Vision-based Xylem Wetness Classification in Stem Water Potential Determination

Pamodya Peiris[1][0000−0003−0318−5971], Aritra Samanta[1][0009−0006−3915−0649],
Caio Mucchiani[1][0000−0001−5471−9270], Cody Simons[1][0009−0009−3363−7538], Amit
Roy-Chowdhury[1][0000−0001−6690−9725], and Konstantinos
Karydis[1][0000−0002−1144−8260]

University of California, Riverside, Riverside CA 92521, USA
{ppeir002,asama004,caiocesr,csimo005,amitrc,karydis}@ucr.edu

Abstract. Water is often overused in irrigation, making efficient management of it crucial. Precision Agriculture emphasizes tools like stem water potential (SWP) analysis for better plant status determination. However, such tools often require labor-intensive in-situ sampling. Automation and machine learning can streamline this process and enhance outcomes. This work focused on automating stem detection and xylem wetness classification using the Scholander Pressure Chamber, a widely used but demanding method for SWP measurement. The aim was to refine stem detection and develop computer-vision-based methods to better classify water emergence at the xylem. To this end, we collected and manually annotated video data, applying vision- and learning-based methods for detection and classification. Additionally, we explored data augmentation and fine-tuned parameters to identify the most effective models. The identified best-performing models for stem detection and xylem wetness classification were evaluated end-to-end over 20 SWP measurements. Learning-based stem detection via YOLOv8n combined with ResNet50-based classification achieved a Top-1 accuracy of 80.98%, making it the best-performing approach for xylem wetness classification.

Keywords: Precision Agriculture · Automation · Stem Water Potential · Stem Detection · Wetness Classification

1 Introduction

Different crops exhibit different characteristics based on the provided amount of water [28]. While for certain types of crops, introducing a deficit can have potential advantages, for others it may have detrimental effects. For example, in one study it was shown untimely irrigation scheduling of avocado plants may lead to plant stress and eventually decreased yields [17]. Another study demonstrated that controlled deficit irrigation in oranges led to increased productivity of about 4% [37]. Refining irrigation practices is thus key to sustainable water usage [23]. Several tools are available for assessing plant water status, such as measuring soil salinity [20] and sap flow [21], and proximal sensing of stem water potential [19].

Stem water potential (SWP) is widely adopted as an indicator of water stress in plants [16], reflecting the tension of water within the plant's stem and serving as a sensitive measure to guide irrigation practices. It is considered one of the most accurate plant-based water status measures, especially for fruit trees and vines [13]. SWP is directly related to water availability and plant transpiration, making it a crucial parameter for agricultural applications, including precision irrigation. The pressure chamber is a widely used tool for measuring SWP in plants [3]. To determine SWP using the pressure chamber, a leaf is first enclosed inside a retroflective bag and left to settle for about 10 minutes. The leaf is then cut at its stem from the plant and placed inside the chamber with the cut end protruding. Pressurized gas, usually nitrogen, is gradually introduced until water just begins to appear at the cut surface. This pressure counteracts the suction force within the xylem, which pulls water up from the roots due to transpiration. The applied pressure at which water appears is then used as a proxy of the plant's water potential at the time of sampling. This method is particularly effective for assessing plant water stress, as higher pressure values correspond to greater water tension and suggest a water scarcity in the plant [30].

The pressure chamber method shows significant potential for automation [19], with one core component being the integration of machine vision and learning to classify xylem wetness in an automated manner. A neural network was able to classify the stem as either "dry" or "wet" in manual [5] and automated [18] pressure chambers. Those works considered only two classes, while leaf samples were taken at relatively similar agronomic conditions. However, the efficacy of the network considered in these works may be challenged when leaf samples are taken at varied agronomic conditions (most crucially, at distinct arid/irrigated field cycles [18]). It also does not capture the presence of bubbles that often occur in the SWP process [33], which, as we show in this work, can significantly challenge visual learning methods. Together, these challenges raise the key question we aim to address in this work: *How to enable generalizable machine vision and learning tools for automated stem detection and xylem wetness classification?*

Conceptually-related works have studied leaf and disease classification [1], xylem tissue segmentation [8,9], and xylem cell type classification [35] using convolutional neural networks (CNNs). Classical methods have been used for plant stress detection (e.g., segmentation and gradient-based decision trees [39] or SWP measurements via leaf length [36]). Given the limited research on detection and classification methods for water observations specifically at the stem cross-section, this study conducts a comprehensive comparative analysis of vision- and learning-based methodologies to identify the distinct observable stages of the stem under increasing pressure inside the chamber. In addition to the "wet" and "dry" classes considered before [5,18], we also consider a third class "bubble."

