



OPEN A multi-spectral and hyperspectral image dataset for evaluating chemical traits and the water status of avocado, olive and grape through leaf dehydration under laboratory conditions

Juan Sebastian Estrada^{1,5}, Rodrigo Demarco^{2,5}, Ciarán Miceal Johnson^{3,5}, Matias Zañartu^{1,5}, Andres Fuentes^{2,5} & Fernando Auat Cheein^{1,4,5}

Assessing the health status of vegetation is of vital importance for all stakeholders. Multi-spectral and hyper-spectral imaging systems are tools for evaluating the health of vegetation in laboratory settings, and also hold the potential of assessing vegetation of large portions of land. However, the literature lacks benchmark datasets to test algorithms for predicting plant health status, with most researchers creating tailored datasets. This work presents a dataset composed of multi-spectral images, hyper-spectral reflectance values, and measurements of weight, chlorophyll, and nitrogen content of leaves at five different drying stages, from avocado, olive, and grape trees, which are common crops in the Valparaíso region of Chile. This dataset is a valuable asset for developing tools in the field of precision agriculture and assessing the general health status of vegetation.

Agriculture in Chile is currently being affected by various factors such as climate change, the reduction in annual total rainfall, increasing maximum and minimum temperatures in interior regions, the high frequency of high temperatures during summer, polar frosts during winter, and changes in seasonal rainfall patterns. These symptoms contribute to a more challenging climate for agriculture, leading to increased instability and higher stress¹. Additionally, water sources have shown signs of exhaustion. The farming industry, alongside others such as mining and power generation, has excessive water demands, resulting in a deficit of this resource². Improving the management of orchards and crops will help address the water deficit and maintain the quality of crops³.

The use of robotic platforms combined with multi-spectral and hyperspectral imaging systems has become a common technique for quickly assessing the health status of vegetation across relatively large areas of land^{3,4}. Spectral information from vegetation is used to compute vegetation indices, which are sensitive to both physical and chemical properties of the vegetation. By integrating vegetation indices derived from spectral information and in-situ measurements, it is possible to train machine-learning models to predict important agronomic indicators, such as chlorophyll and nitrogen content, biomass, and yield, across a range of cultivars, including maize^{5,6}, rice^{7,8}, ratoon rice⁹, cotton^{10,11}, onion¹², potato¹³, wheat¹⁴, sugarcane¹⁵, soybean¹⁶, and sweet potato¹⁷.

There are numerous datasets in the literature containing spectral information about leaves. One notable example is the ANGERS Leaf Optical Properties Database (2003)¹⁸, which includes reflectance and transmittance measurements for 276 leaf samples from 39 different tree species in the Angers region of France. The reflectance values range from 400 to 2400 nm. Data collection occurred in June 2003, and the dataset also includes information on pigment content and physical parameters such as leaf mass and area.

¹Department of Electronic Engineering, Universidad Tecnica Federico Santa Maria, Valparaíso, Chile.

²Department of Industrial Engineering, Universidad Tecnica Federico Santa Maria, Valparaíso, Chile. ³UK National Robotarium, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, Scotland, UK.

⁴Harper Adams University, Newport, England, UK. ⁵Juan Sebastian Estrada, Rodrigo Demarco, Ciarán Miceal Johnson, Matias Zañartu, Andres Fuentes, Fernando Auat Cheein have contributed equally to this work. ✉email: juan.estrada@usm.cl

A similar dataset is the Leaf Optical Properties Experiment Database (LOPEX93)¹⁹, which includes reflectance and transmittance values for 330 leaf samples from 45 different species in the Ispra region, Italy. The experiments were conducted in 1993. The dataset also provides information on the biochemical and physical parameters of the leaves, similar to the ANGERS dataset. Both datasets are highly relevant as they were used to calibrate the PROSPECT-4 and PROSPECT-5 models²⁰, which are employed to predict the contribution of biochemical parameters (chlorophyll a, chlorophyll b, carotenoids, water, and dry matter) to the optical properties of leaves.

Both the LOPEX93 and ANGERS datasetd contain information from trees in the Mediterranean region, in contrast, the dataset presented in^{21,22}, contains leaf spectral data from tropical trees in the Tapajos national forest, Brazil. Besides the spectral information, the authors also provide leaf area, leaf mass, leaf relative water content, and relative canopy environment for each leaf.

The dataset presented in²³ shows the leaf spectra and biochemical traits of agricultural species. The crops included are common sunflower, cotton, pumpkin, cucumber, tomato, kidney bean, soybean, and sweet basil. The dataset comprises 184 reflectance values along with the corresponding biochemical parameters. A similar dataset is detailed in²⁴, which includes information on watered and drought-stressed crops. This dataset features crop species such as Carolina poplar, cayenne pepper, common sunflower, radish, field pumpkin, foxtail millet, and sorghum.

As seen, there are numerous datasets with spectral information from leaves of various species. However, these datasets typically provide only a single spectrum per leaf, which corresponds to a unique area within the leaf. In contrast, a multi-spectral camera can record multiple pixels from a single leaf, making it preferable for covering larger areas of land⁴. In this context, a few datasets in the literature address this need. For example, Ryckewaert et al.²⁵ proposed a dataset of hyperspectral images of grapevine leaves, both healthy and exhibiting disease symptoms. This dataset includes 204 hyperspectral images of grapevine fields, covering the range of 400–900 nm. In addition to the hyperspectral images, the dataset provides information on chlorophyll content, flavonoid content, and nitrogen crop status. Another relevant dataset, presented by Vayssade²⁶, consists of multi-spectral images in 6 bands and is used for the identification of crops and weeds.

More recent datasets concerned with chemical and biological traits of vegetation are the one presented by Przybylska et al.²⁷, where hyperspectral reflectance and functional traits of *Arabidopsis thaliana* are shown. This dataset is still only concerned with hyperspectral reflectance and does not provide images. Other dataset with similar features is the one presented in²⁸, where reflectance and measurements from tundra vegetation are shared to the scientific community.

Other datasets such as the one presented in²⁹, show multi-spectral images of tomato leaves for the assessment of pests and diseases. One disadvantage of this dataset is that it does not include information about chemical or biological traits of vegetation, and it does not include hyperspectral reflectance.

Most datasets in the literature focus on a single type of information, whether it be full spectral data or images. Each approach has its limitations: images may lack spectral resolution, while hyperspectral data collected with a spectrometer may lack spatial resolution. Combining these types of information could leverage the advantages of both approaches-offering the cost-effectiveness, portability, and ability to survey large areas provided by multi-spectral cameras, alongside the rich, detailed information obtained from hyperspectral imaging systems.

Thus, we propose a dataset composed of spectral reflectance ranging from 350 to 2500 nm, multi-spectral images with five bands (blue, green, red, red-edge, and near-infrared), as well as data on chlorophyll content (in SPAD values), leaf nitrogen content, and leaf water content. The samples include leaves from common cultivars in the Valparaíso region of Chile, specifically avocado, olive, and grape. To analyze the water content of the leaves, the samples were subjected to a drying process, resulting in five dehydration stages. The final dataset is composed of 106 leaves of avocado, 111 leaves of olive, and 106 leaves of grape.

Methods

This section describes the methodology used to collect the samples for the database, including the equipment utilized, the dehydration process of the leaves, and the post-processing applied to both the images and numerical data.

Equipment

The equipment consists of:

- *Micasense Red-Edge Camera* Captures multi-spectral images in five bands: blue, green, red, red-edge, and near-infrared.
- *ASD TerraSpec Hi-Res Spectrometer* Measures the spectral reflectance of leaves in the range of 350–2500 nm.
- *TYS-4N Plant Nutrition Meter* Measures chlorophyll and nitrogen content in each leaf.
- *Kern PFB 120-3 Balance* Used to weigh the leaves.
- *Memmert UN30 Universal Oven* Used to dry the leaves, producing five different drying stages. Table 1 provides the main features of the equipment, and Fig. 1 displays images of the instruments used to gather the dataset.

Sample recollection process

Three different trees were chosen for this study: avocado, olive, and grape which are some of the most representative cultivars in the region of Valparaíso, Chile.

For all the trees, the leaf collection process was as follows: Individual branches were cut directly from the trees, and healthy leaves were selected through visual inspection. We considered a healthy leaf to be one with no holes or internal ruptures in the leaf area and, if possible, with no visible symptoms of deterioration, such as brown spots. If a leaf showed any of these symptoms, we ensured that the majority of the leaf area remained unaffected.

