handique et al., 2017).

In contrast, terrestrial laser scanning (TLS) platforms, based on LiDAR (Light Detection and Ranging), provide dense three-dimensional (3D) data that can capture complex vertical and horizontal plant structure, enabling a more detailed representation of crop architecture (Murray et al., 2020). However, LiDAR data are frequently sparse and lack textural information, which complicates the processes of segmentation and classification (Eckart et al., 2018). Teixeira et al. (2023) discussed the challenges of classifying vegetation, concluding that the performance of classification models based on LiDAR data depends heavily on factors related to geometric resolution, annotation quality, and interference from non-vegetative elements.

Recent efforts have focused on reducing these limitations through composite modelling techniques and the fusion of multiple data sources to improve classification under field variability (Bellakaout et al., 2016, Reji and Nidamanuri, 2023). These approaches have shown promising results, particularly when integrating LiDAR with multispectral imagery from satellite platforms, which benefit from broad spatial coverage and accessibility (Reji and Nidamanuri, 2023). However, while effective at medium scales, these combinations often lack the resolution needed for detailed plant-level discrimination.

Meanwhile, Deep Learning (DL) methods have emerged as powerful tools for processing complex agricultural data (Lawin et al., 2017). However, many of these approaches have been adapted from indoor vision applications (Arrizza et al., 2024) and are based predominantly on widely available remote sensing data types, particularly satellite and aerial imagery (Teixeira et al., 2023). There is a recognised research gap in the use of terrestrial sensor data, in particular, those resulting from the fusion of LiDAR with multispectral imagery captured from ground-based platforms. Despite their potential to provide highly detailed structural and spectral information at the level of individual plants, such integrated datasets remain underexploited in crop classification studies.

This paper investigates the feasibility and assessment of using Multi-layer Perceptron (MLP)-based classifiers applied to point clouds acquired with terrestrial sensors. We assess the contribution of different spectral bands and evaluate the added value of spectral information compared to geometry-only approaches.

2. Background

2.1 Multispectral Terrestrial LiDAR Data

Three-dimensional multispectral point clouds combine spatial coordinates and spectral information, enabling the integrated analysis of the physiological and biochemical characteristics of plants (Behmann et al., 2016). While this approach is useful for phenotyping, acquiring 3D multispectral points is challenging due to the complex structure of the canopy and the interaction between light and plants, both of which affect geometric and spectral accuracy (Behmann et al., 2015).

Traditional acquisition methods based on a fixed camera station or from a single Point of View (PoV) are susceptible to issues such as occlusion and poor signal quality, due to limitations in the field of view and depth of field (Liang et al., 2013). Furthermore, multimodal data fusion is limited by compatibility issues and the challenges of large-scale processing (Sun et al., 2022).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-271-2026

There are some promising solutions, including a technique known as Next-Best-View (NBV), which estimates the next best point of view to guide adaptive data acquisition (Vasquez-Gomez et al., 2014). However, these kind of solutions still suffer from high computational costs and low generalisability, particularly in machine and deep learning-based approaches (Wu et al., 2019).

2.2 Point Cloud Classification Using Deep Learning

2.2.1 Geometric Classification

Recent studies have explored the use of geometric information solely in point clouds to classify and segment forest components. Han and Sánchez-Azofeifa (2022) proposed a method involving multivariate time series of geometric features extracted from multiple neighbours. They applied Deep Neural Networks (DNN), such as Fully Convolutional Neural Network (FCNN), Long Short-Term Memory (LSTM) and Residual Network (ResNet), to classify leaves and wood in broadleaved trees. Using this approach, they achieved an accuracy rate of up to 96%, and utilised class activation maps to interpret the model's decisions. This demonstrated the superiority of multivariate series over univariate ones (67–88%).

Shen et al. (2022) used the PointCNN network to segment terrain, foliage, and trunk points in planted forests, achieving an overall accuracy of approximately 95% and precise extraction of forest parameters. However, this method relies on multiple tools and lacks details on computational efficiency, which could limit its use on a large scale.

Dai et al. (2023) developed MDC-Net (Multi-Directional Collaborative Convolutional Neural Network), a multidirectional convolutional neural network that improves directional coding to reduce occlusions, incorporating prior knowledge to discriminate fine structures. This approach resulted in an accuracy of 97.3% in separating wood and leaves in forest plots. Despite its high performance, the model's complexity increases its computational cost.

2.2.2 Radiometric and Geometric Classification

Methods combining radiometric information, such as Laser Return Intensity (LRI), RGB images and multispectral data, with geometric characteristics have been explored to improve the classification of forest and agricultural point clouds. Wu et al. (2020) developed a CNN-based approach that integrates geometric data and LRI. This approach achieves high accuracy in different forest types: 96.2% for broadleaves, 94.98% for conifers, and over 84.26% for mixed forests. While this approach is effective, optimising the hyperparameters requires empirical adjustments.