Our framework (Fig. 1) is structured into three phases: data collection, stem detection, and image/video classification. We compare stem detection efficacy of classical machine vision (Hough Transforms) and visual learning tools ('You Only Look Once' – YOLO [26]) in terms of the intersection-over-union (IoU) against manually annotated images, as well as execution time. We then evaluate state-
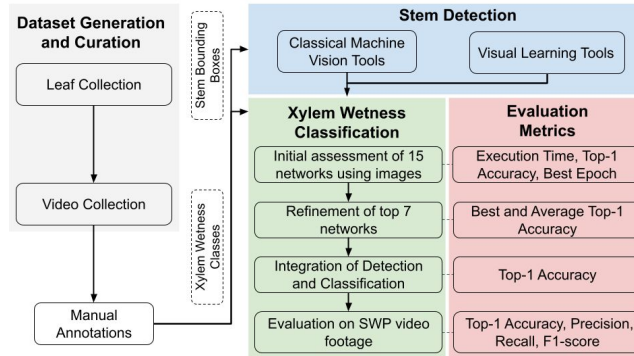
Fig. 1: Our end-to-end framework for vision-based xylem wetness classification has three distinct phases. Data collection, and dataset generation and curation help train and assess methods for stem detection. The best-performing stem detection methods are then used together with the dataset to train and evaluate several plausible learning-based xylem wetness classification methods.

of-the-art CNN models and identify the best methods based on accuracy and classification rates for xylem wetness. Best-performing detection and classification methods are downselected and tested on SWP measurement video footage. Overall, we evaluate fifteen CNN architectures and five detection methods to find that learning-based stem cropping paired with ResNet-50 provides the best Top-1 accuracy in classifying the xylem wetness. To our knowledge, this study is the first to comparatively evaluate various detection and classification methods for stem xylem wetness in SWP. This step is essential for enhancing precision and accuracy and advancing automated solutions for the pressure chamber method.

2   Materials and Methods

This section outlines all components of this study. It describes the experimental setup and data collection procedures, dataset generation and curation, and the methods considered for stem detection and xylem wetness classification.

2.1   Experimental Setup

We used the automated pressure chamber system developed by Mucchiani and Karydis [18]. In summary, a commercially available pressure chamber (PMS 600D) is retrofitted with a single-board computer (Raspberry Pi Model 3B), a microcontroller (Arduino Uno R3) a camera (Raspberry Pi HQ camera V1), a joystick, an air compressor, a set of relays, a pressure sensor, and a solenoid valve (Fig. 2a). The system is rated to withstand a maximum pressure of 20 bar, deemed sufficient for most avocado [2] and citrus [22] leaves, among others [4]. Once a leaf is loaded (manually) by an operator, the process starts and the pressure inside the chamber increases, while simultaneously recording video data. The operator can pause, end, and reset the process at any point.
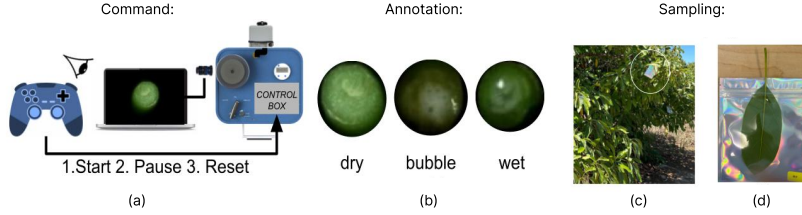
Fig. 2: (a) The data acquisition setup is based on an automated pressure chamber [18], and allows an operator full control of the process. (b) Examples of the three classes considered in this work. (c) Leaves have been bagged for 10 minutes before the excision from the tree. (d) The bagged leaf.

2.2 Data Collections

We sampled leaves from local avocado trees located at UC Riverside's Agricultural Experimental Station (AES; 33°58′04.5″N, 117°20′04.8″W) following the procedure in [18]. Healthy and mature leaves that receive full sun for part of the day were identified (Fig. 2c), placed in a retroreflective bag, left to settle for 10 minutes (Fig. 2d), and then excised cleanly at their stem. To promote sample variability and better test the methods' generalization capability, we conducted four field data collections and sampled leaves from multiple trees (with a maximum of five leaf samples from each tree in all but three field experiments) over different days and with different ambient temperature and soil moisture level (as deduced via the time elapsed post-irrigation). A total of 117 leaves were collected; Table 1 provides specific details.