Instrument	Technical specification
Multi-Spectral camera Micasense Red-Edge M	Bands: blue (B), green (G), red (R) red-edge (RE), near-infrared (NIR)
Spectrometer ASD Terraspec 4 Hi-Res	Range: 350 to 2500 nm
	Resolution: 3 nm at VNIR (350 to 1000 nm) 6 nm at SWIR (1001 to 2500 nm)
	Reproducibility: 0.1 nm
	Accuracy: 0.5 nm
Oven Mettler UN30	Temperature: – 5 to 100 °C with respect to room temperature
	Digital PID control
Balance Kern PFB 120.3	Maximum capacity: 120 g
	Readability: 0.001 g
	Reproducibility: 0.002 g
Plant nutrition analyzer TYS-4N	Chlorophyll range: 00.0 to 99.9 SPAD
	Chlorophyll accuracy: 1.00 SPAD
	Nitrogen range: 00.0 to 99.9 mg/g
	Nitrogen accuracy: 5%

Table 1. Equipment used to create the dataset.

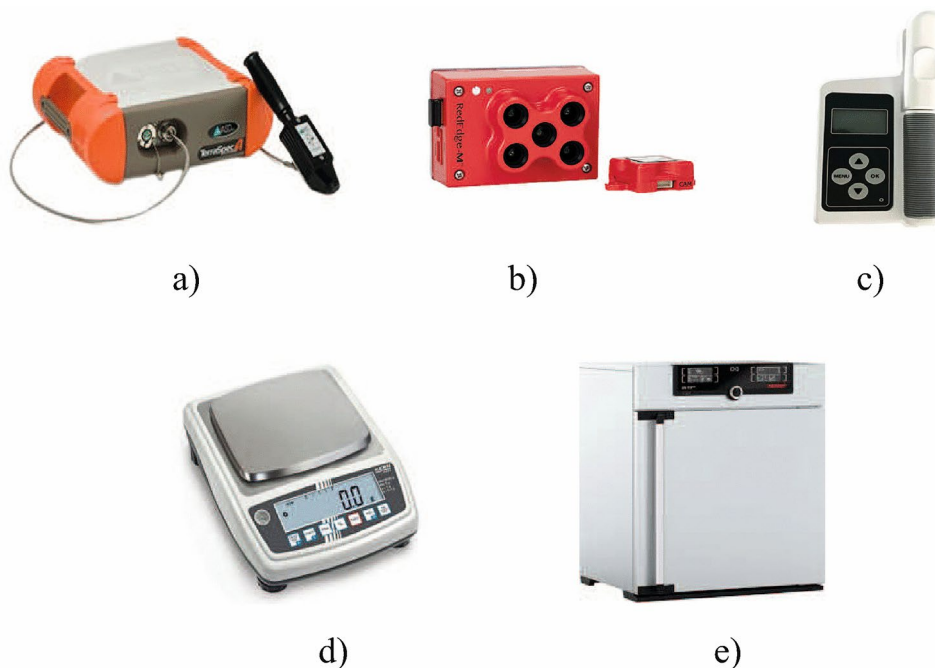


Fig. 1. Instruments used to create the dataset: (a) Spectrometer, (b) multi-spectral camera, (c) plant nutrition analyzer, (d) balance, (e) oven.

These leaves were then placed in bags and transported to the laboratory for analysis and measurement within the first hour of collection^{30,31}.

For the avocado trees, leaf collection occurred during April and May of 2023, with a total of 106 leaves collected. Olive leaf collection took place in July and August of 2023, resulting in a total of 111 leaves collected. Finally, for the grapes, the collection process was carried out between December 2023 and January 2024, with a total of 106 leaves collected. A summary of the total number of leaves and the collection process is provided in Table 2.

Measurement process

The dataset includes the following measurements: leaf weight, spectral reflectance ranging from 350 to 2500 nm, multi-spectral images from the blue, green, red, red-edge, and near-infrared bands, and chlorophyll and nitrogen content. Each measurement was recorded at five different dehydration stages of the leaves.

Tree species	Number of leaves	Months of recollection
Avocado	106	April and May of 2023
Olive	111	July and August of 2023
Grape	106	December of 2023 and January of 2024

Table 2. Leaf recollection summary.

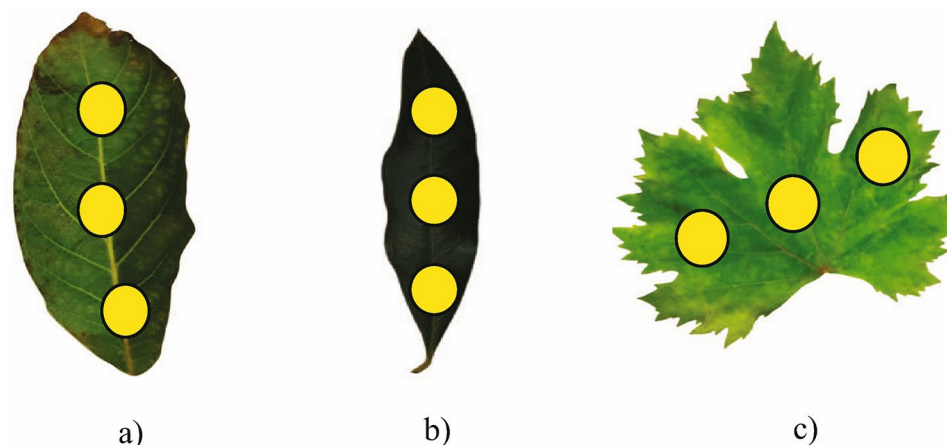


Fig. 2. Approximate positions where the probe was located in each leaf to collect the measurements in (a) avocado, (b) olive and (c) grape leaves.

The drying process follows the guidelines provided by the authors in³⁰, which are standard practices for analyzing fuel moisture content in leaves. The process begins with fresh leaves plucked from the branches. First, the spectral reflectance was recorded using the ASD TerraSpec Hi-Res spectrometer. The probe and light source were turned on for at least 15 minutes to warm up, as per the vendor's instructions. Next, the instrument parameters were optimized, and the measurements were calibrated using a white Spectralon reflectance panel.

Once the calibration and warming up of the instrument were completed, measurements of each leaf were taken. The contact probe was placed flush against each sample, which was positioned on a white background (Spectralon surface). Three individual measurements were collected from different zones of the leaf: one at the bottom, one in the middle, and one at the top. The approximate positions where the probe was placed to collect the measurements are shown in Fig. 2. The final spectral reflectance for each leaf was calculated as the average of these three measurements.

Next, the chlorophyll and nitrogen content were measured using the TYS-4N plant nutrition meter. The leaves were placed in the instrument's probe, and again, three measurements were taken per leaf from the bottom, middle, and top zones. The final measurement was recorded as the average value.

The leaves were then weighed using the electronic balance. Finally, multi-spectral images were captured. The MicaSense RedEdge camera was mounted on a tripod at a height of 45 cm above the leaves. Each image contained small batches of samples (4 leaves for avocado, 6 to 8 leaves for olive, and 1 leaf for grape samples), with the leaves placed against a white background (white paper sheets). To ensure consistent lighting conditions during the measurement process, artificial white light was used, and windows and other natural light sources were blocked. The artificial white light used to capture the multi-spectral images is provided by a fluorescent lamp with a color temperature of 4100 K and a luminous flux of 2850 lumens. The lamps are mounted on the ceiling of the laboratory.

The dehydration proceeded as follows: After the initial measurements were collected, the leaves were placed in an oven set to a temperature of 65 °C. This temperature was selected based on common practices for assessing the flammable content of leaves³⁰. The leaves were dried for 15 minutes. After this process, they were removed from the oven, and a new set of measurements was recorded at the first drying stage.

This process was repeated two more times, resulting in additional sets of measurements at the second and third drying stages. Finally, the leaves were placed in the oven for 24 h to remove all remaining traces of water. After this period, a final set of measurements was recorded with the completely dry leaves.

Both the spectrometer and the nutrition meter were calibrated after each batch of leaves was measured at each drying stage, sixteen leaves for avocado leaves and eight leaves for grape leaves. For olive leaves, calibration was performed after sixteen samples were measured, but a complete batch was of 32 leaves. The spectrometer was calibrated using a white Spectralon surface, and the instrument parameters were optimized. The nutrition meter was calibrated using its internal reference, following the vendor's instructions. Images of the calibration panels were collected after all measurements were completed in each trial.

The laboratory temperature was kept between 20 and °C using an air conditioning system and the relative humidity varied between 55 and 70%.

This measurement process resulted in five distinct sets of measurements at various dehydration stages of the leaves: fresh leaves, dehydration stages one through three, and the completely dehydrated stage. This process allows the determination of the percentual water content of the leaves at each stage.

Multi-spectral image post-processing

Radiometric calibration was performed for each band of the multi-spectral images. To achieve this, images of a calibrated reflectance panel were captured during each trial. The images were then post-processed using the metadata from each file, following the vendor's instructions. The radiometric calibration process was conducted using the free Python software provided by Micasense (<https://github.com/micasense/imageprocessing>).

With the calibrated images, a segmentation process was conducted to extract individual leaves from the batches. The segmentation process is as follows:

- The binarization of the image was performed using a threshold method. The value of the threshold was chosen using the Otsu method³².
- With the binarized image, the individual contours were found for each leaf.
- The original image was masked with each contour.
- Each leaf was cropped from the original image, and saved in a separate file. The segmentation algorithm was implemented in a Python v3.11 environment using the OpenCV library. The results of the segmentation algorithm can be seen in Fig. 3.

