

Article

High-Throughput Identification and Prediction of Early Stress Markers in Soybean Under Progressive Water Regimes via Hyperspectral Spectroscopy and Machine Learning

Caio Almeida de Oliveira ¹, Nicole Ghinzelli Vedana ¹ , Wesley Augusto Mendonça ¹, João Vitor Ferreira Gonçalves ¹ , Dheyne Heyre Silva de Matos ¹, Renato Herrig Furlanetto ², Luis Guilherme Teixeira Crusiol ³, Amanda Silveira Reis ¹, Werner Camargos Antunes ⁴ , Roney Berti de Oliveira ¹ , Marcelo Luiz Chicati ¹ , José Alexandre M. Demattê ⁵ , Marcos Rafael Nanni ¹  and Renan Falcioni ^{1,4,*} 

- ¹ Graduate Program in Agronomy, State University of Maringá, Av. Colombo, 5790, Maringá 87020-900, Paraná, Brazil; pg55482@uem.br (C.A.d.O.); pg405864@uem.br (N.G.V.); pg405896@uem.br (W.A.M.); pg403805@uem.br (J.V.F.G.); ra143742@uem.br (D.H.S.d.M.); asreis@uem.br (A.S.R.); rboliveira@uem.br (R.B.d.O.); mlchicati@uem.br (M.L.C.); mrnanni@uem.br (M.R.N.)
- ² Gulf Coast Research and Education Center, University of Florida, Wimauma, FL 33598, USA; re.herrigfurlane@ufl.edu
- ³ Embrapa Soja (National Soybean Research Center—Brazilian Agricultural Research Corporation), Rodovia Carlos João Strass, s/n°, Distrito de Warta, Londrina 86001-970, Paraná, Brazil; luis.crusiol@colaborador.embrapa.br
- ⁴ Department of Biology, State University of Maringá, Av. Colombo, 5790, Maringá 87020-900, Paraná, Brazil; wcantunes@uem.br
- ⁵ Department of Soil Science, Luiz de Queiroz College of Agriculture, University of São Paulo, Av. Pádua Dias, 11, Piracicaba 13418-260, São Paulo, Brazil; jamdemat@usp.br
- * Correspondence: renanfalcioni@gmail.com or rfalcioni2@uem.br; Tel.: +55-443-0118-940



Academic Editor: Ruyin Cao

Received: 17 July 2025

Revised: 8 October 2025

Accepted: 9 October 2025

Published: 11 October 2025

Citation: Oliveira, C.A.d.; Vedana, N.G.; Mendonça, W.A.; Gonçalves, J.V.F.; Matos, D.H.S.d.; Furlanetto, R.H.; Crusiol, L.G.T.; Reis, A.S.; Antunes, W.C.; Oliveira, R.B.d.; et al. High-Throughput Identification and Prediction of Early Stress Markers in Soybean Under Progressive Water Regimes via Hyperspectral Spectroscopy and Machine Learning. *Remote Sens.* **2025**, *17*, 3409. <https://doi.org/10.3390/rs17203409>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Highlights

What are the main findings?

- Hyperspectral spectroscopy combined with machine learning enables high-accuracy, nondestructive prediction of early stress markers (pigments, osmolytes, antioxidants, cell wall compounds, and water status) in soybean under progressive drought via remote sensing and machine learning models.
- Tree-based ensemble and neural network models (e.g., random forest, MLP) achieved >95% accuracy in classifying drought severity, outperformed distance- and probability-based classifiers, and effectively distinguished eleven water regimes across the full or range UV–VIS–NIR–SWIR spectrum.

What is the implication of the main finding?

- The integration of hyperspectral sensors and machine learning provides a rapid, field-deployable solution for early drought detection and precision irrigation management in soybean, potentially reducing the reliance on time-consuming laboratory assays via remote sensing tools.
- Selecting minimal and informative spectral bands paves the way for simplified, cost-effective proximal or UAV-mounted sensors for large-scale drought phenotyping and smart agriculture applications.

Abstract

The soybean *Glycine max* (L.) Merrill is a key crop in Brazil's agricultural sector and is essential for both domestic food security and international trade. However, water

stress severely impacts its productivity. In this study, we examined the physiological and biochemical responses of soybean plants to various water regimes via hyperspectral reflectance (350–2500 nm) and machine learning (ML) models. The plants were subjected to eleven distinct water regimes, ranging from 100% to 0% field capacity, over 14 days. Seventeen key physiological parameters, including chlorophyll, carotenoids, flavonoids, proline, stress markers and water content, and hyperspectral data were measured to capture changes induced by water deficit. Principal component analysis (PCA) revealed significant spectral differences between the water treatments, with the first two principal components explaining 88% of the variance. Hyperspectral indices and reflectance patterns in the visible (VIS), near-infrared (NIR), and shortwave-infrared (SWIR) regions are linked to specific stress markers, such as pigment degradation and osmotic adjustment. Machine learning classifiers, including random forest and gradient boosting, achieved over 95% accuracy in predicting drought-induced stress. Notably, a minimal set of 12 spectral bands (including red-edge and SWIR features) was used to predict both stress levels and biochemical changes with comparable accuracy to traditional laboratory assays. These findings demonstrate that spectroscopy by hyperspectral sensors, when combined with ML techniques, provides a nondestructive, field-deployable solution for early drought detection and precision irrigation in soybean cultivation.