After sampling leaves, we performed SWP measurements. We manually focused the camera on the stem to get clear images (Fig. 2b). The aperture was adjusted to focus the lighting on the stem while minimizing reflections on other visible parts. The camera focus was occasionally re-adjusted to refine the view during experiments as needed. We ensured the pressure chamber was securely sealed before each experiment. Then, the pressure inside the chamber was increased until a clear xylem water expression. Once a clear "wet" xylem status was observed, the pressure was maintained constant for at least 2-3 seconds before releasing it. For safety purposes, the air compressor was stopped once the pressure reached 18 bar irrespective of the xylem wetness status. After complete pressure release, the recording was ended and the leaf was removed from the chamber. The camera was stationary throughout the entire process. Three researchers familiar with the process and hardware conducted these experiments.[1]

---

[1] In one instance we performed SWP measurements on two separate sessions. To keep the air compressor's temperature within safe limits, a maximum of 30 leaves can be tested at each SWP measurement session. Hence, the rest 20 leaves were stored in a cool and dry place overnight, and measurements resumed 23 hours post-excision. Images from the corresponding videos were used in stem detection; however, they were excluded from the xylem wetness classification as they were considerably drier.

Table 1: Information on Leaf Data Collection and SWP Video Generation.

Date	Time	Temp (F)	# Trees	# Leaves	Time elapsed post irrigation (h)
05/01	11:00 am	63	1	8	unknown
05/05	12:00 pm	61	10	50	55 (30 leaves) & 78 (20 leaves)
05/12	12:30 pm	72	6	29	55.5
07/03	11:00 am	91	6	30	6

2.3 Dataset Generation and Curation

Videos from SWP measurements were processed to create three distinct datasets: the Stem Detection dataset, the Xylem Wetness Classification dataset, and the Video Evaluation dataset. The first two datasets comprise images extracted from SWP video footage cropped between several frames (but not overly many) before stem water expression and until the SWP cycle ends. Images were extracted at 30 fps with a resolution of 640×480 pixels and were stored in *jpg* format. The *Stem Detection dataset* comprises images from 30 videos of mostly dry measurements, collected on May 5th and July 3rd. The videos were split into training (18), validation (6), and testing (6) sets. Training and validation images were 4,500 and 1,500, respectively. For testing, we used the first 200 frames of the testing set videos. The *Xylem Wetness Classification dataset* comprises images sampled from 40 videos collected on May 1st, 5th, and 12th. The videos were split into training (24), validation (8), and testing (8) sets. Frames were sampled so that a balanced distribution between all the classes was maintained, thus resulting in 13,500/4,500/4,500 images, respectively. Images in this dataset were cropped around the stem using bounding boxes derived via manual annotation and automated stem detection developed based on the Stem Detection dataset.

To determine the ground truth for classification, we initially annotated the xylem wetness into "dry," "wet," and "bubble" of the first 30 videos collected on May 5th. Every video was annotated independently by three different annotators. After some initial annotations and observed disagreements, the team discussed in depth to define "wet" as being more transparent, while "bubble" appears to be whiter, due to the scattering of light in different directions. [2] The refined definition was used to guide the remaining annotations.

The *Video Evaluation dataset* comprises 20 videos collected on July 3rd that were used for end-to-end evaluation of the overall pipeline (i.e. joint stem detection and xylem wetness classification). Videos were annotated with both bounding boxes and wetness classification as in the other two datasets.

2.4 Stem Detection

We explored both classical vision-based methods and learning-based methods for stem detection. The distinct strengths and weaknesses of methods belonging to these two categories have been studied across different application contexts [24]. However, evaluating the efficacy of these tools within the context of this work

---

[2] We refer the reader to `https://bit.ly/arcs_isvc24` for supplementary information.

(stem detection in the SWP process) is new, and it is thus important to understand how tools in both categories compare against each other. Different methods were evaluated in terms of (1) the intersection over union (IoU) with respect to manually annotated ground truth bounding boxes, and (2) their execution time to successfully detect the bounding box around the stem.

Hough Transform [7] is a classical method used for line and curve detection. It has been used in agronomy to detect the diameter of tree trunks [14]; that result inspired us to investigate it for stem detection. The OpenCV [3] built-in Hough Circle Transform function was used for detecting Hough circles across all frames. Any frames with multiple detections were discarded considering that exactly one stem is present in all cases. For the remaining frames, we rounded the center and radius of each circle to the nearest pixel and created a histogram of the position and radius. We then took the mode of all detections and evaluated the cases with the radius padded by 20 (H20), 30 (H30), and 40 (H40) pixels.

YOLOv5nu [4] and YOLOv8n [5] are the two learning-based methods employed in this work. Their selection was based on their widespread use across applications. To create the YOLO-derived bounding boxes we used the coordinates corresponding to the top-left and bottom-right corners.