A visual assessment was performed for each leaf segmented by the algorithm. In some cases, two or more leaves were too close to each other, resulting in them being segmented together. In other instances, parts of the leaf were mistakenly identified as the background. In such cases, manual segmentation was performed using the Image Segmenter app in MATLAB.

Leaf water content

Based on the weight of the leaves, the leaf water content can be expressed as fuel moisture content using either a dry basis or a fresh basis³⁰. The equations to calculate the fuel moisture content are shown below:

$$FMC_f = \frac{W_{f,t} - W_d}{W_{f,t}} \times 100 \quad (1)$$

$$FMC_d = \frac{W_{f,t} - W_d}{W_d} \times 100 \quad (2)$$

In the equations FMC_f is the fuel moisture content using a fresh base, FMC_d is the fuel moisture content using the dry base, $W_{f,t}$ is the leaf weight at the drying stage t , W_d is the dry leaf weight after the 24 h in the oven.

Data records

All of the data can be found in this repository³³.

The dataset is organized into three main folders: Avocado, Olive, and Vineyard, each containing data specific to the respective species (the Vineyard folder contains the files for the grape samples). Within each of these folders, there is a subfolder called Multi-spectral images. This subfolder contains the multi-spectral images for each leaf sample, with five bands and five drying stages, resulting in a total of 25 images per leaf.

Each image is a .tif file, with a resolution of 16 bits. The size of the files and the dimensions of the images vary due to the fact that the leaves were cropped and the background was subtracted.

The naming convention is as follows, each file has the name *leaf###dy_x.tif*, where ### is a three-digit number identifying the number of the sample, y is the drying stage (ranging from 0 to 4, with 0 representing the fresh stage and 4 representing the completely dehydrated stage), and x is the multi-spectral band, which can take the following values: 1 (blue), 2 (green), 3 (red), 4 (near-infrared), and 5 (red edge). The Avocado and Vineyard Multi-spectral image folders contain 2,600 files each, while the Olive folder contains 2,775 files.

There are other files in the folders for each species, these files are: **FMC_species.mat**, **Nitrogen_species.mat**, **Chlorophyll_species.mat**, **Weight_species.mat**, **Spectral_species.mat**, with *species* being either avocado, olive or vineyard. These files can be read using Matlab software. Once these files are loaded, matrices containing the chlorophyll content, nitrogen content, weight, and fuel moisture content with a dry and fresh basis will appear in the working space, all these matrices are of size $N \times 5$, N number of samples, and five drying stages. When the *Spectral_species* is loaded, five variables will appear in the workspace, these matrices are of size $N \times 2151$, N is the number of samples which depends on the species, and the 2151 are the reflectance values in the wavelength range from 350 to 2500 nm. Each matrix contains a drying stage.

The tree structure of the dataset and a Matlab workspace when all the variables have been loaded, can be seen in Fig. 4, and a sample of the complete 25 images can be seen in Fig. 5.

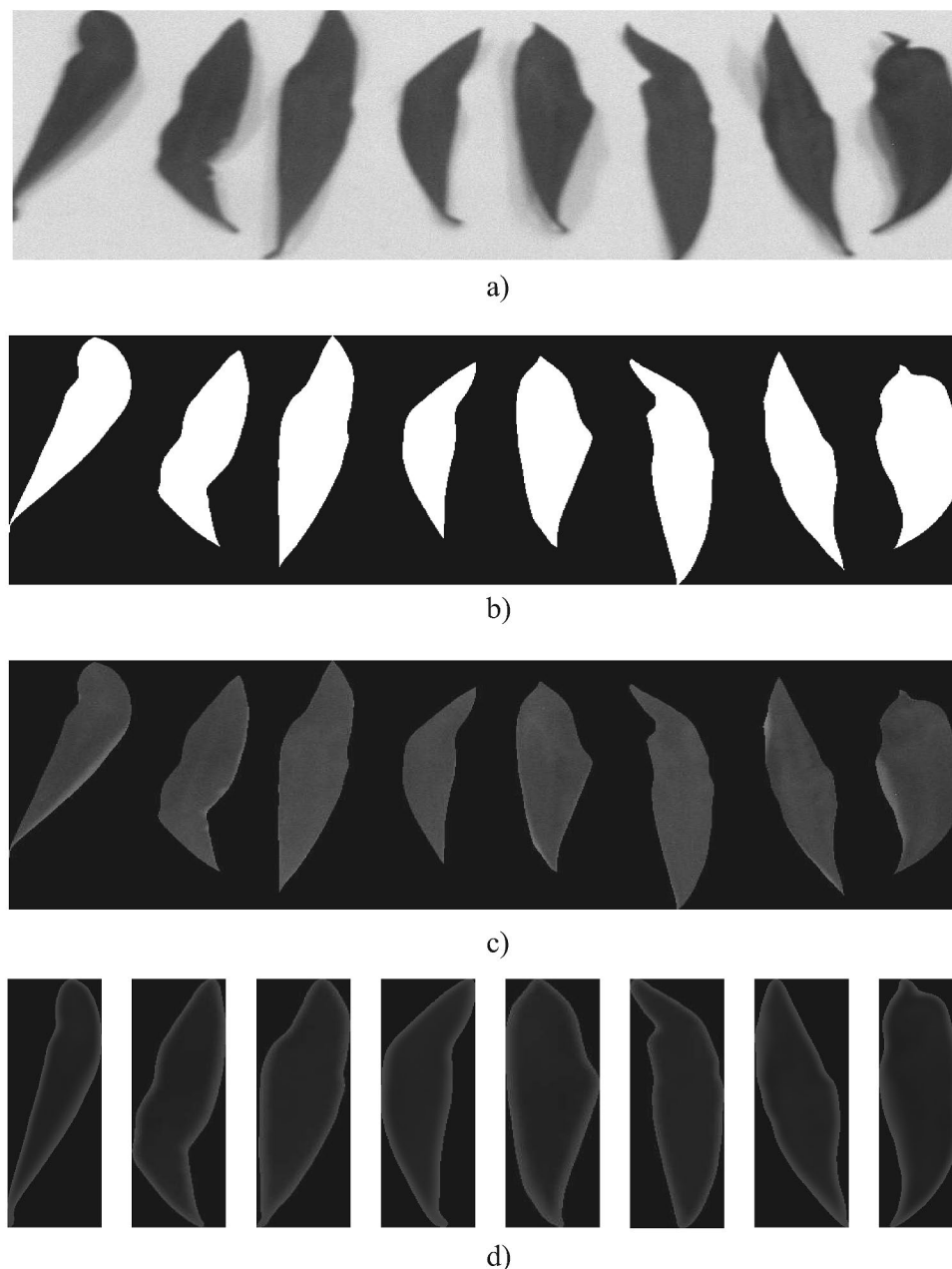


Fig. 3. Segmentation algorithm: (a) A multispectral image of a batch of olive leaves, (b) the binarization result using the Otsu threshold, (c) the result of masking the original image with the binarized image, and (d) the individual leaves generated.

Technical validation

All experiments conducted to gather the dataset followed the protocols specified by previous authors^{30,31,34}. The drying process was implemented to achieve various levels of leaf water content, which will aid in analyzing the spectral response of leaves to different levels of water stress. The drying temperature for all species was selected based on established protocols to ensure that leaves do not lose any flammable gases during the process. These temperatures have been tested for forest trees such as *Eucalyptus globulus* and *Pinus radiata*^{30,31,35,36}, as well as for crops such as potatoes³⁷.

The results of the spectral reflectance values for all species and drying stages are shown in Figs. 6, 7, and 8. Additionally, Fig. 9 displays the mean values at different drying stages for the three species. As depicted in the figures, the areas in the short-wave infrared (SWIR) region exhibit changes in response across the different drying stages. This phenomenon is attributed to the response of infrared wavelengths to water molecules^{30,34}. Particularly, reflectance values at bands at around 975 nm, 1200 nm, 1400 nm and 1900 nm systematically increase across the drying stages due to the vibrational bonds of water molecules³⁸.



Fig. 4. (a) File tree with all the files and folders contained in the dataset, (b) variables created in a Matlab workspace once all the .mat files have been loaded.

For the water content of the leaves, it is evident that as dehydration progresses, each leaf in all species shows a considerable loss of water content. Table 3 presents information on the water content for each species at each dehydration stage, including the mean, maximum (max), minimum (min), and standard deviation (std). The completely dehydrated stage was excluded from the table since the water content is zero in all cases.

Figure 10 shows the violin plots for additional measurements from the dataset, including chlorophyll content and leaf water content expressed as fuel moisture content on both a fresh and dry basis. As expected, at the final dehydration stage, all values of fuel moisture content are zero, due to the complete removal of water traces during the drying process in the oven.