Keywords: agricultural management; biochemical prediction; computational intelligence; phenotyping in plants; precision agriculture; spectroscopy in plants; UV–VIS–NIR–SWIR sensors

1. Introduction

Soybean (*Glycine max* (L.) Merrill) plays a central role in Brazilian agriculture, underpinning both domestic food security and international trade [1,2]. Over the past decade, Brazil has solidified its position as the world's leading soybean producer, with output reaching a record 169.0 million tons in the 2024/25 marketing year, cultivated on some 47.4 million ha (USDA, 2024) [3]. This oilseed accounts for nearly 60% of global soybean exports, representing a cornerstone of the national economy and a critical source of foreign exchange [3]. Consequently, such prominence renders soybean both a driver of agronomic innovation and a sensitive indicator of environmental stressors.

In this context, drought is defined as a sustained soil water deficit that impairs plant physiological function and poses one of the greatest challenges to soybean productivity [4]. Water limitation disrupts carbon assimilation through stomatal closure, reduces the chlorophyll content, and accelerates the generation of reactive oxygen species (ROS), precipitating oxidative damage to the photosynthetic apparatus and membrane lipids [5,6]. Moreover, the accumulation of osmolytes such as proline and antioxidative compounds (e.g., flavonoids and phenolics) reflects a plant's protective response, linking water status to both photochemical efficiency and oxidative homeostasis [7].

Hyperspectral proximal sensing, which spans ultraviolet (UV), visible (VIS), near-infrared (NIR), and shortwave-infrared (SWIR) regions (350–2500 nm), offers a nondestructive window into these physiological and biochemical changes [8–10]. Narrowband features in the VIS region capture pigment absorption peaks (chlorophylls, carotenoids), whereas SWIR bands correspond to overtones of water and cell wall constituents (e.g., cellulose, lignin, and other compounds) [11]. Moreover, recent studies have demonstrated that spectroscopy via hyperspectral indices and continuum-removed reflectance metrics can reliably predict leaf water content and stress markers, with accuracies exceeding those of broadband sensors [12,13].

Narrowband reflectance vegetation indices (VIs) synthesize key wavelengths into simple ratios or differences, such as the NDVI, PRI and WBI, that increase sensitivity to chlorophyll content, photosynthetic efficiency and water status [14,15]. By condensing thousands of bands into targeted indices, VIs reduce noise and highlight physiological shifts, serving both as interpretable proxies of plant health and as robust input features for downstream models [16–18]. When integrated within machine learning frameworks, these indices improve prediction and characterization by emphasizing the most stress-responsive spectral signals while limiting redundant information [19].

Complementing spectral approaches, machine learning (ML) and deep learning (DL) techniques enable the classification and quantification of stress responses from high-dimensional reflectance data [20,21]. Algorithms such as random forest, support vector machines, and convolutional neural networks have achieved classification accuracies above 95% in distinguishing drought-stressed vs. well-watered soybeans by recognizing subtle spectral shifts associated with pigment degradation and osmotic adjustment [2,11,22]. These models not only automate stress detection but also provide real-time decision support in precision agriculture. For example, partial least squares or CNN-based classifiers can rapidly flag drought conditions with minimal false alarms, allowing timely irrigation or management interventions. The integration of ML thus adds robustness and scalability to hyperspectral drought diagnostics [23–25].

Critical to both spectral and ML-based methods is the targeted selection of informative wavelengths [26]. Meta-analyses have identified discrete bands at approximately 550–750 nm (the VIS to red-edge region) and in the SWIR near 1450 nm and 1900 nm as especially predictive of plant water status and structural changes [27]. The reflectance in the VIS to red-edge transition zone is highly sensitive to chlorophyll content and stress-induced pigment changes, whereas the reflectances at ~1450 nm and ~1900 nm correspond to major water absorption features where the LWC strongly influences the signal [2]. Leveraging these “spectral pin-points” can drastically reduce the data volume while preserving predictive power, enabling the development of simplified handheld sensors and UAV-mounted instruments tailored for drought monitoring. By focusing on a minimal set of diagnostic bands, one can design field-deployable multispectral systems that approximate the performance of full hyperspectral instruments in detecting water stress [28].

A comprehensive characterization of the drought response thus benefits from integrating multiple analytical layers, spectral indices, machine-learning classification, and biochemical assays to capture both macroscopic reflectance patterns and underlying molecular dynamics [29,30]. Such a multilayer approach can detect stress onset and quantify key metabolites before visible symptoms appear. Recent work has shown that hyperspectral models can predict drought-induced changes in leaf metabolites (e.g., proline, abscisic acid, and electrolyte leakage) with good accuracy, indicating that optical data contain biochemical information about stress [11,31,32]. By combining optical and biochemical perspectives, one can achieve robust early warning of water deficit and a mechanistic understanding of the stress response [33]. This framework promises more precise estimation of plant health and stress status, advancing both fundamental knowledge and practical management of drought in soybean [2,3,34].