In Zhu et al. (2018), seven radiometric features were combined with six geometric features to classify soil, wood, and foliage, achieving an accuracy of 84.4%, which underlines the greater importance of geometric information. Sun et al. (2021) proposed a hybrid approach using intensity for initial classification, and K-nearest neighbours and voxelization techniques for segmentation refinement, resulting in high accuracy (95.5%) and improving computational efficiency.

More recently, Jayakumari et al. (2021) used a modified Deep Convolutional Neural Network (DCNN), called CropPointNet, with RGB data fused with LiDAR for plant segmentation. This model is an adaptation of PointNet for the segmentation of indoor objects. This approach resulted in 90% of accuracy for eggplants and cabbages, and 70% for tomatoes. Reji and Nidamanuri (2023) applied DL to fuse multispectral orbital images with LiDAR point clouds, achieving 92% accuracy in crop classification. Despite variations in model architectures, the authors emphasise that multispectral-LiDAR fusion consistently exhibits superior performance in crop discrimination.

3. Materials and methods

3.1 Materials

3.1.1 Terrestrial Laser Scanner and Camera

The LiDAR point cloud was collected using the FARO Focus Premium TLS, which has a beam divergence of 0.3 mrad, and a 1553.5 nm wavelength. This device captures targets from 0.5 to 70 m with 1 mm precision and an angular accuracy of 19 arcseconds in both axes. Additionally, point cloud colourisation can be performed by integrating data from its internal camera or other sensors, such as multispectral cameras.

In this study, we evaluate the fusion of LiDAR data with imagery obtained with an Agrowing multispectral camera (Agrowing Development Team, 2025), which combines six lenses and spectral filters mounted on a Sony Alpha 7 IVR sensor (ILCE-7RM4A) (Table 1).

Field of View	Diagonal: 45.90°, Horizontal: 35°, Vertical: 26.60°
Spectral bands (nm)	405; 430; 450; 490; 525; 550; 560; 570; 630; 650; 685; 710; 735; 850
Effective focal length (mm)	21.6
Pixel size (mm)	0.0037

Table 1 – General information about the Agrowing multispectral camera.

To each point in the acquired point clouds, Digital Numbers (DN) values were assigned, using selected bands from each Sub-image (S), captured by the six lenses of the Agrowing camera. Each sub-image corresponds to a specific set of spectral bands provided by the sensor lenses and spectral filters, as mentioned in Table 1 and as shown in Figure 1.

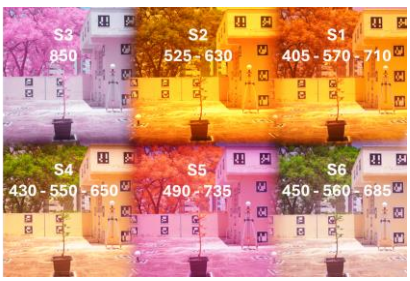

Figure 1 – Example of one image frame acquired by the Agrowing multispectral camera showing six sub-images (S1, S2, S3, S4, S5 and S6) from six lenses.

3.2 Methods

3.2.1 Field Data Acquisition

The dataset was collected at São Paulo State University (Unesp). In the environment field, coded markers were attached to both the wall and floor, serving later as reference and control points. Spheres were also strategically placed to assist with the registration of multiple point cloud scans in an outdoor setting.

Existing previous calibration of the camera, defining the interior orientation parameters, could not be used, because the objective with the lens is removed for storage. TLS Faro scanner and

camera were systematically rotated in two complete 360° turns around the apple tree (height: 1.65 m), with a short interval between acquisitions, maintaining a sensor-to-object distance of approximately 2 m. To investigate the limitations of traditional single PoV acquisitions (Section 2.1), six PoVs were selected to ensure suitable coverage of the tree (Figure 2).

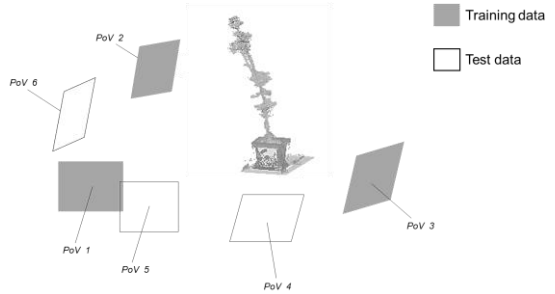

Figure 2 – Distribution of PoVs around one tree.

The model was trained using the data from the first turn, considering three PoVs (1, 2 and 3), while the tests were performed using the MLP architecture with the data exclusively from the second turn (PoVs 4, 5 and 6). This design enables to capture variations in perspective and illumination between the two collections, ensuring independence and accurate performance metric calculations across multiple viewing angles.