As it can be seen the variation in the spectra, is related to the variation in the physical traits presented by the leaves, particularly the leaf moisture content. Variations in the leaf moisture content presented in the violin plots are also reflected in the spectra measurements, from Figs. 6, 7, and 8. Suggesting that leaf moisture content can be estimated using hyperspectral information as suggested by³⁰.

Another notable result is that in all the species the median value of the chlorophyll content increases with the dehydration stage. SPAD values is computed using two transmittance values, from red light at 650 nm and infrared light at 940 nm. The equation is shown below³⁹:

$$SPAD = k \times [\log(T_{940}) - \log(T_{650})] + C \quad (3)$$

where T_x is the transmittance at x wavelength, $\log$ is the common logarithm function, k is a confidential proportional coefficient and C is a compensation value in the instrument software³⁹.

It has been demonstrated that the SPAD values are highly correlated with the chlorophyll content per area⁴⁰, however, the SPAD value can be affected by external conditions such as illumination conditions, nitrogen treatment⁴⁰ and water stress induced by floodings^{41,42}.

Thus, it is reasonable to expect variations in the SPAD value, with the different drying stages of the leaves. Figure 9 shows the mean reflectance values across the spectrum. In the avocado, and olive plots there is a clear change in the region from around 750 nm to 1000 nm, where the fresh stage shows a steep change in the reflectance values, whereas in the dehydration stages a smoother change in the values is observed. This changes in the reflectance values shows a correlation with the increase in the SPAD values that is shown in the violin plots from Fig. 10, particularly from olive and avocado where the biggest change in the mean SPAD value occurs from the fresh stage to the first dehydration stage. In the grape leaves, such changes in the reflectance values from region 750–1000 nm are not so evident, thus the SPAD measurements remain with little variation to their

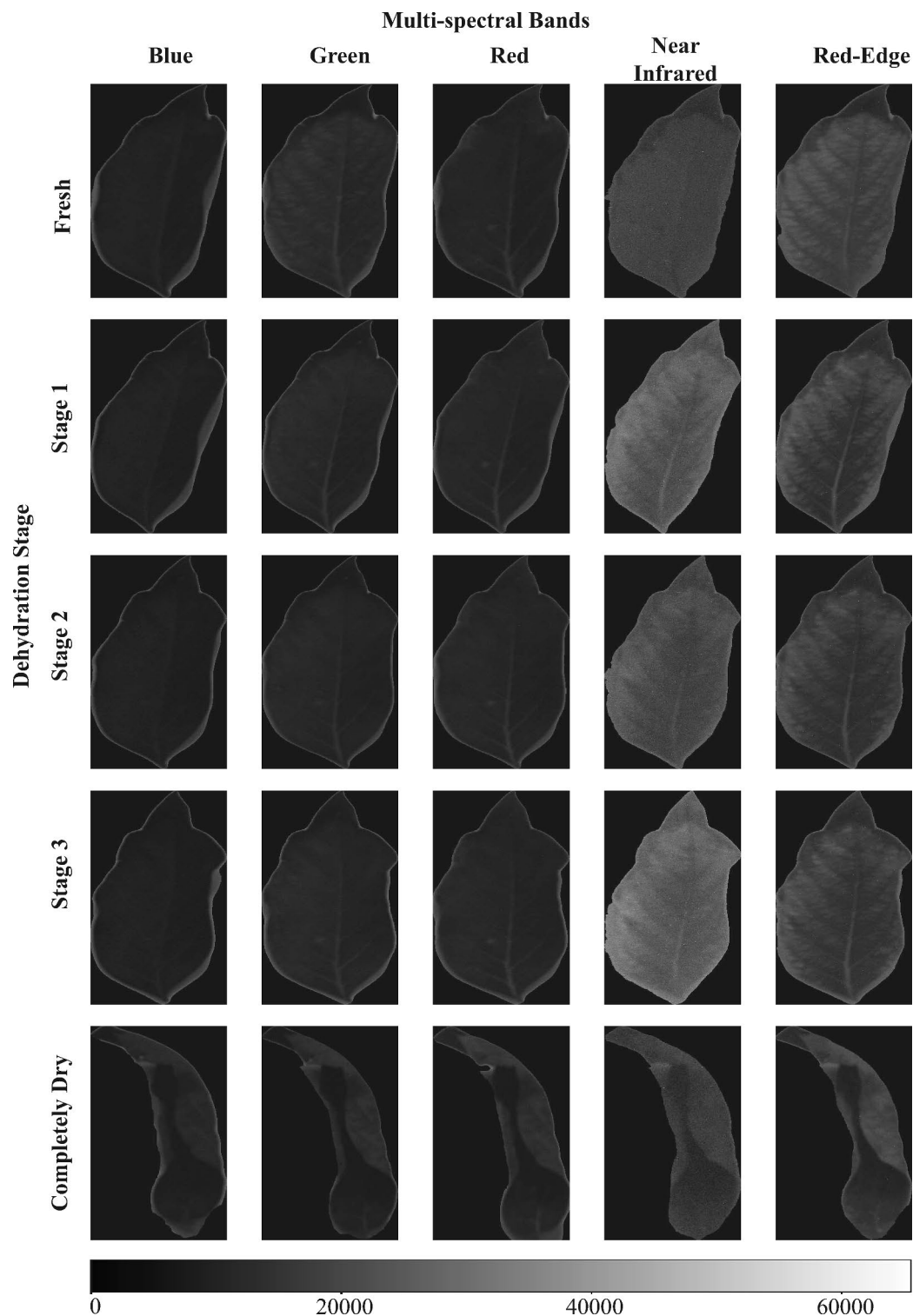


Fig. 5. Images of the five bands and five dehydration stages from an avocado leaf. to d.5) dehydration stage 3, e.1) to e.5) completely dry leaf.

mean value. Table 4 shows the mean, maximum (max), minimum (min), and the standard deviation (std) for the chlorophyll content for each species at each drying stage.

A t-test was conducted to determine if there are significant differences between the means of measurements at different drying stages. The test showed that the expected values for FMCf and FMCd are significantly different from each other at every drying stage for each of the three species. However, for chlorophyll content, dehydration stages 1 through 3 are not significantly different from each other in the avocado and grape samples.

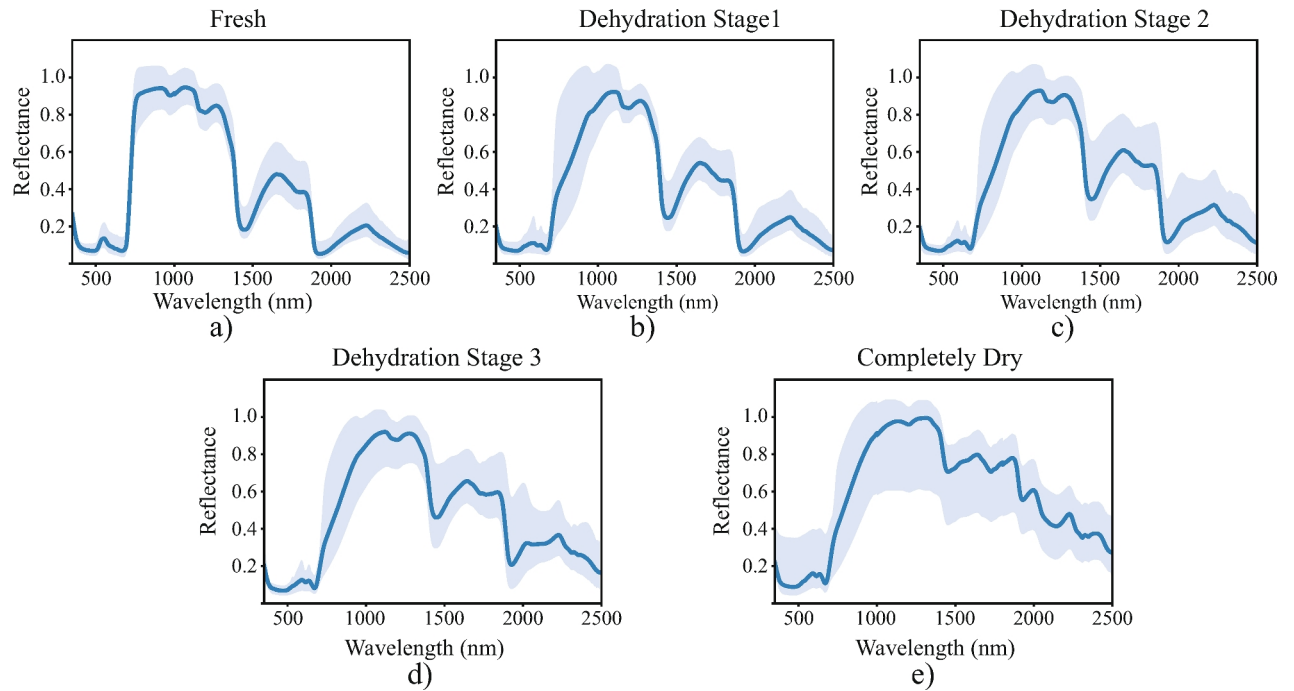


Fig. 6. Spectral reflectance for the avocado leaves, the mean value is darker and the shaded area corresponds to the maximum and minimum values, (a) fresh leaves, (b) first dehydration stage, (c) second dehydration stage, (d) third dehydration stage, and (e) completely dehydrated leaves.