In this study, we evaluated the efficacy of combined hyperspectral and ML methods for the nondestructive prediction of key physiological parameters (leaf water status and biochemical and metabolic parameters in cells) and oxidative stress biomarkers (flavonoids, proline, and phenolics) in soybean (*Glycine max* (L.) Merrill) plants under a gradient of water regimes. We hypothesize that a minimal suite of ~12 strategically selected wavelengths, coupled with ensemble learning models, can predict both drought severity and biochemical composition with accuracy comparable to that of traditional laboratory assays. Validating this hypothesis would demonstrate a rapid, field-deployable approach to soybean drought phenotyping, enabling early stress detection and informing water-management decisions.

2. Material and Methods

2.1. Plant Materials

Soybean seeds (*Glycine max* (L.) Merrill) were initially germinated under controlled laboratory conditions. Vigorous, morphologically uniform seedlings were selected and transplanted into 1 L plastic pots containing sterilized substrate. The plants were grown in a controlled-environment chamber, where the day/night temperatures were maintained at 26 °C/23 °C, the relative humidity was 70%, and a 16/8 h photoperiod was used. The photosynthetically active radiation was set at 500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and was calibrated via a LI-COR 1800 quantum sensor (LI-COR Inc., Lincoln, NE, USA).

After transplanting, the seedlings were irrigated to full substrate capacity for seven days to ensure acclimation. The plants were subsequently subjected to eleven distinct water regimes, defined as 100% (W100), 90% (W90), 80% (W80), 70% (W70), 60% (W60), 50% (W50), 40% (W40), 30% (W30), 20% (W20), 10% (W10), and 0% (W0) of the substrate field capacity. Watering volumes were determined gravimetrically, simulating a gradient of water restriction. Hoagland's solution was applied every two days to ensure an adequate nutrient supply.

The experiment was carried out in a completely randomized design, with eleven water regimes, eight plants per treatment, and three technical replicates per plant, resulting in a total of 264 analysed samples. Pots were randomly distributed within the chamber. The water regimes were maintained for 14 days, with the environmental and irrigation parameters continuously monitored to ensure experimental consistency.

2.2. Hyperspectral Reflectance Data

Hyperspectral reflectance spectra were obtained from the adaxial surface of fully expanded leaves via a FieldSpec 3 spectroradiometer (ASD Inc., Boulder, CO, USA) equipped with a PlantProbe[®] leaf clip. The measurements covered the 350–2500 nm spectral range (UV–VIS–NIR–SWIR). Prior to each measurement session, the instrument was calibrated with both a Spectralon[®] white reference and a dark reference to ensure accuracy. For each leaf, 50 consecutive scans were averaged to reduce noise and improve signal quality. All reflectance data were processed via ViewSpec Pro[®] software version 5 (ASD Inc., Boulder, CO, USA), following standard protocols for baseline correction and interpolation. Only reflectance values were considered for subsequent statistical and multivariate analyses [35]. All the measurements were conducted under controlled ambient light conditions.

2.3. Chlorophylls and Carotenoids Extraction

Simultaneous quantification of total chlorophyll (Chl), total carotenoids (Car), and flavonoids (Flv) was performed as described by Gitelson and Solovchenko (2018) [36] and Falcioni et al. (2022) [37], with adaptations for soybean. Leaf segments (0.5 cm²) were homogenized in 1.5 mL microtubes containing chloroform–methanol solution (2:1, *v/v*) in the presence of CaCO₃. After complete pigment extraction, distilled water (20% of the total extract volume) was added to promote phase separation. The samples were subsequently centrifuged at 15,000 rpm for 5 min to ensure clear separation of the polar and apolar phases. The absorbance was measured in a 96-well microplate using a Biochrom Asys UVM-340 microplate reader (Biochrom Ltd., Cambridge, UK) with ScanPlus VisibleWell[®] software version 1.0.2 (Biochrome Ltd., Milton Road, Cambridge, UK).

2.3.1. Quantification of Chlorophylls and Carotenoids

Chlorophyll a (Chl a), chlorophyll b (Chl b), total chlorophyll (Chl a + b), and carotenoids (Car; carotenes + xanthophylls) were quantified from the acetone phase by adding 200 μL of extract to each well. The absorbance was measured at 470, 652, and

665 nm, and 100% methanol was used as the blank. The base area (mg m^{-2}) and mass (mg g^{-1}) concentrations were calculated via the following equations:

$$\begin{aligned}\text{Chl a} &= 16.72 \times \text{Abs}_{665} - 9.16 \times \text{Abs}_{652} \\ \text{Chl b} &= 34.09 \times \text{Abs}_{652} - 15.28 \times \text{Abs}_{665} \\ \text{Chl a + b} &= \text{Chl a} + \text{Chl b} \\ \text{Car} &= \frac{1000 \times \text{Abs}_{470} - 1.63 \times \text{Chl a} - 104.96 \times \text{Chl b}}{221}\end{aligned}$$

2.3.2. Quantification of Flavonoids

Flavonoids (Flv) were quantified in the polar methanol extract. The upper phase, containing extrachloroplastic pigments, was used for total flavonoid content determination by absorbance at 358 nm ($\epsilon_{358} = 25 \text{ mM}^{-1} \text{ cm}^{-1}$; Gitelson & Solovchenko, 2018) [36].