3.2.2 Colourisation of the LiDAR point cloud

The adopted methodology is based on the orientation of the sub-images with respect to the point clouds. This is achieved by estimating the camera's position and orientation by bundle adjustment (BA), which enables the estimation of camera position and orientation for each PoV. Spectral values from the images were then projected onto the point cloud, restricted to the points visible from each viewpoint. To mitigate double mapping effects, we applied the method proposed by Katz et al. (2007), which allows for the identification and labelling of occluded points. These occluded points were then excluded from the projection. Each image is therefore treated as an independent observation of the point cloud, corresponding to a distinct point of view.

In total, six PoV were selected based on visual assessment. Three of these, acquired during the first turn around the apple tree, were used for training (PoVs: 1, 2 and 3), while the remaining three comprised the test set (PoVs: 4, 5 and 6). The multispectral point clouds were labelled individually at each PoV.

3.2.3 MLP-Based Classification

The labelling process was performed manually in CloudCompare software using 3D point clouds. Each point was classified into one of three categories: leaf, wood or fruit. Following the terminology commonly used in forestry studies, the term 'wood' was adopted to refer to all non-photosynthetically active components of the vegetation, such as branches and stems. Distinguishing between photosynthetic (leaf) and non-photosynthetic (wood) components is important, as misclassification can significantly affect the estimation of tree attributes (Arrizza et al., 2024).

Feature values for the leaf, wood, and fruit classes were standardised using z-score normalisation applied globally on the training data to ensure uniform scaling across variables and reduce the influence of differing units and magnitudes on model training and to prevent data leakage (Fei et al., 2021). The same normalisation parameters (mean and standard deviation) were subsequently applied to the validation and test sets. Moreover, the class labels were re-indexed starting from zero to conform to standard computational conventions. Semantically, related categories were merged based on expert human judgment, aiming to reduce class imbalance and to streamline the classification task, consolidating the output into three well-defined and meaningful classes. This consolidation not only mitigates potential issues arising from sparse class distributions but also enhances the interpretability and robustness of the classification model.

To address class imbalance in the dataset, random undersampling was applied, which involves the random removal of instances from the majority class, until a more balanced distribution among categories is achieved. This approach reduces the predominance of overrepresented classes, thereby mitigating bias during model training. After balancing, PoVs 1, 2 and 3 were combined into a single dataset and split into 75% for training and 25% for validation. The data acquired from PoVs 4, 5 and 6 was used exclusively for testing, and the quality assessment was reported as the average performance across these three PoVs.

The proposed architecture consists of an MLP-based approach with five hidden layers containing 128, 64, 32, 16, and 8 neurons, respectively. Each hidden layer employs the ReLU activation function and is followed by Batch Normalisation, which mitigates internal covariate shift by normalising mini-batch inputs to have zero mean and unit variance, thereby stabilising and accelerating training (Ioffe and Szegedy, 2015). The ReLU function was chosen due to its ability to mitigate the vanishing gradient problem by maintaining non-zero gradients for positive inputs. Additionally, ReLU function introduces sparsity in each layer by outputting zero for negative inputs, which enhances generalisation and computational efficiency (Nair and Hinton 2010, Glorot et al. 2011). Dropout regularisation was applied selectively after the first two hidden layers with a rate of 30%, and after the final hidden layer with a rate of 20%, helping to prevent overfitting by randomly deactivating neurons during training (Salehin and Kang, 2023). The architecture was determined based on empirical heuristics and iterative experimentation and is presented in Figure 3. The final models have a size of $(n \times 144) + 12{,}028$ parameters, of which 496 are non-trainable, where n represents the number of input features.

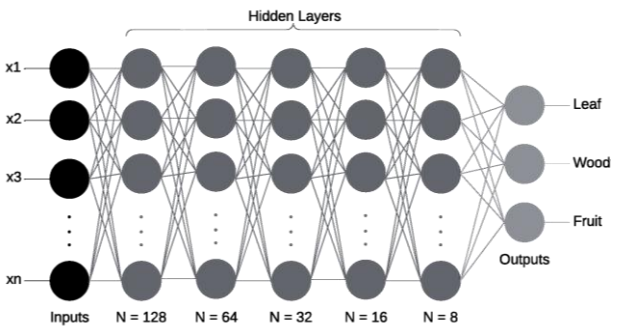

Figure 3 – Schematic representation of a MLP used in this paper. Adapted from Rumelhart et al. (1986).

The output layer comprises three neurons with SoftMax activation, which transforms raw logits into normalised class probabilities, facilitating multi-class classification interpretability. The model was compiled using the Adam optimiser and categorical cross-entropy as the loss function, suitable for multi-class problems. Additionally, the performance metrics monitored during training included accuracy, precision, and recall, allowing for a more comprehensive evaluation of the model's predictive quality. During testing, evaluation metrics included accuracy, precision, recall, and F1-Score.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-271-2026 | © Author(s) 2026. CC BY 4.0 License.