2.5   Xylem Wetness Classification

Given that different networks with good performance have been developed for classification tasks, we elected to follow a tiered approach. Specifically, from a large set of networks, we downselected the best-performing ones which were ultimately evaluated using video footage of SWP measurements.

In the first phase, we considered fifteen CNN model architectures known to be computationally efficient and used in edge computing. These include Efficient-Net B0/B1 [32], Inception V3 [31], MobileNetv2 [27], MobileNetv3-small [11], ResNet18/34/50 [10], Shufflenet v2 1.0/2.0 [15], SqueezeNet 1.0/1.1 [12], VGG16/19 [29], and YOLOv8n-cls. All models were implemented using PyTorch [25] and were pretrained on the ImageNet [6] dataset. We trained all models except YOLOv8n-cls for 20 epochs with a batch size of 16, a learning rate of $1e - 3$, and via Stochastic Gradient Descent (SGD) optimizer. The images were resized as needed to meet each model's input image size requirement. YOLOv8n-cls was trained for 30 epochs with a 64×64 image size, 16 batch size, and 0.01 initial and final learning rate with SGD optimizer. We compared the Top-1 testing accuracy of the best and last trained model, the time to train an epoch, and whether the validation accuracy increased during the training process. The criteria to downselect networks to consider in the next phase included: (1) Top-1 accuracy greater than 70%, (2) training time less than 100 s, and (3) the epoch with the best validation accuracy being greater than one. In the case that multiple network variants belonging to the same family (e.g., the three ResNet variants) passed the bar, we kept only the best-performing one.

---

[3] https://opencv.org/
[4] https://doi.org/10.5281/zenodo.3908559
[5] https://github.com/ultralytics/ultralytics

Table 2: Hyperparameters Used for Tuning Downselected Models in Phase 2.

	Optim	LR	LRS Milestones	Batch Size	Epoch	Aug	LR_{final}
CNN	SGD Adam	1e-3	None [5, 10, 20-th epoch] [1, 3, 5, 8-th epoch]	16 32 64	15 20 30	Random Crop	-
YOLOv8n-cls	SGD Adam	1e-2	-	-	10 30 50	Built-in	1e-2 1e-6

Table 3: Hyperparameters Used in the Models Downselected for Phase 3.

Case/Criteria	B	E	LR_{init}	LR_{final}	LSR Milestones	Optim	Image Size
F1	16	30	1e-3	1e-3	None	SGD	224×224
F2	64	20	1e-3	1e-7	[1, 3, 5, 8-th epoch]	SGD	224×224
F3	16	50	1e-2	1e-6	-	SGD	64×64
F4	16	50	1e-2	1e-6	-	SGD	224×224

Seven networks moved on to the second phase. We then performed a hyperparameter search for those models (Table 2). We employed most of the built-in parameters for the downselected YOLOv8-cls network, tuned the optimizer (Optim), and varied the number of epochs (E) and the final learning rate (LR_{final}). These led to five different variants for YOLOv8-cls to be evaluated. For the remaining CNN models, we tuned the optimizer, learning rate (LR), learning rate scheduler (LRS) milestones for different number of epochs, and batch size (B). We also augmented (Aug) the dataset using random cropping and froze top layers during training to further improve model performance. These led to 18 different variants for each of the six other CNN networks to be evaluated. The models' best and average Top-1 testing accuracies were the evaluation metrics.

To further refine the evaluation, we chose the top two models in Top-1 accuracy and the top two in average Top-1 accuracy to move to the third phase. These four models were subsequently trained using two hyperparameter setups outlined in cases F1–F4 in Table 3. The best-performing configuration in terms of Top-1 accuracy for each of these four models was then moved on to the overall end-to-end framework evaluation with SWP video footage. Note that during the first two phases, we used cropped images based on the manually annotated stem bounding boxes. In the third phase, we used these, and in addition, the cropped images provided by the vision- and learning-based stem detection methods.

Lastly, the final four downselected network configurations were combined with the downselected stem detection methods and were evaluated using the Video Evaluation dataset. During this evaluation, the stem detection method was run first on frames 501–700. Once the stem has been detected, the classification network determines xylem wetness from frame 701 onwards until all the frames are exhausted. The end-to-end framework's classification was evaluated using the Top-1 accuracy, Precision, Recall, and F1-score.