2.4. Quantification of Proline

The proline content in the leaf samples was determined following the methods of Falcioni et al. (2025) [11], with adaptations for microplate analysis. Fresh leaf segments (100 mg) were homogenized in 2 mL of 3% (*w/v*) sulfosalicylic acid and centrifuged at 15,000 rpm for 10 min. A 100 μL aliquot of the supernatant was transferred to a 96-well microplate, to which 100 μL of acid ninhydrin solution (prepared by mixing 1.25 g of ninhydrin in 30 mL of glacial acetic acid and 20 mL of 6 M phosphoric acid) and 100 μL of glacial acetic acid were added. The plate was sealed and incubated at 95 °C for 1 h. After cooling to room temperature, 200 μL of toluene was added to each well, and the mixture was agitated for 30 s. The absorbance of the chromophore-containing toluene phase was measured at 520 nm via a microplate reader. The proline concentration was calculated from a standard curve constructed with L-proline and expressed as $\mu\text{mol proline per gram of fresh mass}$ ($\mu\text{mol g}^{-1}$ FM). All analyses were performed in triplicate.

2.5. Quantification of Soluble Phenolic Compounds (Phe)

The soluble phenolic compounds (Phe) in the leaf samples were quantified via an adapted Folin–Ciocalteu method (Ragae, 2006) [38]. The methanolic extracts (150 μL) were mixed with 70 μL of Folin–Ciocalteu reagent (1 M), 140 μL of sodium carbonate (Na_2CO_3 , 3.56 M), and 850 μL of deionized water in 2 mL tubes. The reaction mixture was incubated in the dark for 50 min at room temperature and then centrifuged at 15,000 rpm for 2 min. The absorbance of the supernatant was measured at 725 nm via a spectrophotometer. The phenolic content was calculated against a gallic acid standard curve ($\hat{Y} = 87.651x + 1.6515$; $R^2 = 0.993$) and expressed as gallic acid equivalents per sample. All measurements were performed in triplicate.

2.6. Preparation of Protein-Free Cell Wall Fraction (PFCW) and Quantification of Lignin and Cellulose

Protein-free cell wall fractions (PFCWs) were prepared from dried, powdered leaf tissue. Aliquots of 150 mg were weighed into 2 mL microtubes. The samples were sequentially washed five times with 50 mM potassium phosphate buffer (pH 7.0), five times with Triton X-100 (pH 7.0), four times with 1 M NaCl (pH 7.0), four times with distilled water, and three times with acetone. After each wash, the samples were centrifuged at 15,000 rpm for 3 min. The final pellets were oven-dried at 60 °C for 24 h and used as the PFCW fraction, which was free of both water-soluble polar and apolar compounds.

2.6.1. Lignin Content Determination

The lignin content in the PFCW fraction was quantified via the acetyl bromide method. For each sample, 20 mg of PFCW was transferred to a new microtube and mixed with 130 μL of freshly prepared acetyl bromide solution (25% *v/v* in glacial acetic acid). The samples were incubated at 70 °C for 30 min and then cooled rapidly on ice. Subsequently, 0.24 mL of 2 M NaOH, 0.02 mL of 5 M hydroxylamine-HCl, and 1.6 mL of glacial acetic acid were added for complete solubilization of the lignin extract. The samples were subsequently centrifuged at $1400 \times g$ for 5 min. The lignin content was determined spectrophotometrically at 280 nm ($\epsilon = 22.1 \text{ g L}^{-1} \text{ cm}^{-1}$) via a FlexStation 3 plate spectrophotometer (Molecular Devices LLC., San Jose, CA, USA) and SoftMax[®] Pro Software version 5 (Molecular Devices LLC., San Jose, CA, USA). The results are expressed as mg lignin per g of PFCW. All analyses were performed in triplicate.

2.6.2. Cellulose Content Determination

The cellulose was quantified according to standard protocols with adaptations for leaf tissue. The dried tissue samples were incubated at 70 °C for 1 h. The ethanol was then replaced with acetic/nitric acid solution for extraction, and the mixture was subsequently discarded. The samples were washed with distilled water and then treated with freshly prepared anthrone in sulfuric acid. Quantification was performed at 620 nm using a Biochrom Asys UVM-340 microplate reader (Biochrom Ltd., Cambridge, UK). The cellulose content was expressed as glucose equivalents (μmol glucose per g dry mass) and was calculated according to a glucose standard curve. All analyses were performed in triplicate.

2.7. Antioxidant Activity (RSA%)

Antioxidant activity, expressed as relative scavenging activity (RSA%), was determined via the DPPH (2,2-diphenyl-1-picrylhydrazyl) free radical assay following Falcioni et al. (2025) [11] with adaptations for soybean leaves. Methanolic leaf extracts (50 μL) were added to 200 μL of 1 mM DPPH solution in each well of a quartz 96-well microplate. The mixture was agitated and incubated in the dark at room temperature for 60 min. The absorbance was then measured at 515 nm via a microplate spectrophotometer (Biochrom Asys UVM-340). The RSA (%) was calculated as follows:

$$\text{RSA (\%)} = \left[1 - \left(\frac{A_{\text{sample}}}{A_{\text{blank}}} \right) \right] \times 100$$

where A_{sample} is the absorbance of the reaction mixture and A_{blank} is the absorbance of the DPPH solution without extraction. Higher RSA values indicate greater antioxidant activity. All measurements were performed in triplicate.