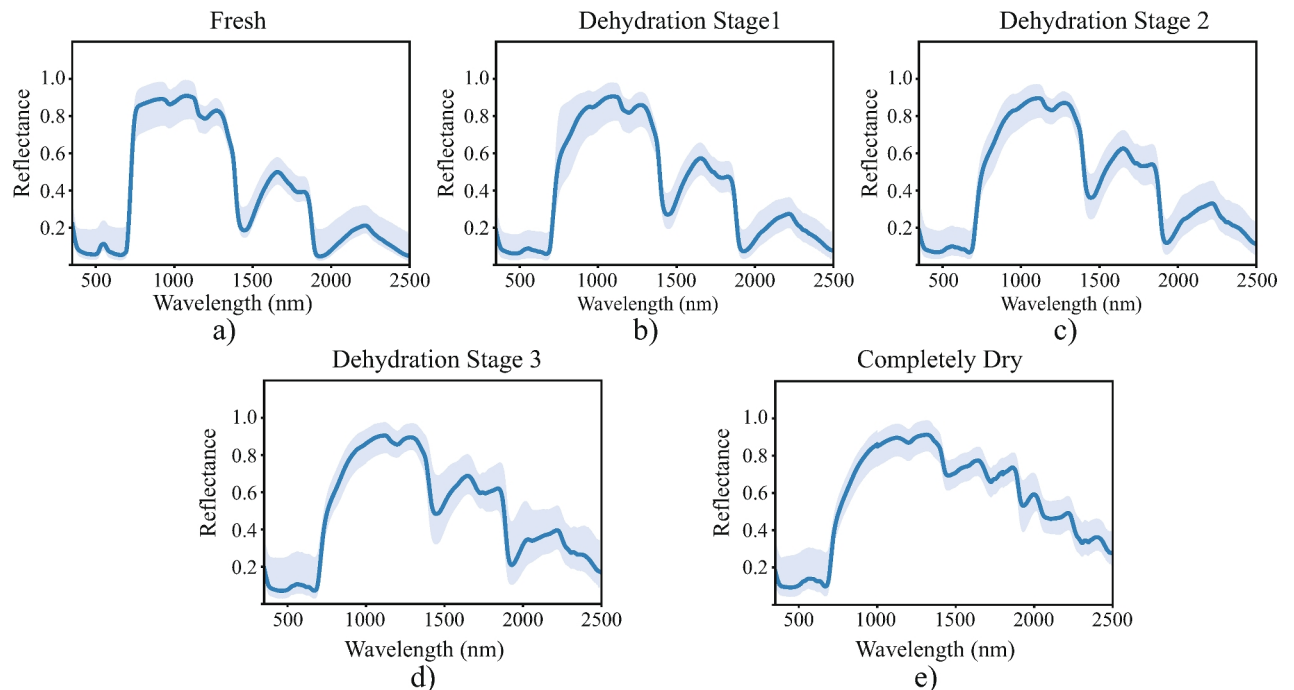


Fig. 7. Spectral reflectance for the olive leaves, the mean value is darker and the shaded area corresponds to the maximum and minimum values, (a) fresh leaves, (b) first dehydration stage, (c) second dehydration stage, (d) third dehydration stage, and (e) completely dehydrated leaves.

In contrast, for the olive dataset, the means of the three stages are significantly different from each other, except between stages 2 and 3 of dehydration.

Finally, to further validate the spectral dataset, we replicated the studies conducted by Villacres et al.³⁰, in which water-related vegetation indices were computed and linear regressions were used to assess their

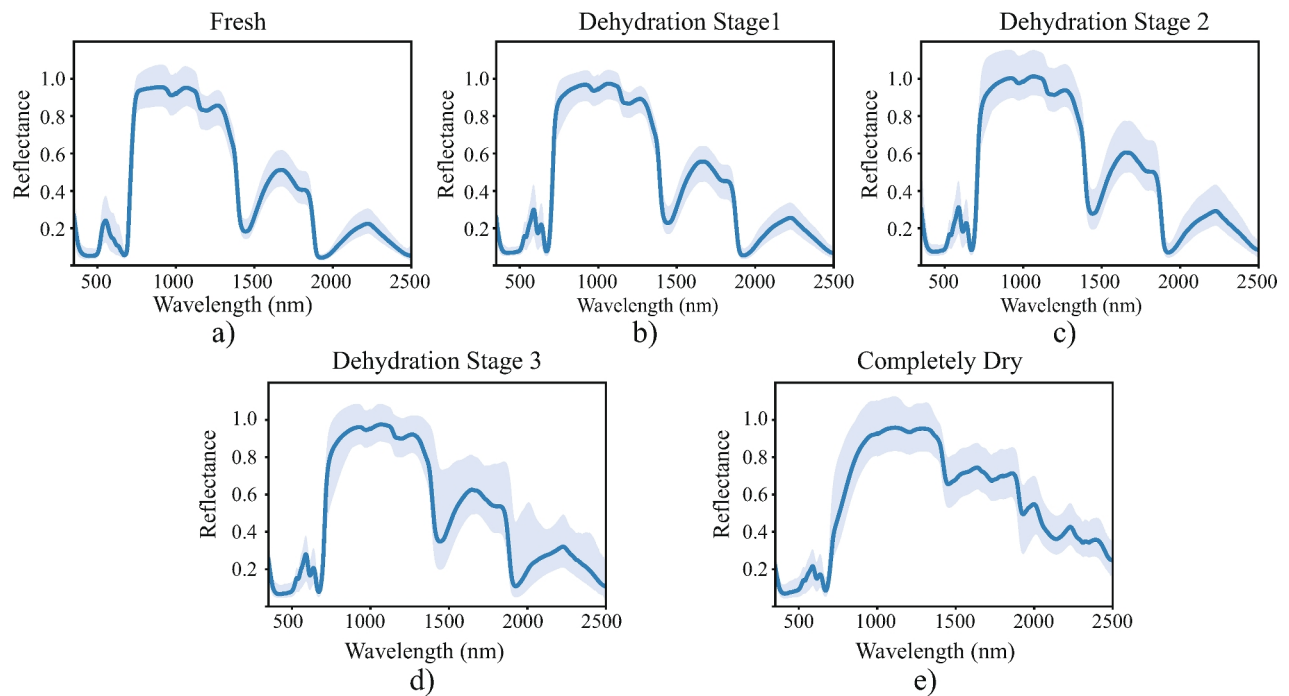


Fig. 8. Spectral reflectance for the grape leaves, the mean value is darker and the shaded area corresponds to the maximum and minimum values, (a) fresh leaves, (b) first dehydration stage, (c) second dehydration stage, (d) third dehydration stage, and (e) completely dehydrated leaves.

relationships. For our analysis, we selected the top two performing vegetation indices based on the correlation factor (R^2): the Leaf Water Index (LWI) and the Double Difference Index (DDI), as well as one of the least performing indices, the Normalized Difference Water Index 1 (NDWI1). The equations for these indices are provided below, where R_x represents the reflectance value at the x wavelength:

$$LWI = \frac{R_{1300}}{R_{1450}} \quad (4)$$

$$DDI = 2R_{1530} - R_{1005} - R_{2055} \quad (5)$$

$$NDWI1 = \frac{R_{860} - R_{1240}}{R_{860} + R_{1240}} \quad (6)$$

Table 5 shows the equations ($y = ax + b$) and the evaluation metrics of correlation factor R^2 and root mean squared error. The linear plots for the three indices and the corresponding tree species are shown in Fig. 11. As can be seen, the results obtained from Villacres et al.,³⁰ and in this technical validation, imply that there is a linear relationship between the fuel moisture content in vegetation and some vegetation indices, particularly LWI and DDI, however, the metrics in³⁰, reach higher values, particularly the highest correlation factor achieved in³⁰ in the LWI is of 0.96 and our result is of 0.84. This can be attributed to different factors, such as variations in the reflectance values across the different species, and a more widespread result in the computation of the VIs in this study.

Usage notes

There are numerous applications for this dataset. First, the data collection process was carried out following strict protocols, and it included information not provided by other datasets particularly multi-spectral images. Also, the various levels of leaf water content allow the exploration of the relationship between this biotic factor and the respective spectral signature in the range from 350 to 2500 nm, and also with the limited information provided by the multi-spectral cameras.

The use of multi-spectral images can be used to perform classification according to the dehydration level such as the study conducted in⁴³, but applied to agricultural species.

Finally, the addition of SPAD values can be explored through the use of vegetation indices and multi-spectral images. Furthermore, the relationship between the SPAD values and the leaf moisture content can be explored with the use of the current dataset.