Training was performed with a batch size of 8 (i.e., 2^3) and was configured to run up to a maximum of 10,000 epochs, which served merely as an upper bound to accommodate Early Stopping. Training was automatically halted if no improvement in validation loss was observed for 20 consecutive epochs. This approach reduces the risk of overfitting and enhances computational efficiency by avoiding unnecessary iterations. At the end of the training, the model weights corresponding to the epoch with the best validation performance were restored, ensuring that the final model corresponded to the most generalisable state identified during the training process.

4. Results and Discussions

4.1 Multispectral Point Cloud Generation

4.1.1 Colourisation

The acquisition was performed with a PSOSU (Pixel Size in Object Space Units), or GSD (Ground Sample Distance), of 0.35 mm. After the bundle adjustment, the Root Mean Square Error (RMSE) for the check points, considering the central bands of each sub-image, ranged from 0.41 to 1.16 PSOSU for the planimetric coordinates, and from 0.18 to 0.82 PSOSU for the altimetric component. These values were suitable for generating the multispectral point cloud and ensured geometric accuracy consistent with the research objectives.

The point clouds were separated into point of views, with three PoVs from the first acquisition (PoVs 1, 2 and 3) used in the training–validation split. The second data collection presented different spectral variations compared to the first one, due to changes in lighting conditions (PoVs 4, 5 and 6). The resulting multispectral point clouds are shown in Figure 4.

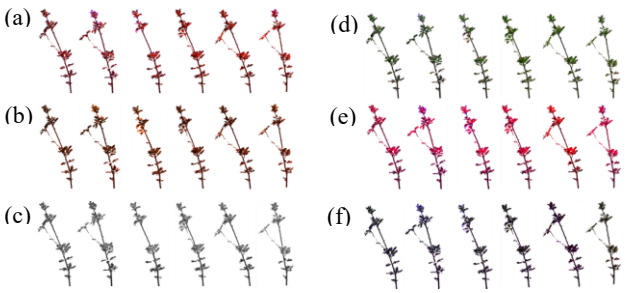

Figure 4 – Multispectral point clouds for each sub-image sets: (a) S1 (Red Edge, G, B), (b) S2 (RG), (c) S3 (Infrared), (d) S4 (RGB), (e) S5 (Red Edge, B), and (f) S6 (RGB). In each case, the first 3 are from the PoVs:1,2, and 3 (used from training-validation) and the last 3 from PoVs: 4, 5, and 6 (for model prediction).

4.2 Classification

4.2.1 Model Analysis

The number of training epochs varied across multispectral point cloud datasets: S1 (112 epochs), S2 (65), S3 (80), S4 (42), S5 (75), and S6 (42). For comparison, the XYZ dataset was trained for 70 epochs while the full multispectral dataset (All Bands – AB) was trained for 75 epochs.

As Karp (2024) previously discussed, the relationship between training loss and validation loss can provide insights into a model's generalisation capability. Closely aligned values typically reflect better performance on unseen data. Figure 5 presents the loss gap (difference between training and validation losses) for each epoch.

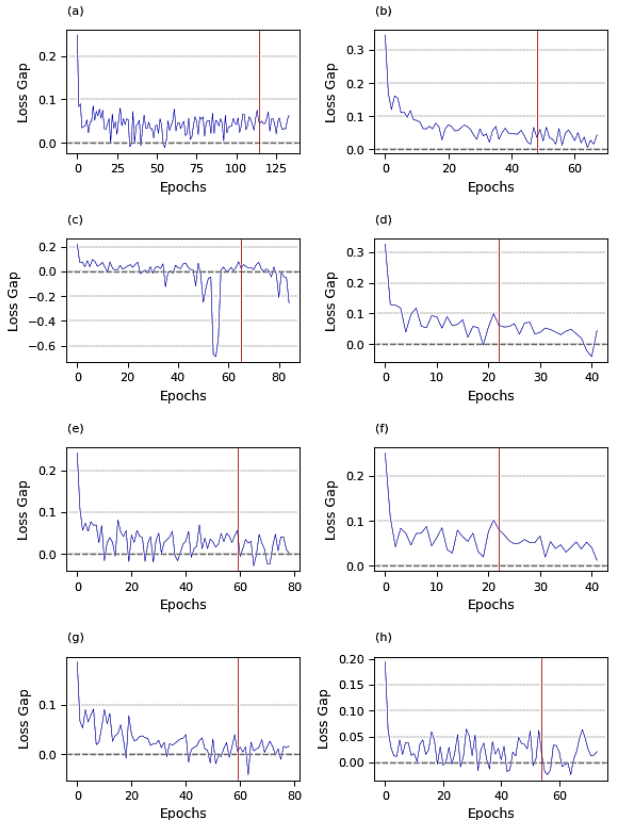

Figure 5 – Loss gap (blue line) and final loss gap considered by early stop (red line) analysis of datasets: (a) S1, (b) S2, (c) S3, (d) S4, (e) S5, (f) S6, (g) AB and (h) XYZ.