2.8. Electrolyte Leakage (ELK%)

Electrolyte leakage (ELK%) was determined to assess membrane integrity, following standard protocols with adaptations for soybean leaves. Fresh leaf discs (0.5 cm diameter) were rinsed thoroughly with deionized water to remove surface-adhered electrolytes. The discs were placed in test tubes containing 10 mL of deionized water and incubated at room temperature for 24 h. The initial conductivity (C_1) of the bath solution was measured via a conductivity meter (model, manufacturer). The samples were then autoclaved at 121 °C for 20 min to ensure complete membrane rupture and allowed to cool to room temperature, after which the final conductivity (C_2) was recorded. Electrolyte leakage was calculated as the percentage ratio of initial conductivity to total conductivity:

$$\text{ELK (\%)} = \left(\frac{C_1}{C_2} \right) \times 100$$

All measurements were performed in triplicate. The results are expressed as the percentage of total electrolytes released, reflecting the degree of cell membrane damage.

2.9. Relative Water Content (RWC%)

The relative water content (RWC) was determined to evaluate the leaf water status. Fresh leaf discs were collected and immediately weighed to obtain the fresh mass (FM). The discs were then floated on distilled water in Petri dishes for 24 h at 4 °C in the dark to reach full turgidity. After gentle blotting to remove surface water, the turgid mass (TM) was recorded. The samples were subsequently dried at 70 °C for 48 h and weighed to determine the dry mass (DM). The RWC was calculated as follows:

$$\text{RWC (\%)} = \left(\frac{\text{FM} - \text{DM}}{\text{TM} - \text{DM}} \right) \times 100$$

All measurements were performed in triplicate. The results are presented as the percentage of water content relative to full turgor.

2.10. Statistical Analyses

2.10.1. Analysis of Variance and Descriptive Statistics

Descriptive statistical analysis was conducted for all agronomic, physiological, biochemical, and spectral datasets, including the calculation of the mean, standard error of the mean (SEM), minimum and maximum values, and coefficient of variation (CV%). Treatment differences were evaluated by analysis of variance (ANOVA), with statistical significance considered at $p < 0.05$. Multiple mean comparisons were performed via Duncan's test at the same significance level. Relationships between physiological variables, vegetation indices, and growth parameters were investigated via Pearson's correlation test. All univariate analyses and the generation of summary tables were performed via custom Statistica 10[®] (StatSoft Inc., Tulsa, OK, USA), SigmaPlot 10.0[®] (Systat Inc., Santa Clara, CA, USA), CorelDraw 2020[®] (Corel Corp., Ottawa, ON, Canada) and Python scripts (version 3.11) [39].

2.10.2. Principal Component Analyses (PCA)

Principal component analysis (PCA) was applied to physiological and spectral data to reduce dimensionality and identify major groupings and response patterns to water treatments. The optimal number of principal components was determined according to the cumulative variance explained. All PCA analyses were conducted in Python, adopting a significance level of $p < 0.05$. In addition to scores and clustering, the proportion of explained variance (scree plot), spectral loadings, and regression coefficients for each principal component were extracted to determine the most informative wavelength regions. These outputs supported subsequent feature selection and modelling steps.

2.10.3. Vegetation Indices (VIs)

A comprehensive set of vegetation indices was calculated to evaluate their sensitivity in detecting differences among water regimes, foliar pigmentation, photochemical efficiency, water status, pigment content, and structural–physiological alterations in soybean leaves. The indices analysed were, in order, NDVI (normalized difference vegetation index), GNDVI (green normalized difference vegetation index), EVI (enhanced vegetation index), SAVI (soil-adjusted vegetation index), OSAVI (optimized soil-adjusted vegetation index), MSAVI2 (modified soil-adjusted vegetation index 2), SIPI (structure insensitive pigment index), PSSRc (pigment specific simple ratio—carotenoids), RARS (red-edge anthocyanin reflectance signal), WBI (water band index), MSI (moisture stress index), NDII (normalized

difference infrared index), NDMI (normalized difference moisture index), NDDI (normalized difference drought index), NMDI (normalized multi-band drought index), NDWI1640 (normalized difference water index with 1640 nm band), NDWI2130 (normalized difference water index with 2130 nm), ARI1 (anthocyanin reflectance 1), ARI2 (anthocyanin reflectance 2), CRI1 (carotenoid reflectance 1), CRI2 (carotenoid reflectance 2), VOG1 (Vogelmann red edge 1), VOG2 (Vogelmann red edge 2), NPQI (normalized phaeophytinization index), and PRI (photochemical reflectance index).

The mean values for each index were compared across treatments, and their relationships with physiological parameters were assessed via Pearson's correlation coefficient. This integrated approach enabled robust discrimination of plant responses to water deficit and supported the identification of informative spectral markers for physiological and structural changes in the leaves.

2.10.4. Hierarchical and Cluster Analysis

Pairwise Euclidean distances were calculated between the mean spectral profiles of each water regime. A hierarchical clustering algorithm was applied to these distances to assess the similarity and grouping of treatments. The resulting Euclidean distance matrix was visualized as a heatmap, and a corresponding dendrogram was constructed to illustrate hierarchical relationships among treatments. All calculations and visualizations were performed in Python.