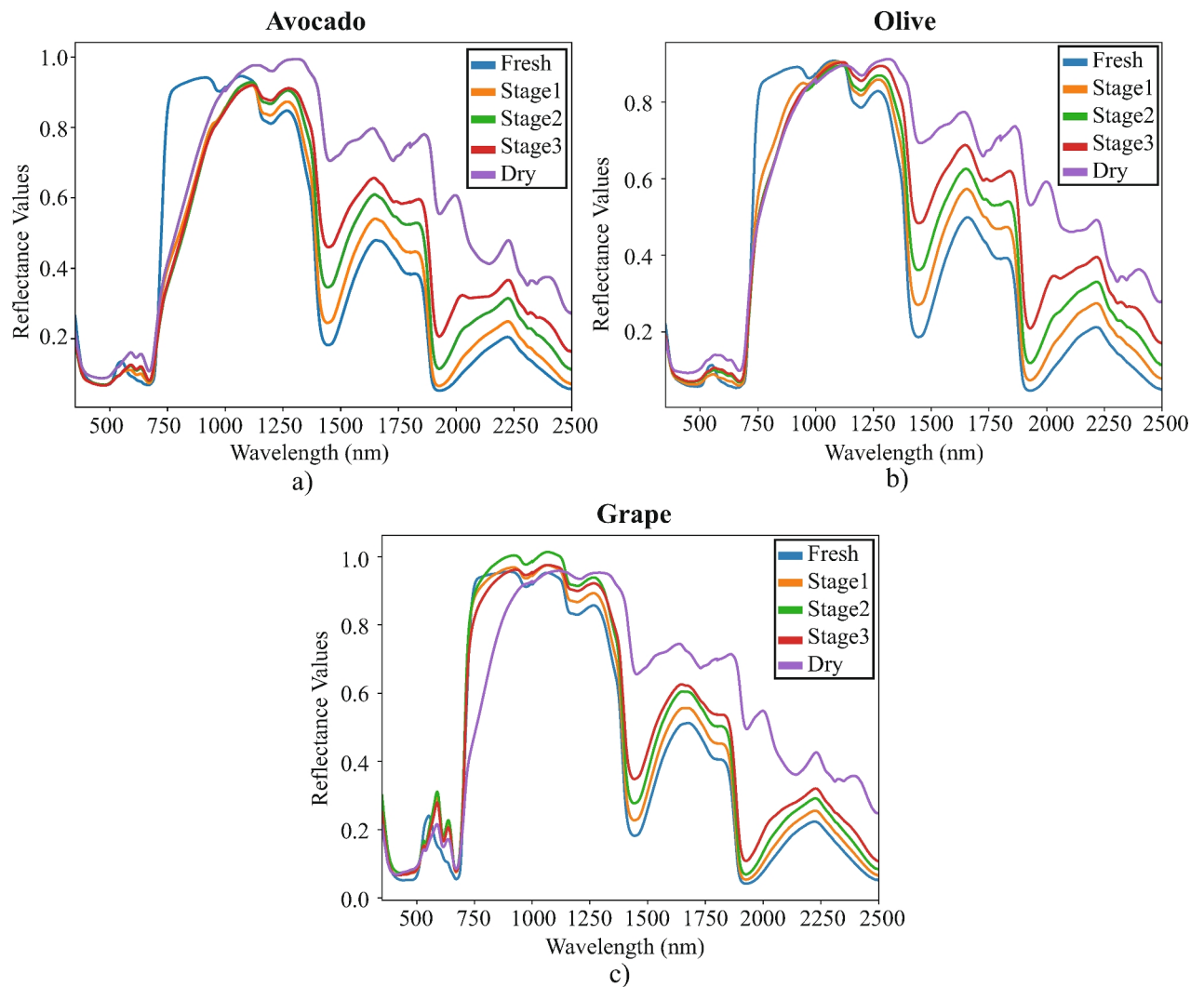


Fig. 9. Mean reflectance value for the full spectrum at the five different drying stages: (a) avocado samples, (b) olive samples, (c) grape samples.

Species	Drying stage	FMC_d				FMC_f			
		Mean	Max	Min	Std	Mean	Max	Min	Std
Avocado	Fresh	135.00	196.36	86.50	19.25	57.16	66.26	46.38	3.60
Avocado	Stage 1	96.48	132.21	37.42	16.32	48.73	56.93	27.23	4.60
Avocado	Stage 2	61.04	92.14	9.82	16.34	37.21	47.98	8.94	7.01
Avocado	Stage 3	32.66	66.37	2.15	14.36	23.73	39.89	2.10	8.34
Olive	Fresh	120.81	154.11	100.84	11.37	54.60	60.65	50.21	2.28
Olive	Stage 1	80.60	114.21	62.96	11.09	44.43	53.32	38.64	3.32
Olive	Stage 2	50.41	87.42	23.77	11.74	33.12	46.65	19.20	5.15
Olive	Stage 3	25.17	55.41	3.28	10.94	19.52	35.65	3.17	6.90
Grape	Fresh	246.18	377.41	103.33	54.40	70.27	79.05	50.82	5.57
Grape	Stage 1	170.50	233.33	68.35	34.70	62.28	70.00	40.60	6.03
Grape	Stage 2	113.05	155.90	32.27	30.91	51.87	60.92	24.40	8.42
Grape	Stage 3	67.24	124.20	3.70	30.74	37.89	55.39	3.57	12.98

Table 3. Descriptive statistics about the leaf moisture content at each drying stage.

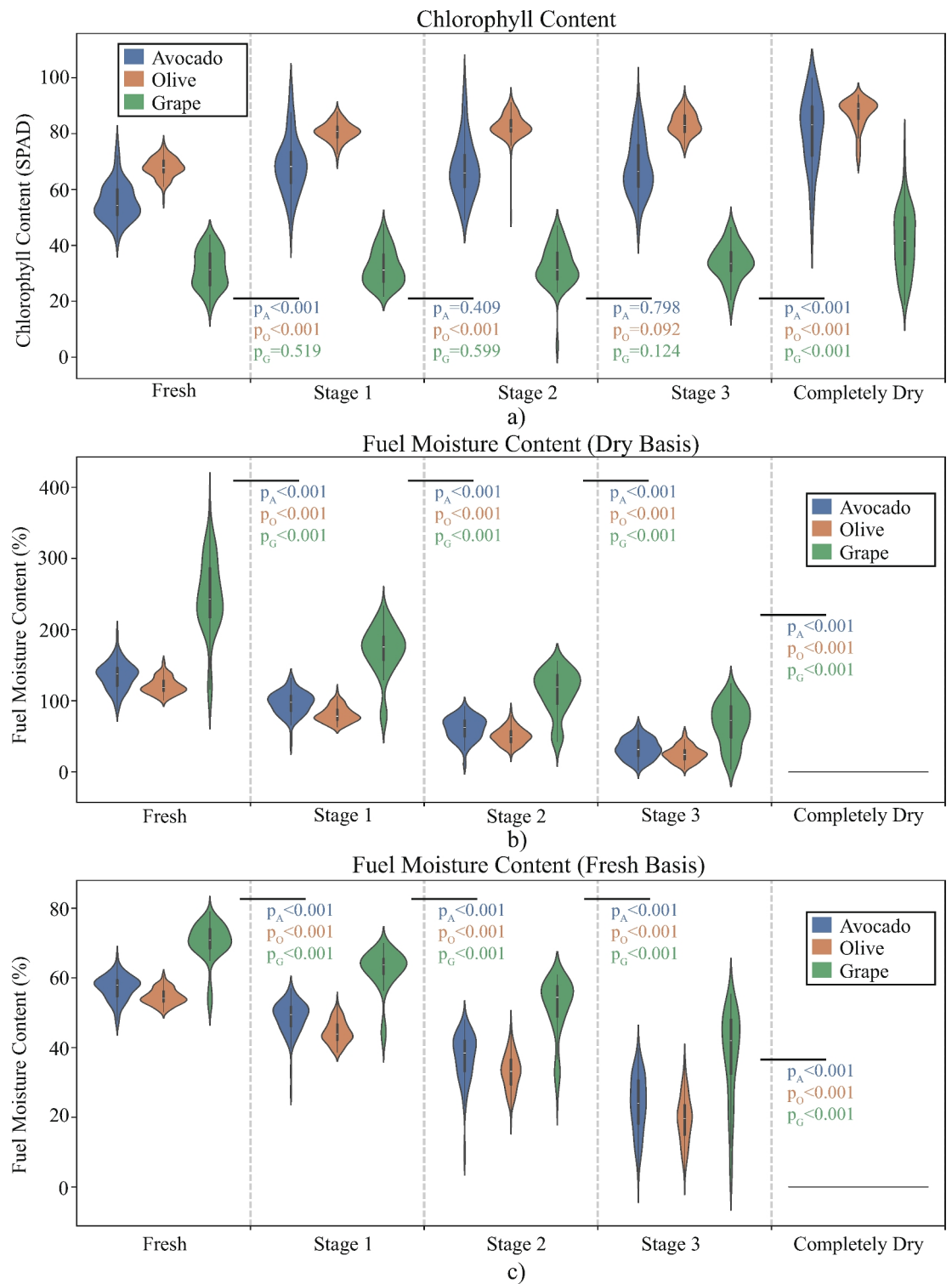


Fig. 10. Violin plots for (a) chlorophyll content, (b) fuel moisture content using a dry basis, and (c) fuel moisture content using a fresh basis. The graphs also show the p values for a t-test to check if the expected values between the means of each drying stage are significantly different from each other.