3   Results and Discussion

3.1   Evaluation of Stem Detection

The Hough Transform with 20-pixel padding (H20) performs slightly worse but is generally comparable to YOLO network variants, with a maximum performance drop of about 5% in terms of IoU (Table 4). The execution time of the different Hough Transform paddings is nearly identical and approximately six times faster than that of YOLO, with both YOLO variants having similar execution times. The YOLO networks maintain high and consistent performance, with about a 3% drop in IoU. The main difference is that YOLOv8n can provide tighter bounding boxes (see Fig. 3). Therefore, YOLOv8n was chosen as the learning-based method to be tested in the overall end-to-end framework.

The Hough Transform can be quite fragile. As the padding increases the performance drops significantly (about 28% max IoU drop between H20 and H40). This is likely because of the increased background noise added to the image part used in IoU computation, resulting in lower matching accuracy. Additionally, we tested the Hough Transform in the first 200 frames of the entire Stem Detection dataset (third row in Table 4). In this case, there is higher variability in terms of the location of the stem in the provided images. This caused a significant drop in the H20 case (exceeding 10%), a small increase in H40, and a not significant change in the H30 case. These findings suggest that H30 is more stable, which is important for generalization, and was thus selected as the vision-based method to be tested in the overall end-to-end framework.

Table 4: Comparison of IoU for Different Stem Detection Methods.

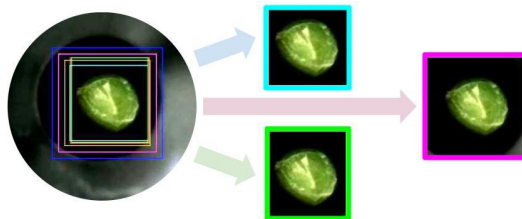|	H20	**H30**	H40	YOLOv5nu	**YOLOv8n**
Test IoU	80.56	71.62	57.64	84.86	82.26
time/video (s)	1.327	1.334	1.351	8.167	7.976
Full Stem Dataset IoU	71.96	71.68	61.67	-	-

Fig. 3: Sample stem detection bounding boxes via manual annotation (cyan), H20 (yellow), H30 (magenta), H40 (blue), YOLOv5nu (red), and YOLOv8n (green). The dimensions and placement of the bounding boxes differ between the tested methods. The two selected methods (H30 and YOLO8vn) along with the manually-annotated bounding box are highlighted.

3.2   Evaluation of Xylem Wetness Classification

Results from the evaluation of all 15 models considered initially are shown in Table 5. T_E represents the time took to train the model for one epoch. $P1_{BA}$ reports the Top-1 accuracy between the best and the last trained model. B_E is the epoch where the model demonstrated the best validation accuracy. A value of 1 indicates that the validation accuracy did not improve during training.

Out of the 15 models, most exceeded the set threshold of 70% Top-1 accuracy. Those models that failed to pass the bar (SqueezeNet 1.1 and VGG16/19) were able to classify only one class correctly (most often the "bubble" class). Several models with good Top-1 accuracy, however, did not show improvement past the first epoch of training and were thus discarded. Two exceptions were made to the latter case; Inception V3 and SqueezeNet 1.0. This is because both models had very high accuracy (over 90% in classifying "dry" classes and marginal accuracy (over 65%) in classifying the other two classes; hence, we hypothesized that fine-tuning can further improve the performance of these models. Downselected models for the second phase testing are indicated with a checkmark in Table 5.

Table 5: First Two Phases of Xylem Wetness Classification Network Evaluation.

CNN Model	T_E (s) ↓	$P1_{BA}$ % ↑	B_E	Phase 2	$P2_{BA}$ % ↑	$P2_{AVG}$ %	Phase 3
EfficientNet B0	63.76	75.53	10	✓	83.11	**76.65**	✓
EfficientNet B1	86.15	73.67	1	✗	-	-	✗
Inception V3	133.25	77.86	1	✓	82.33	73.54	✗
MobileNetv2	54.48	70.51	19	✗	-	-	✗
MobileNetv3-small	45.47	89.07	2	✓	**85.44**	75.91	✓
ResNet18	54.57	72.86	2	✗	-	-	✗
ResNet34	64.30	74.78	1	✗	-	-	✗
ResNet50	90.87	73.15	20	✓	**84.86**	76.39	✓
ShuffleNet v2 1.0	49.60	81.93	8	✓	84.09	75.89	✗
ShuffleNet v2 2.0	52.28	74.73	1	✗	-	-	✗
SqueezeNet 1.0	46.49	80.67	1	✓	83.16	70.47	✗
SqueezeNet 1.1	40.66	33.33	1	✗	-	-	✗
VGG16	136.07	33.33	1	✗	-	-	✗
VGG19	150.05	33.33	1	✗	-	-	✗
YOLOv8n-cls	43.12	82.42	-	✓	83.15	**78.79**	✓

The different parameterizations (Table 2) were then implemented for all downselected models. The overall best $P2_{BA}$ and average $P2_{AVG}$ Top-1 accuracy from these evaluations are listed in Table 5. We noted that applying data augmentations did not improve training. Our datasets comprise images and videos that are tightly centered around the stem, and all look similar. As such, although there is some background, resizing and randomly cropping may

Table 6: Top-1 Accuracy of Models Integrated with Stem Detection Methods.