2.10.5. Machine Learning Models

Predictive classification models were developed to discriminate among water regimes via hyperspectral reflectance data. The algorithms employed were support vector machine (SVM), random forest (RF), k-nearest neighbor (KNN), naive Bayes (NB), decision tree (DT), logistic regression (LR), gradient boosting (GBoost), and multilayer perceptron (MLP classifier). All the data were split such that 60% of the samples were used for model training (calibration), 40% were used for internal validation, and an independent external set was reserved for model testing. The training and test groups were defined on the basis of both random sampling and hierarchical clustering outcomes to ensure representativeness across treatments.

Model performance was evaluated via the external test set, with efficiency defined as models achieving training accuracy above 95%. The performance metrics included the confusion matrix, accuracy, precision, recall (sensitivity), and F1 score for each class (i.e., water regime). The confusion matrices of all the models are presented for direct visual comparison of classification performance and error patterns across all the treatments.

2.10.6. Correlation and Heatmap Analyses

Pearson's correlation coefficients ($p < 0.001$) were calculated for all the physiological, biochemical, structural, and spectral variables. For each trait, a correlation matrix was generated with all individual wavelengths, and the results were visualized as heatmaps, with color scales representing the strength and direction of the correlation (from -1 to $+1$). Additionally, a second heatmap was produced to visualize the correlation matrix among all measured variables, allowing for an integrated assessment of relationships between biochemical, physiological, and structural traits.

2.10.7. Selection of Responsive Spectral Bands

To identify the most responsive wavelengths for each physiological and biochemical parameter, multiple variable selection algorithms were applied via custom Python scripts. The following methods were implemented: partial least squares regression (PLSR), variable importance in projection (VIP), interval partial least squares (iPLS-VIP), the genetic algo-

rithm (GA), random forest (RF), competitive adaptive repeated sampling (CARS), Boruta, Lasso, mutual information, recursive feature elimination (RFE), and linear discriminant analysis (LDA). For each variable, the top 50 wavelengths were selected on the basis of the importance ranking or selection criterion of each algorithm. The convergence and specificity of selected bands across algorithms are visualized as dot plots, where wavelength positions (x-axis) are grouped by algorithm (y-axis). This integrative approach enabled robust identification of the spectral regions most informative for each target trait.

2.10.8. Partial Least Squares Regression (PLSR)

Partial least squares regression (PLSR) was used to predict physiological, biochemical, and agronomic parameters from the hyperspectral reflectance data. For each variable, the spectral and target data were synchronized and standardized via z score normalization prior to model fitting. The datasets were randomly split into calibration (training) and validation (testing) sets, typically using 60% of the samples for model calibration and 40% for validation; the exact proportion was user-defined.

The optimal number of PLSR components was selected by the user on the basis of data dimensionality and prediction stability. The PLSR models were fitted via the NIPALS algorithm. For each target variable, model performance was evaluated in both the training and testing sets by calculating the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and prediction bias. Additionally, linear regression analysis between the observed and predicted values was performed to estimate the slope and intercept of the regression line. All observed and predicted values, together with the minimum and maximum ranges, were recorded.

Model predictions and metrics were visualized in scatter plots of observed versus predicted values for each variable, highlighting the 1:1 line, fitted regression, and color-coded gradients by the observed variable.

2.10.9. Analysis of Optimized Hyperspectral Vegetation Indices

To identify the most informative spectral regions for each physiological and biochemical parameter, all possible combinations of two wavelengths within the hyperspectral reflectance range were evaluated via a generalized normalized difference vegetation index (NDVI) formula:

$$\text{HVI}(\lambda_1, \lambda_2) = \frac{R(\lambda_1) - R(\lambda_2)}{R(\lambda_1) + R(\lambda_2)}$$

where $R(\lambda)$ is the standardized reflectance at wavelength λ .

For each pairwise combination of wavelengths, the resulting HVI was calculated across all samples and correlated with the target variable via the coefficient of determination (R^2). This generated a two-dimensional spectral map (contour map) of R^2 values, highlighting spectral regions where the NDVI-like index was most predictive of the trait. All calculations were performed via custom Python scripts. The resulting R^2 maps were visualized for each variable, with the colour scale representing the strength of the correlation and the axes corresponding to λ_1 and λ_2 . This approach enabled the identification of hyperspectral index combinations most responsive to variations in physiological, biochemical, and structural leaf parameters (Figure 1 and Table 1).