Species	Drying stage	Mean	Max	Min	Std
Avocado	Fresh	55.63	77.53	41.13	6.78
Avocado	Stage 1	68.85	96.97	43.83	10.26
Avocado	Stage 2	67.65	99.90	49.23	10.56
Avocado	Stage 3	80.32	68.02	95.50	10.35
Avocado	Dry	80.33	99.90	12.20	14.24
Olive	Fresh	67.85	76.60	56.10	11.37
Olive	Stage 1	80.69	88.80	70.10	11.09
Olive	Stage 2	82.62	92.80	50.40	11.74
Olive	Stage 3	83.62	93.90	74.60	10.94
Olive	Dry	87.14	93.90	70.20	5.42
Grape	Fresh	31.33	43.80	16.40	6.86
Grape	Stage 1	32.01	47.60	21.70	6.40
Grape	Stage 2	32.51	47.00	3.70	7.36
Grape	Stage 3	34.00	48.60	16.60	6.48
Grape	Dry	41.41	76.00	6.48	11.53

Table 4. Descriptive statistics about the chlorophyll content at each drying stage.

Species	VI	Linear Model	R ²	RMSE
Avocado	LWI	$y = 0.0578x + 0.9613$	0.82	0.562
Avocado	DDI	$y = -0.00835x + 0.1337$	0.84	0.079
Avocado	NDWI1	$y = 0.002253x - 0.2112$	0.11	0.120
Olive	LWI	$y = 0.05598x + 0.9219$	0.84	0.489
Olive	DDI	$y = -0.00794x + 0.1159$	0.92	0.048
Olive	NDWI1	$y = 0.001943x - 0.1213$	0.47	0.042
Grape	LWI	$y = 0.04393x + 1.21$	0.84	0.507
Grape	DDI	$y = -0.00583x + 0.02765$	0.86	0.064
Grape	NDWI1	$y = 0.001745x - 0.06844$	0.77	0.026

Table 5. Linear regression results for the prediction of fuel moisture content using different vegetation indices.

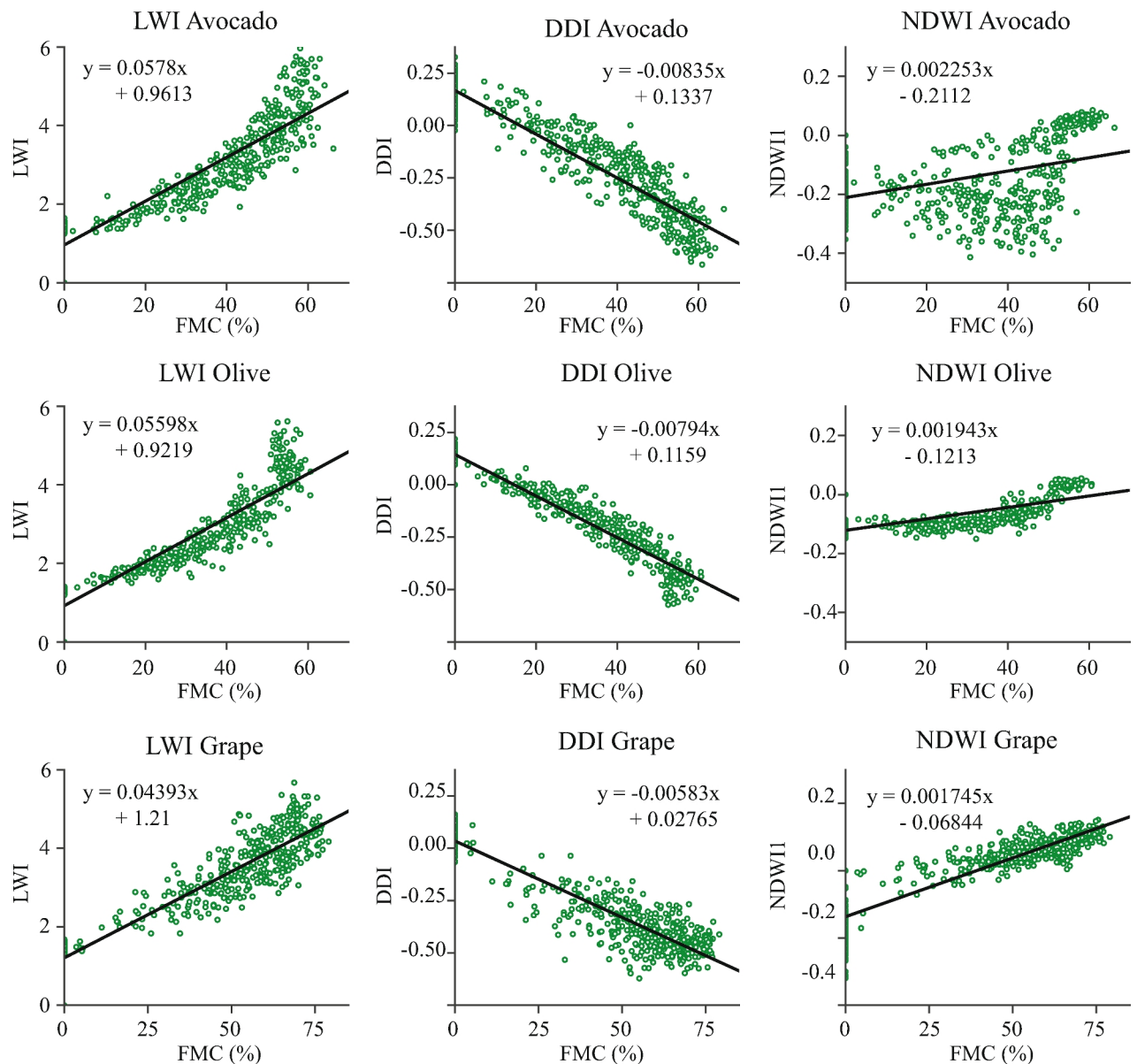


Fig. 11. Linear regression results for the tested VIs for fuel moisture content prediction.

Data availability

All the data is available at this repository DOI: <https://doi.org/10.6084/m9.figshare.26950660.v2>.

Code availability

A sample code is available in the project, and it is used to open the spectral information and plot the reflectance values and their respective wavelength. The code is provided as a Matlab Script, within the dataset this file is named *Sample_code_avocado.mlx*.

Received: 13 September 2024; Accepted: 6 January 2025

Published online: 23 January 2025

References

1. Chilean Ministry of Environment. Chile's second biennial update report on climate change. https://unfccc.int/files/national_reports/non-annex_i_parties/biennial_update_reports/application/pdf/bur2_chile_english2017.pdf (2016).
2. Santibanez, F. El cambio climático y los recursos hídricos de Chile. <https://www.odepa.gob.cl/wp-content/uploads/2018/01/cambioClim12parte.pdf> (2018).
3. Vougioukas, S. G. Agricultural robotics. *Annu. Rev. Control Robot. Auton. Syst.* **2**, 365–392. <https://doi.org/10.1146/annurev-control-053018-023617> (2019).
4. Estrada, J. S., Fuentes, A., Reszka, P. & Cheein, F. A. Machine learning assisted remote forestry health assessment: A comprehensive state of the art review. *Front. Plant Sci.* **14**, 1139232. <https://doi.org/10.3389/fpls.2023.1139232> (2023).