In our experiments, we observed variations in the final loss gap considered by early stop across the datasets S1–S6, ranging from 0.0259 (S3) to 0.0811 (S6). Dataset S3 exhibited the lowest final loss gap considered by early stop (0.0259), suggesting that the model achieved good generalisation, with the validation loss closely following the training loss. Conversely, S6 presented the highest final loss gap considered by early stop (0.0811), indicating a larger discrepancy between training and validation, which may reflect overfitting or instability in the training process. Intermediate values were observed for datasets S1, S2, S4, and S5 (0.0483, 0.0356, 0.0619, and 0.0562, respectively), showing moderate generalisation performance.

The dataset AB, which integrates all available features, yielded a very low final loss gap, considering an early stop of 0.0062, demonstrating good stability and balance between training and validation. Similarly, the dataset XYZ, which uses only geometric information, presented a slightly higher final loss gap considering an early stop of 0.0139, indicating that while the absence of additional features slightly impacts generalisation, the model still maintains stable performance. Overall, the final loss gap considered by early stop analysis suggests that the model training was stable across most datasets, with early stopping effectively preventing significant overfitting.

In the training, the Cohen's Kappa statistic for each dataset was also evaluated. Among the multispectral subsets (S1–S6), S2 and S3 datasets resulted in the highest Kappa values (0.950). The datasets S1 and S5 demonstrated high levels of agreement, with Kappa values of 0.945, followed by S4 (0.940) and S6 (0.935). This indicates that there is reliable discriminative capacity across different perspectives.

The dataset XYZ, which contains only geometric features, yielded the lowest Kappa value (0.875), suggesting that the

absence of spectral information reduces classification reliability. The dataset AB, which integrates all spectral bands, resulted in a Kappa value of 0.9. This outperforms the dataset XYZ but is slightly below the values achieved by most individual multispectral subsets.

4.2.2 Model Prediction

Table 2 provides a comprehensive overview of the performance metrics achieved by the model during the testing phase for the PoVs 4–6, for each individual dataset evaluated.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
S1	89.81	86.75	84.71	84.95
S2	90.86	88.16	86.26	86.50
S3	86.67	84.73	80.00	80.07
S4	94.00	91.24	91.01	91.06
S5	94.24	91.62	91.37	91.44
S6	93.36	90.27	90.04	90.10
XYZ	82.00	81.73	73.00	71.65
AB	90.41	86.41	85.64	85.78

Table 2 – Results analysis of the datasets.

The evaluation of model predictions shows a consistent trend: multispectral point cloud datasets (S1–S6) generally outperform the XYZ and AB datasets across all metrics. Datasets S4 and S5 resulted in the highest F1-scores (91.06% and 91.44%, respectively), demonstrating strong classification performance with a balanced precision and recall. Conversely, datasets S1, S3, and AB resulted in the lowest F1-scores (84.95%, 80.07% and 85.78%, respectively). The XYZ dataset, lacking spectral information, showed moderate performance (F1-score of 71.65%), highlighting the importance of multispectral features for accurate classification. The superior results observed in intermediate datasets, such as S4, S5, and S6, indicate that specific spectral combinations, RGB (S4 and S6) and Red Edge and Blue (S5), yield more informative feature representations than other colour combinations.

Figure 6 shows the confusion matrices for different combinations. Datasets S1 to S6 exhibit a well-balanced classification across the three classes, with highly correct classification rates for 'Fruit', for instance, S4 (95%), S5 (93%) and S6 (94%), indicating the model's effectiveness in detecting this class in these subsets. Datasets S2 and S1 performed well in classifying fruits, with S2 reaching 77% of correct identification and no confusion with the "Wood' class. Dataset S1 also showed high accuracy for 'Leaf' (90%) and 'Wood' (89%), but 24% of 'Fruit' samples were misclassified as 'Leaf'. In contrast, the Datasets S3 and XYZ showed the weakest performance for the 'Fruit' class, correctly identifying only 61% and 39% of samples, respectively. Most errors involved misclassification as 'Leaf', suggesting that the lack of spectral components (in XYZ) or suboptimal input combinations (in S3), limits the discrimination capability. The dataset AB, despite aggregating all spectral combinations, correctly classified 84% of 'Fruit' samples but showed increased confusion with 'Leaf' (16%).

To further investigate class-level misclassification, a post-hoc Dunn's test was applied to compare classification errors among classes. A statistically significant difference was observed between 'Leaf' and 'Fruit' ($p < 0.001$ in all positions), confirming the consistent misclassification of fruits as leaves seen in the confusion matrices (Fig. 6). In contrast, no significant differences were found between 'Leaf' and 'Wood' or between 'Wood' and 'Fruit' ($p > 0.05$), indicating that errors involving 'Wood' components are of similar magnitude to those of the other classes and do not exhibit a distinct pattern.