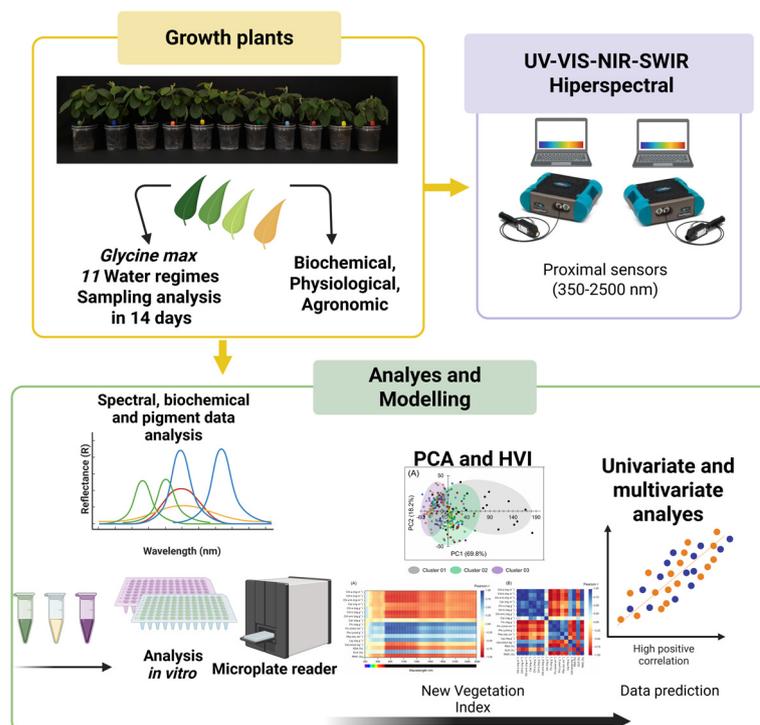


Figure 1. Experimental workflow for the assessment of the biochemical, physiological, and spectral responses of *Glycine max* under eleven water regimes. The diagram illustrates the experimental pipeline, including the cultivation of soybean plants under eleven distinct water regimes with sampling every 14 days. Biochemical, physiological, and agronomic parameters were analysed in parallel with hyperspectral data acquisition (UV–VIS–NIR–SWIR range, 350–2500 nm) via proximal sensors. Spectral, biochemical, and pigment analyses were performed both *in vitro* and through multivariate modelling approaches, including principal component analysis (PCA), hyperspectral vegetation indices (HVI), selected most responsive wavelengths and prediction models. Univariate and multivariate analyses were applied to predict and interpret the relationships between variables and to discriminate among treatments.

Table 1. Narrowband vegetation indices calculated from the leaf spectral reflectance.

Vegetation Index	Formula	Reference
NDVI	$NDVI = \frac{R_{NIR} - R_{Red}}{R_{NIR} + R_{Red}}$	[40]
GNDVI	$GNDVI = \frac{R_{NIR} - R_{Green}}{R_{NIR} + R_{Green}}$	[41]
EVI	$EVI = 2.5 \times \frac{R_{NIR} - R_{Red}}{R_{NIR} + 6 R_{Red} - 7.5 R_{Blue} + 1}$	[42]
SAVI	$SAVI = (1 + L) \frac{R_{NIR} - R_{Red}}{R_{NIR} + R_{Red} + L} \quad (L = 0.5)$	[42]
OSAVI	$OSAVI = \frac{R_{NIR} - R_{Red}}{R_{NIR} + R_{Red} + 0.16}$	[42]
MSAVI2	$MSAVI2 = \frac{2 R_{NIR} + 1 - \sqrt{(2 R_{NIR} + 1)^2 - 8 (R_{NIR} - R_{Red})}}{2}$	[43]
SIPI	$SIPI = \frac{R_{800} - R_{445}}{R_{800} - R_{680}}$	[44]
PSSRc	$PSSRc = \frac{R_{800}}{R_{470}}$	[44,45]
RARS	$RARS = \frac{R_{675}}{R_{700}}$	[44]
WBI	$WBI = \frac{R_{900}}{R_{970}}$	[46]
MSI	$MSI = \frac{R_{1600}}{R_{820}}$	[47]
NDII	$NDII = \frac{R_{819} - R_{1649}}{R_{819} + R_{1649}}$	[46]
NMDI	$NMDI = \frac{R_{860} - (R_{1640} - R_{2130})}{R_{860} + (R_{1640} - R_{2130})}$	[48]

Table 1. Cont.

Vegetation Index	Formula	Reference
NDDI	$NDDI = \frac{NDVI - NDWI}{NDVI + NDWI}$	[49]
NMDI	$NMDI = \frac{R_{860} - (R_{1640} - R_{2130})}{R_{860} + (R_{1640} - R_{2130})}$	[50]
NDWI ₁₆₄₀	$NDWI_{1640} = \frac{R_{858} - R_{1640}}{R_{858} + R_{1640}}$	[50]
NDWI ₂₁₃₀	$NDWI_{2130} = \frac{R_{858} - R_{2130}}{R_{858} + R_{2130}}$	[50]
ARI1	$ARI1 = \frac{1}{R_{550}} - \frac{1}{R_{700}}$	[51]
ARI2	$ARI2 = R_{800} \times \left(\frac{1}{R_{550}} - \frac{1}{R_{700}} \right)$	[51]
CRI1	$CRI1 = \frac{1}{R_{510}} - \frac{1}{R_{550}}$	[50]
CRI2	$CRI2 = \frac{1}{R_{510}} - \frac{1}{R_{700}}$	[50]
VOG1	$VOG1 = \frac{R_{740}}{R_{720}}$	[52]
VOG2	$VOG2 = \frac{R_{734} - R_{747}}{R_{715} + R_{726}}$	[52]
NPQI	$NPQI = \frac{R_{415} - R_{435}}{R_{415} + R_{435}}$	[53]
PRI	$PRI = \frac{R_{531} - R_{570}}{R_{531} + R_{570}}$	[54]

3. Results

3.1. Photosynthetic and Protective Pigments, Stress Markers, and Leaf Biochemical Parameters

The mean concentrations of photosynthetic pigments expressed per unit area were 391.51 mg m⁻² for chlorophyll a, 187.01 mg m⁻² for chlorophyll b, 578.52 mg m⁻² for total chlorophyll (a + b), and 69.10 mg m⁻² for carotenoids. The coefficients of variation ranged from 42.29% to 67.05%. When expressed per unit mass, chlorophyll a had a mean of 21.18 mg g⁻¹, chlorophyll b 9.49 mg g⁻¹, total chlorophyll (a + b) 30.66 mg g⁻¹, and carotenoids 3.86 mg g⁻¹, with coefficients of variation ranging from 28.35% to 55.00%.