5. Lu, J. et al. Combining plant height, canopy coverage and vegetation index from UAV-based RGB images to estimate leaf nitrogen concentration of summer maize. *Biosyst. Eng.* **202**, 42–54. <https://doi.org/10.1016/j.biosystemseng.2020.11.010> (2021).
6. Milas, A. S., Romanko, M., Reil, P., Abeyasinghe, T. & Marambe, A. The importance of leaf area index in mapping chlorophyll content of corn under different agricultural treatments using UAV images. *Int. J. Remote Sens.* **39**, 5415–5431. <https://doi.org/10.1080/01431161.2018.1455244> (2018).
7. Eugenio, F. C. et al. Estimated flooded rice grain yield and nitrogen content in leaves based on RPAS images and machine learning. *Field Crops Res.* **292**, 108823. <https://doi.org/10.1016/j.fcr.2023.108823> (2023).
8. Brinkhoff, J., Dunn, B. W. & Robson, A. J. Rice nitrogen status detection using commercial-scale imagery. *Int. J. Appl. Earth Obs. Geoinf.* **105**, 102627. <https://doi.org/10.1016/j.jag.2021.102627> (2021).
9. Longfei, Z. et al. Improved yield prediction of ratoon rice using unmanned aerial vehicle-based multi-temporal feature method. *Rice Sci.* **30**, 247–256. <https://doi.org/10.1016/j.rsci.2023.03.008> (2023).
10. Marang, I. J. et al. Machine learning optimised hyperspectral remote sensing retrieves cotton nitrogen status. *Remote Sens.* **13**, 1428. <https://doi.org/10.3390/rs13081428> (2021).
11. Shanmugapriya, P. et al. Cotton yield prediction using drone derived LAI and chlorophyll content. *J. Agrometeorol.* **24**, 348–352. <https://doi.org/10.54386/jam.v24i4.1770> (2022).
12. Messina, G. et al. Monitoring onion crop “cipolla rossa di tropea calabria IGP” growth and yield response to varying nitrogen fertilizer application rates using UAV imagery. *Drones* **5**, 61. <https://doi.org/10.3390/drones5030061> (2021).
13. Peng, J., Manevski, K., Kørup, K., Larsen, R. & Andersen, M. N. Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. *Field Crops Res.* **268**, 108158. <https://doi.org/10.1016/j.fcr.2021.108158> (2021).
14. Liu, J., Zhu, Y., Tao, X., Chen, X. & Li, X. Rapid prediction of winter wheat yield and nitrogen use efficiency using consumer-grade unmanned aerial vehicles multispectral imagery. *Front. Plant Sci.* **13**, 1032170. <https://doi.org/10.3389/fpls.2022.1032170> (2022).
15. Narmilan, A. et al. Predicting canopy chlorophyll content in sugarcane crops using machine learning algorithms and spectral vegetation indices derived from UAV multispectral imagery. *Remote Sens.* **14**, 1140. <https://doi.org/10.3390/rs14051140> (2022).
16. Teodoro, P. E. et al. Predicting days to maturity, plant height, and grain yield in soybean: A machine and deep learning approach using multispectral data. *Remote Sens.* **13**, 4632. <https://doi.org/10.3390/rs13224632> (2021).
17. Tedesco, D. et al. Use of remote sensing to characterize the phenological development and to predict sweet potato yield in two growing seasons. *Eur. J. Agronomy* **129**, 126337. <https://doi.org/10.1016/j.eja.2021.126337> (2021).
18. Jacquemoud, S., Bidel, L., Francois, C. & Pavan, G. Angers leaf optical properties database (2003).
19. Hosgood, B. et al. Leaf optical properties experiment database (lopex93) (1993).
20. Feret, J.-B. et al. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sens. Environ.* **112**, 3030–3043. <https://doi.org/10.1016/j.rse.2008.02.012> (2008).
21. Wu, J. et al. Convergence in relationships between leaf traits, spectra and age across diverse canopy environments and two contrasting tropical forests. *New Phytol.* **214**, 1033–1048. <https://doi.org/10.1111/nph.14051> (2016).
22. Wu, Jin and Prohaska, Neill and Saleska, Scott. 2012-leaf-reflectance-spectra-of-tropical-trees-in-tapajos-national-forest (2020).
23. Ely, K.S., Serbin, S.P., Lieberman-Cribbin, W. & Rogers, A. Leaf spectra, structural and biochemical leaf traits of eight crop species. <https://doi.org/10.21232/C2GM2Z> (2018).
24. Burnett, A.C., Serbin, S.P., Davidson, K.J., Ely, K.S. & Rogers, A. Hyperspectral leaf reflectance, biochemistry, and physiology of droughted and watered crops. <https://doi.org/10.21232/UTK8ZAW4> (2020).
25. Ryckewaert, M. et al. Hyperspectral images of grapevine leaves including healthy leaves and leaves with biotic and abiotic symptoms. *Sci. Data* **10**, 743. <https://doi.org/10.1038/s41597-023-02642-w> (2023).
26. Vayssade, J.-A. Multi-spectral leaf segmentation for crop/weed identification. <https://doi.org/10.15454/JMKP9S> (2021).
27. Przybylska, M. S. et al. Aradiv: A dataset of functional traits and leaf hyperspectral reflectance of *Arabidopsis thaliana*. *Sci. Data* **10**, 314. <https://doi.org/10.1038/s41597-023-02189-w> (2023).
28. Huemmrich, K. & Campbell, P. Tundra plant leaf-level spectral reflectance and chlorophyll fluorescence, 2019–2021. <https://doi.org/10.3334/ORNDAAC/2005> (2022).
29. Georgantopoulos, P. et al. A multispectral dataset for the detection of *Tuta absoluta* and *Leveillula taurica* in tomato plants. *Smart Agric. Technol.* **4**, 100146. <https://doi.org/10.1016/j.atech.2022.100146> (2023).
30. Villacrés, J., Arevalo-Ramirez, T., Fuentes, A., Reszka, P. & Cheein, F. A. Foliar moisture content from the spectral signature for wildfire risk assessments in Valparaíso-Chile. *Sensors* **19**, 5475. <https://doi.org/10.3390/s19245475> (2019).
31. Kumar, L. High-spectral resolution data for determining leaf water content in eucalyptus species: Leaf level experiments. *Geocarto Int.* **22**, 3–16 (2007).
32. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66. <https://doi.org/10.1109/TSMC.1979.4310076> (1979).
33. Estrada, J. S. A multi-spectral and hyperspectral image dataset for evaluating the health status of avocado, olive and vineyard. https://figshare.com/articles/dataset/_b_A_multi-spectral_and_hyperspectral_image_dataset_for_evaluating_the_health_status_of_avocado_olive_and_vineyard_b_/26950660. <https://doi.org/10.6084/m9.figshare.26950660.v2> (2024).
34. Arevalo-Ramirez, T., Villacrés, J., Fuentes, A., Reszka, P. & Auat Cheein, F. A. Moisture content estimation of *Pinus radiata* and *Eucalyptus globulus* from reconstructed leaf reflectance in the SWIR region. *Biosyst. Eng.* **193**, 187–205. <https://doi.org/10.1016/j.biosystemseng.2020.03.004> (2020).
35. Simeoni, A. et al. Flammability studies for wildland and wildland-urban interface fires applied to pine needles and solid polymers. *Fire Saf. J.* **54**, 203–217 (2012).
36. Jervis, F. X. & Rein, G. Experimental study on the burning behaviour of *Pinus halepensis* needles using small-scale fire calorimetry of live, aged and dead samples. *Fire Mater.* **40**, 385–395 (2016).
37. Liu, N. et al. Spectral characteristics analysis and water content detection of potato plants leaves. *IFAC-PapersOnLine* **51**, 541–546. <https://doi.org/10.1016/j.ifacol.2018.08.152> (2018).
38. Purwadi, I., Erskine, P. D. & van der Ent, A. Reflectance spectroscopy as a promising tool for ‘sensing’ metals in hyperaccumulator plants. *Planta* **258**, 41. <https://doi.org/10.1007/s00425-023-04167-3> (2023).
39. Nauš, J., Prokopová, J., Rebiček, J. & Špundová, M. SPAD chlorophyll meter reading can be pronouncedly affected by chloroplast movement. *Photosynth. Res.* **105**, 265–271. <https://doi.org/10.1007/s11120-010-9587-z> (2010).
40. Xiong, D. et al. SPAD-based leaf nitrogen estimation is impacted by environmental factors and crop leaf characteristics. *Sci. Rep.* **5**, 13389. <https://doi.org/10.1038/srep13389> (2015).
41. Lao, Z. et al. Retrieval of chlorophyll content for vegetation communities under different inundation frequencies using UAV images and field measurements. *Ecol. Indic.* **158**, 111329. <https://doi.org/10.1016/j.ecolind.2023.111329> (2024).
42. Saghoui el idrissi, I. et al. Water stress effect on durum wheat (*Triticum durum* desf.) advanced lines at flowering stage under controlled conditions. *J. Agric. Food Res.* **14**, 100696. <https://doi.org/10.1016/j.jafr.2023.100696> (2023).
43. Estrada, J. S., Zanartu, M., Demarco, R., Fuentes, A. & Cheein, F. A. Water content classification on leaves based on multi-spectral imagery and machine learning techniques for wildfire prevention. In *2024 IEEE International Conference on Industrial Technology (ICIT)* 1–7 (2024). <https://doi.org/10.1109/ICIT58233.2024.10540771>.

Acknowledgements

The authors acknowledge and thank ANID AFB240002, basal center AC3E, ANID SCIA-ANILLO ACT210052, and ANID national doctorate scholarship, folio No. 21231129 for funding the present work.

Author contributions

J.S.E contributed to the data recollection, data analysis, software, and writing process. R.D conceived the project, data revision and manuscript revision. C.J. contributed with data analysis and software. M.Z contributed with manuscript revision. A.F conceived the project. F.A.C conceived the project, designed the experiments, data analysis, and manuscript revision.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.S.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025