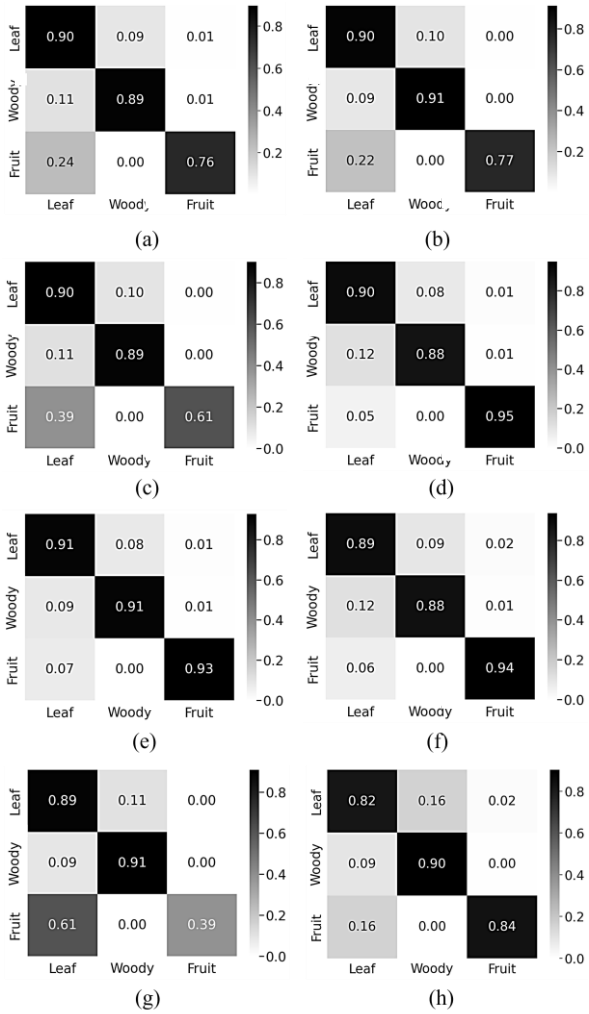

Figure 6 – The confusion matrices, computed as the average of the three test point clouds (PoVs: 4-6) for each dataset, are presented for: (a) S1, (b) S2, (c) S3, (d) S4, (e) S5, (f) S6, (g) XYZ, and (h) for all bands (AB).

4.3 Discussions

The analysis of the boxplots of absolute errors by each PoVs 4 to 6 demonstrates that acquisition geometry had a significant influence on model performance (Fig. 7) in this case study. PoV 5 consistently exhibits greater variability and dispersion across nearly all datasets analysed, with longer upper whiskers, indicating a higher susceptibility to inconsistencies in object representation from this viewing angle.

Figure 7 shows boxplots of test datasets. Each dataset presents distinct behaviour: S1 displays relatively compact error distributions; S2 shows increased variability, specifically in PoV 5; S3 highlights greater asymmetry in PoV 6; and S4 yields the best results for PoVs 4 and 6. Dataset S5 demonstrates the most stable overall performance, with low medians and reduced dispersion. In contrast, S6 exhibits the most critical scenario among the point clouds colourised by sub-images, with median errors ranging from 7,000 to 10,000 and PoV 5 reaching upper whiskers beyond 15,000.

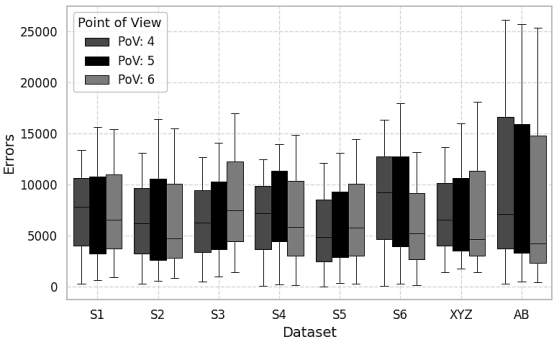

Figure 7 – Boxplot of test dataset.

The dataset XYZ, which lacks spectral information, maintains relatively low median errors; however, PoV 5 still shows higher variability with longer whiskers, suggesting that even in the absence of spectral data, viewpoint geometry affects the error rate. In the AB dataset, which combines all multispectral bands, a high variability is observed across all PoVs. This degradation in performance can be further analysed, but one plausible hypothesis can be associated to inaccuracies in the interior and exterior orientation parameters, which can cause misalignments when all spectral bands were merged.

Furthermore, aside from the differences in error distributions observed across datasets, the proposed method also stands out due to its computational efficiency and ease of use. According to Qi et al. (2017), the Vanilla PointNet architecture comprises approximately 0.8 million parameters, while the full version exceeds 3.5 million. In contrast, the MLP architectures proposed in this study have an average of 13,252 parameters. This corresponds to a reduction of approximately 98.3% and 99.6% in the number of parameters when compared to the Vanilla and full versions of PointNet, respectively. The significant parameter reduction underscores the efficiency and practicality of the proposed models, with minimal impact on classification performance. Furthermore, by operating directly on point-wise data, the method eliminates the need for pre-processing steps, commonly required by CNN-based approaches, such as patch extraction or spatial chunking, further contributing to its practical applicability.