CNN Model	Manual crop	H30 crop	YOLO crop
EfficientNet B0 (F1)	81.29	76.2	77.44
EfficientNet B0 (F2)	**81.64**	77.33	80.33
MobileNetv3-small (F1)	82.11	80.64	76.27
MobileNetv3-small (F2)	81.11	81.98	**82.22**
ResNet50 (F1)	82.24	82.69	77.31
ResNet50 (F2)	76.18	**84.24**	70.13
YOLOv8n-cls (F3)	77.38	77.76	76.48
YOLOv8n-cls (F4)	**82.93**	81.42	82.4

lead to images that miss critical information. Further, the SGD optimizer led to better performance; this finding is consistent with other similar observations in the literature [38]. It was also observed that a faster decrease in the learning rate led to models with improved generalization accuracy. Overall, the difference in best performance between the seven fine-tuned networks is subtle (less than 4% max difference in $P2_{BA}$). Differences in average performance are more pronounced (about 12% max difference in $P2_{AVG}$).

The top two networks in each metric category were then selected for further evaluation using two parametric setups out of those listed in Table 3. Downselected detectors were also integrated. The specific selections, as well as results from this third phase of evaluation, are reported in Table 6. The best-performing combinations for each method are highlighted. It can be observed that networks trained using the H30 stem detector have higher accuracy with learning rate scheduling (F2). This is because H30 crops the image around the stem more tightly than manual and YOLO-based crops which, in turn, leads to less background noise in the image. As our approach follows a more generalized learning method to create a network that performs better on unseen data, smaller learning rates (F2) contribute to achieving a higher accuracy [34]. Overall, EfficientNet B0 and MobileNetv3-small appear to be more stable when hyperparameters vary, while YOLOv8n-cls is the network listed affected by the employed stem detection means. ResNet50 has the widest variation in Top-1 accuracy among the different parameterizations considered herein.

3.3   Video Evaluation

The best-performing hyperparameter cases for each of the four networks in Table 6 were combined with either the H30 or the YOLOv8n stem detector and were then evaluated in SWP video footage. Table 7 contains the results of this end-to-end xylem wetness classification pipeline over 20 different videos. It can be readily observed that YOLOv8n consistently performs better as a stem detector in comparison to H30. The case of the YOLOv8n detector integrated with the ResNet-50 classifier (C7) demonstrated the best performance (about 22%

Table 7: End-to-end SWP Video Footage Evaluation.

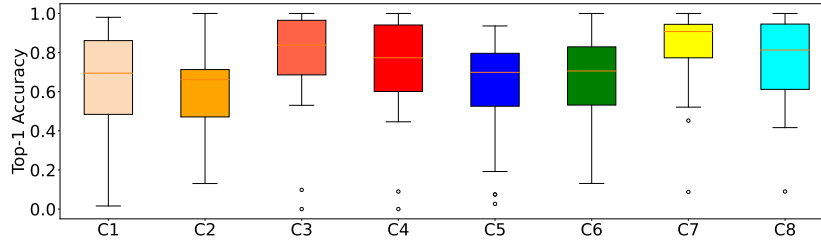| Case	Stem Detection and Classification	Top-1 Acc.	Precision	Recall	F1-score
C1	H30 & EfficientNet B0 (F2)	60.13	87.66	60.13	64.07
C2	H30 & MobileNetv3-small (F2)	60.70	86.27	60.70	60.37
C3	H30 & ResNet50 (F2)	74.93	88.89	74.93	77.15
C4	H30 & YOLOv8n-cls (F4)	71.33	84.89	71.33	72.13
C5	YOLOv8n & EfficientNet B0 (F2)	60.35	**92.41**	60.35	65.20
C6	YOLOv8n & MobileNetv3-small (F2)	67.33	89.68	67.33	67.30
C7	YOLOv8n & ResNet50 (F2)	**80.98**	90.17	**80.98**	**82.58**
C8	YOLOv8n & YOLOv8n-cls (F4)	74.98	89.94	74.98	76.06

Fig. 4: Top-1 Accuracy for end-to-end xylem wetness classification.

on average higher Top-1 Accuracy in comparison to the other cases). This case synergistically leverages the strong generalization capabilities of both YOLO and ResNet-50. Even though ResNet-50 is a CNN framework with considerable depth, it effectively retains information owing to its residual learning capacity. Finally, Fig. 4 presents the distribution of the Top-1 Accuracy of all eight cases listed in Table 7. It can be observed that case C7 has the least amount of variability and the highest average value. Case C3 (which also uses ResNet-50 but the H30 detector) is also performing well overall and could be considered a variable alternative, but its best output is less strong compared to when using the YOLO-based detector. The outliers in the boxplots correspond to no detection in some cases. Further investigation into these cases revealed that the video brightness was too low and hindered detection. In all, these findings suggest that a combination of a learning-based stem detection method and a deep CNN wetness classifier, when trained meticulously, can provide accurate predictions, even for previously unseen data.

4   Conclusion

We introduced a framework for automated vision-based classification of xylem wetness during stem water potential (SWP) experiments. Our approach combined two parts; stem detection and xylem wetness classification. We conducted an extensive tiered evaluation of a total of two different methods for stem detec-

tion and 15 different methods for xylem wetness classification as well as different parameterizations thereof via hyperparameter tuning. Evaluation considered both images and video footage provided by SWP measurements. Our findings revealed several viable combinations of detectors and classifiers and established key similarities and differences. In sum, the combination of a YOLO-based detector and a ResNet-50 classifier was identified as the best-performing configuration, owing to the synergies afforded by those models.

To the best of our knowledge, this is the first comprehensive study to perform joint stem detection and xylem wetness classification in support of SWP measurement via the pressure chamber, a process that is critical for precision irrigation. This work can support further automation of the pressure chamber method to measure SWP. It can also yield crucial information for similar precision agriculture efforts that seek to leverage the recent advances in visual learning. Future work will seek to include different plant species (e.g., citrus) as well as investigate the use of other sensing modalities (e.g., near-infrared imaging).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Al Bashish, D., Braik, M., Bani-Ahmad, S.: A framework for detection and classification of plant leaf and stem diseases. In: International Conference on Signal and Image Processing. pp. 113–118. IEEE (2010)
2. Bonomelli, C., Arias, M.I., Celis, V., Venegas, D., Gil, P.M.: Effect of application of sulfuric and humic acid in the mitigation of root asphyxiation stress in potted avocado plants. Communications in Soil Science and Plant Analysis **50**(13), 1591–1603 (2019)
3. Boyer, J.: Leaf water potentials measured with a pressure chamber. Plant Physiology **42**(1), 133–137 (1967)
4. Chone, X., Van Leeuwen, C., Dubourdieu, D., Gaudillère, J.P.: Stem water potential is a sensitive indicator of grapevine water status. Annals of Botany **87**(4), 477–483 (2001)
5. Dechemi, A., Chatziparaschis, D., Chen, J., Campbell, M., Shamshirgaran, A., Mucchiani, C., Roy-Chowdhury, A., Carpin, S., Karydis, K.: Robotic assessment of a crop's need for watering: Automating a time-consuming task to support sustainable agriculture. IEEE Robotics and Automation Magazine **30**(4), 52–67 (2023)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009)
7. Duda, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. Communications of the ACM **15**(1), 11–15 (1972)

8. Garcia-Pedrero, A., García-Cervigón, A., Caetano, C., Calderón-Ramírez, S., Olano, J.M., Gonzalo-Martín, C., Lillo-Saavedra, M., García-Hidalgo, M.: Xylem vessels segmentation through a deep learning approach: a first look. In: IEEE international work conference on bioinspired intelligence (IWOBI). pp. 1–9 (2018)

9. Garcia-Pedrero, A., Garcia-Cervigon, A.I., Olano, J.M., Garcia-Hidalgo, M., Lillo-Saavedra, M., Gonzalo-Martin, C., Caetano, C., Calderon-Ramirez, S.: Convolutional neural networks for segmenting xylem vessels in stained cross-sectional images. Neural Computing and Applications **32**, 17927–17939 (2020)

10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)

11. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for mobilenetv3. In: IEEE/CVF International Conference on Computer Vision. pp. 1314–1324 (2019)

12. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. arXiv preprint arXiv:1602.07360 (2016)

13. Levin, A.D.: Re-evaluating pressure chamber methods of water status determination in field-grown grapevine (vitis spp.). Agricultural water management **221**, 422–429 (2019)

14. Li, D., Jia, W., Guo, H., Wang, F., Ma, Y., Peng, W., Zhang, S.: Use of terrestrial laser scanning to obtain the stem diameters of larix olgensis and construct compatible taper-volume equations. Trees **37**(3), 749–760 (2023)