5. Conclusions

The results obtained in this paper demonstrate that the Multi-layer Perceptron approach for classification, utilising geometric information of a terrestrial LiDAR point cloud, combined with multispectral data, resulted in levels of precision competitive with those of reference methods in the literature. Datasets S4 and S5 resulted in accuracies exceeding 94%, values close to the 95-97% reported by studies employing specialised architectures such as CNNs (Wu et al., 2020; Dai et al., 2023).

Reji and Nidamanuri (2023) reported 92% accuracy by fusing satellite multispectral imagery (GSD of 0.5 m) with terrestrial LiDAR point clouds for discriminating cabbage, eggplant, and tomato under different nitrogen levels. Although this represents a different context, involving horticultural crops and lower-resolution satellite imagery, it highlights the potential of integrating structural and spectral information. Our approach exceeded this result with 94.24% (S5), employing exclusively terrestrial data with higher resolution (PSOSU of 0.35 mm). This performance can be attributed to the intrinsic characteristics of terrestrial acquisition. The challenges identified by Reji and Nidamanuri (2023) were that plants of relatively low height do not present robust shapes and profiles for fusion between LiDAR and orbital data did not occur in the present case. Strictly speaking, terrestrial collection allows for more precise delimitation of targets, resulting in better discrimination between leaf, wood, and fruit classes.

Amongst the datasets evaluated, dataset S5 (bands combination: 490 and 735 nm with LiDAR point cloud) demonstrated the best balance between accuracy and generalisation capacity. Although S4 and S6 presented similar results in terms of accuracy, the loss gap analysis indicates that S5 showed less tendency towards overfitting (0.0562) compared to S4 (0.0619) and S6 (0.0811). The dataset S5 proved to be consistent in discriminating the three classes, with more than 90% of accuracy for all classes and stable overall performance.

MLP-based approach demonstrated computational efficiency, as evidenced by low memory demand, and ease of implementation, while still maintaining competitive performance. Many studies on point cloud classification continue to focus on forestry applications. In contrast, research on fruit detection is recent, making this work an important contribution to precision agriculture applications. The S5 dataset model demonstrated greater consistency for this set of apple tree data, but its applicability to other species and conditions requires further investigation.

Recommendations for future research include the implementation of a rigorous geometric camera calibration during field collection (*in-situ* calibration) using more targets and markers. This suggestion is made since the results from dataset AB showed that the colourisation process and, consequently, the model's performance could be affected by the absence of a previous calibration or by movements of leaves and fruits. A more accurate *in-situ* calibration would enable to verify whether increasing the number of spectral bands truly enhances accuracy and improves classification metrics.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) (Grants: 88887.961781/2024-00, 88887.821780/2023-00, 88887.961773/2024-00), and by the São Paulo Research Foundation, FAPESP (Grants: 2021/06029-7).

References

Agrowing Development Team. 2025. (https://agrowing.com/).

Arrizza, S.; Marras, S.; Ferrara, R.; Pellizzaro, G. 2024. Terrestrial Laser Scanning (TLS) for tree structure studies: a review of methods for wood-leaf classifications from 3D point clouds. *Remote Sensing Applications: Society and Environment* 36: 101364.

Avola, G.; Di Gennaro, S.F.; Cantini, C.; Riggi, E.; Muratore, F.; Tornambè, C.; et al. 2019. Remotely sensed vegetation indices to discriminate field-grown olive cultivars. *Remote Sensing* 11: 1242.

Behmann, J.; Mahlein, A.-K.; Paulus, S.; Kuhlmann, H.; Oerke, E.-C.; Plümer, L. 2015. Calibration of hyperspectral close-range pushbroom cameras for plant phenotyping. *ISPRS Journal of Photogrammetry and Remote Sensing* 106: 172–182.

Behmann, J.; Mahlein, A.-K.; Paulus, S.; Dupuis, J.; Kuhlmann, H.; Oerke, E.-C.; et al. 2016. Generation and application of hyperspectral 3D plant models: methods and challenges. *Machine Vision and Applications* 27: 611–624.

Bellakaout, A.; Cherkaoui, M.; Ettarid, M.; Touzani, A. 2016. Automatic 3d extraction of buildings, vegetation and roads from

lidar data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 41: 173–180.

Dai, W.; Jiang, Y.; Zeng, W.; Chen, R.; Xu, Y.; Zhu, N.; et al. 2023. MDC-Net: a multi-directional constrained and prior assisted neural network for wood and leaf separation from terrestrial laser scanning. *International Journal of Digital Earth* 16: 1224–1245.

Eckart, B.; Kim, K.; Jan, K. 2018. EOE: Expected overlap estimation over unstructured point cloud data. *2018 International Conference on 3D Vision (3DV)*: 747–755.