15. Ma, N., Zhang, X., Zheng, H.T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: European Conference on Computer Vision (ECCV). pp. 116–131. Springer (2018)

16. McCutchan, H., Shackel, K.: Stem-water potential as a sensitive indicator of water stress in prune trees (prunus domestica l. cv. french). Journal of the American Society for Horticultural Science **117**(4), 607–611 (1992)

17. Michelakis, N., Vougioucalou, E., Clapaki, G.: Water use, wetted soil volume, root distribution and yield of avocado under drip irrigation. Agricultural Water Management **24**(2), 119–131 (1993)

18. Mucchiani, C., Karydis, K.: Development of an automated and artificial intelligence assisted pressure chamber for stem water potential determination. Computers and Electronics in Agriculture **222**, 109016 (2024)

19. Mucchiani, C., Zaccaria, D., Karydis, K.: Assessing the potential of integrating automation and artificial intelligence across sample-destructive methods to determine plant water status: A review and score-based evaluation. Computers and Electronics in Agriculture **224**, 108992 (2024)

20. Muhammad, M., Waheed, A., Wahab, A., Majeed, M., Nazim, M., Liu, Y.H., Li, L., Li, W.J.: Soil salinity and drought tolerance: An evaluation of plant growth, productivity, microbial diversity, and amelioration strategies. Plant Stress p. 100319 (2023)

21. Nadezhdina, N.: Sap flow index as an indicator of plant water status. Tree Physiology **19**(13), 885–891 (1999)

22. Ortuño, M., García-Orellana, Y., Conejero, W., Ruiz-Sánchez, M., Mounzer, O., Alarcón, J., Torrecillas, A.: Relationships between climatic variables and sap flow, stem water potential and maximum daily trunk shrinkage in lemon trees. Plant and Soil **279**, 229–242 (2006)

23. Oster, J.D., Wichelns, D.: Economic and agronomic strategies to achieve sustainable irrigation. Irrigation Science **22**, 107–120 (2003)

24. O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G.V., Krpalkova, L., Riordan, D., Walsh, J.: Deep learning vs. traditional computer vision. In: Advances in Computer Vision: Proceedings of the Computer Vision Conference (CVC), Volume 1 1. pp. 128–144. Springer (2019)
25. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems **32** (2019)
26. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 779–788 (2016)
27. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4510–4520 (2018)
28. Shackel, K., Moriana, A., Marino, G., Corell, M., Pérez-López, D., Martin-Palomo, M.J., Caruso, T., Marra, F.P., Agüero Alcaras, L.M., Milliron, L., et al.: Establishing a reference baseline for midday stem water potential in olive and its use for plant-based irrigation management. Frontiers in Plant Science **12**, 791711 (2021)
29. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
30. Suter, B., Triolo, R., Pernet, D., Dai, Z., Van Leeuwen, C.: Modeling stem water potential by separating the effects of soil water availability and climatic conditions on water status in grapevine (vitis vinifera l.). Frontiers in Plant Science **10**, 1485 (2019)
31. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2818–2826 (2016)
32. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114. PMLR (2019)
33. Wei, C., Tyree, M., Bennink, J.: The transmission of gas pressure to xylem fluid pressure when plants are inside a pressure bomb. Journal of Experimental Botany **51**(343), 309–316 (2000)
34. Wilson, D.R., Martinez, T.R.: The need for small learning rates on large problems. In: IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222). vol. 1, pp. 115–119. IEEE (2001)
35. Wu, S., Dabagh, R.A., Jacobsen, A.L., Holmlund, H.I., Scalzo, F.: Deep learning-based classification of plant xylem tissue from light micrographs. In: International Symposium on Visual Computing. pp. 237–248. Springer (2022)
36. Yamane, T., Habaragamuwa, H., Sugiura, R., Takahashi, T., Hayama, H., Mitani, N.: Stem water potential estimation from images using a field noise-robust deep regression-based approach in peach trees. Scientific Reports **13**(1), 22359 (2023)
37. Zapata-Sierra, A.J., Manzano-Agugliaro, F.: Controlled deficit irrigation for orange trees in mediterranean countries. Journal of Cleaner Production **162**, 130–140 (2017)
38. Zhou, P., Feng, J., Ma, C., Xiong, C., Hoi, S.C.H., et al.: Towards theoretically understanding why sgd generalizes better than adam in deep learning. Advances in Neural Information Processing Systems **33**, 21285–21296 (2020)
39. Zhuang, S., Wang, P., Jiang, B., Li, M., Gong, Z.: Early detection of water stress in maize based on digital images. Computers and Electronics in Agriculture **140**, 461–468 (2017)