Fei, N.; Gao, Y.; Lu, Z.; Xiang, T. 2021. Z-score normalization, hubness, and few-shot learning. *Proceedings of the IEEE/CVF International Conference on Computer Vision*: 142–151.

Glorot, X.; Bordes, A.; Bengio, Y. 2011. Deep sparse rectifier neural networks. *Proceedings of the fourteenth international conference on artificial intelligence and statistics*: 315–323.

Han, T.; Sánchez-Azofeifa, G.A. 2022. A deep learning time series approach for Leaf and wood classification from Terrestrial LiDAR point clouds. *Remote Sensing* 14: 3157.

Handique, B.K.; Khan, A.Q.; Goswami, C.; Prashnani, M.; Gupta, C.; Raju, P.L.N. 2017. Crop discrimination using multispectral sensor onboard unmanned aerial vehicle. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences* 87: 713–719.

Ioffe, S.; Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International conference on machine learning*: 448–456.

Jayakumari, R.; Nidamanuri, R.R.; Ramiya, A.M. 2021. Object-level classification of vegetable crops in 3D LiDAR point cloud using deep learning convolutional neural networks. *Precision Agriculture* 22: 1617–1633.

Karp, S. 2024. *Adapting to Structure and Using Structure to Adapt: Toward Explaining the Success of Modern Deep Learning*. PhD Thesis, Carnegie Mellon University, Machine Learning Department.

Katz, S.; Tal, A.; Basri, R. 2007. Direct visibility of point sets. In: *ACM SIGGRAPH 2007 Papers*, Vol. 26, p.11.

Lawin, F.J.; Danelljan, M.; Tosteberg, P.; Bhat, G.; Khan, F.S.; Felsberg, M. 2017. Deep projective 3D semantic segmentation. *Computer Analysis of Images and Patterns: 17th International Conference, CAIP 2017, Ystad, Sweden, August 22-24, 2017, Proceedings, Part I*: 95–107.

Liang, J.; Zia, A.; Zhou, J.; Sirault, X. 2013. 3D Plant Modelling via Hyperspectral Imaging. *2013 IEEE International Conference on Computer Vision Workshops*: 172–177.

Murray, J.; Fennell, J.T.; Blackburn, G.A.; Whyatt, J.D.; Li, B. 2020. The novel use of proximal photogrammetry and terrestrial LiDAR to quantify the structural complexity of orchard trees. *Precision Agriculture* 21: 473–483.

Nair, V.; Hinton, G.E. 2010. Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on machine learning (ICML-10)*: 807–814.

Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*: 652–660.

Reji, J.; Nidamanuri, R.R. 2023. Deep learning-based fusion of LiDAR point cloud and multispectral imagery for crop classification sensitive to nitrogen level. *2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)* 1: 1–4.

Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. 1986. Learning representations by back-propagating errors. *Nature* 323: 533–536.

Salehin, I.; Kang, D.-K. 2023. A review on dropout regularization approaches for deep neural networks within the scholarly domain. *Electronics* 12: 3106.

Shen, X.; Huang, Q.; Wang, X.; Li, J.; Xi, B. 2022. A Deep Learning-Based Method for Extracting Standing Wood Feature Parameters from Terrestrial Laser Scanning Point Clouds of Artificially Planted Forest. *Remote Sensing* 14: 3842.

Sun, D.; Robbins, K.; Morales, N.; Shu, Q.; Cen, H. 2022. Advances in optical phenotyping of cereal crops. *Trends in Plant Science* 27: 191–208.

Sun, Y.; Luo, Y.; Chai, X.; Zhang, P.; Zhang, Q.; Xu, L.; et al. 2021. Double-threshold segmentation of panicle and clustering adaptive density estimation for mature rice plants based on 3D point cloud. *Electronics* 10: 872.

Teixeira, I.; Morais, R.; Sousa, J.J.; Cunha, A. 2023. Deep learning models for the classification of crops in aerial imagery: a review. *Agriculture* 13: 965.

Vasquez-Gomez, J.I.; Sucar, L.E.; Murrieta-Cid, R.; Lopez-Damian, E. 2014. Volumetric next-best-view planning for 3D object reconstruction with positioning error. *International Journal of Advanced Robotic Systems* 11: 159.

Wu, B.; Zheng, G.; Chen, Y. 2020. An improved convolution neural network-based model for classifying foliage and woody components from terrestrial laser scanning data. *Remote Sensing* 12: 1010.

Wu, C.; Zeng, R.; Pan, J.; Wang, C.C.L.; Liu, Y.-J. 2019. Plant Phenotyping by Deep-Learning-Based Planner for Multi-Robots. *IEEE Robotics and Automation Letters* 4: 3113–3120.

Zhu, X.; Skidmore, A.K.; Darvishzadeh, R.; Niemann, K.O.; Liu, J.; Shi, Y.; et al. 2018. Foliar and woody materials discriminated using terrestrial LiDAR in a mixed natural forest. *International Journal of Applied Earth Observation and Geoinformation* 64: 43–50.