For protective compounds, the mean values observed were 42.26 mg g⁻¹ for flavonoids (mass basis), 67.91 nmol cm⁻² for flavonoids (area basis), 23.38 μmol g⁻¹ for proline, and 135.88 mL cm⁻² for phenolic compounds. The coefficients of variation for these compounds ranged from 22.02% to 44.59%.

The analysis of the stress markers revealed a mean lignin concentration of 27.53 mg g⁻¹, cellulose content of 103.92 nmol mg⁻¹, antioxidant activity (RSA) of 64.75%, electrolyte leakage (ELK) of 39.70%, and relative water content (RWC) of 70.46%. The coefficients of variation for these variables varied between 16.54% and 37.98%. The distribution of values (minimum, median, and maximum) for each parameter is detailed in Table 2.

3.2. Spectral Reflectance Profiles Under Water Regimes

The mean reflectance spectra (350–2500 nm) of fully expanded *Glycine max* leaves exhibited distinct patterns across the eleven water regime treatments (W100 to W0). In the VIS region (350–700 nm), all the treatments presented low reflectance values, with a clear separation between the regimes: leaves under relatively high water availability (W100, W90, W80) consistently presented relatively low reflectance, whereas those under relatively severe water restriction (notably W30, W20, W10, and especially W0) presented a progressive increase in reflectance, particularly near the red edge (approximately 700 nm).

In the NIR region (700–1350 nm), all the treatments demonstrated a pronounced increase in reflectance. Compared with those under greater water deficit, the treatments with greater water availability (W100 to W60) maintained higher NIR reflectance values, with W0 consistently presenting the lowest NIR reflectance values.

Table 2. Descriptive statistics for physiological and biochemical parameters assessed in leaf samples. The table presents the mean, median, minimum, maximum, and coefficient of variation (CV, %) for each parameter measured across 264 samples. Photosynthetic pigments are expressed both per area (mg m^{-2} for chlorophyll a, chlorophyll b, total chlorophyll (a + b), and carotenoids) and per dry mass (mg g^{-1}). Protective compounds include flavonoids (mg g^{-1} and nmol cm^{-2}), proline ($\mu\text{mol g}^{-1}$), and phenolic compounds (mL cm^{-2}). The stress markers included lignin (mg g^{-1}), cellulose (nmol mg^{-1}), radical scavenging activity (RSA, %), electrolyte leakage (ELK, %), and relative water content (RWC, %). The CV indicates the degree of variability for each parameter within the sample set.

Physiological Groups	Parameters	Count (n)	Mean	Median	Min	Max	CV (%)
Photosynthetic pigments (area)	Chl a (mg m^{-2})	264	391.51	428.75	75.72	680.70	42.29
	Chl b (mg m^{-2})	264	187.01	187.92	3.77	484.47	67.05
	Chl a + b (mg m^{-2})	264	578.52	662.99	84.90	1145.01	47.20
	Car (mg m^{-2})	264	69.10	60.53	13.91	180.98	55.78
Photosynthetic pigments (mass)	Chl a (mg g^{-1})	264	21.18	21.21	5.45	36.48	28.35
	Chl b (mg g^{-1})	264	9.49	10.11	0.29	22.96	55.00
	Chl a + b (mg g^{-1})	264	30.66	33.06	6.11	58.18	31.49
	Car (mg g^{-1})	264	3.86	3.56	0.60	9.89	48.95
Protective compounds	Flv (mg g^{-1})	264	42.26	36.54	15.14	105.20	44.59
	Flv (nmol cm^{-2})	264	67.91	67.55	37.43	109.67	22.02
	Pro ($\mu\text{mol g}^{-1}$)	264	23.38	24.16	4.93	43.32	37.98
	Phe (mL cm^{-2})	264	135.88	130.45	68.71	238.30	27.93
Stress markers	Lig (mg g^{-1})	264	27.53	27.11	10.55	49.20	26.63
	Cel (nmol mg^{-1})	264	103.92	103.98	64.17	144.28	16.54
	RSA (%)	264	64.75	66.38	37.62	83.48	17.39
	ELK (%)	264	39.70	41.55	21.82	53.12	20.62
	RWC (%)	264	70.46	68.69	38.20	108.52	21.68

In the SWIR region (1350–2500 nm), two marked absorption features were observed (SWIR1 and SWIR2), with all the treatments resulting in a reduction in reflectance at these wavelengths. The separation among water regimes was most apparent in this domain: the W0 treatment presented the highest reflectance values across SWIR1 and SWIR2, whereas intermediate treatments followed a gradation corresponding to the severity of water deficit but did not progressively change.

Statistically significant differences in reflectance were detected among all the treatments throughout the entire spectral range ($F = 26.97$, $p < 0.001$; $n = 24$), as presented in Figure 2.

3.3. Principal Component Analysis of Leaf Reflectance

Principal component analysis (PCA) of the leaf reflectance spectra of *Glycine max* under different water regimes revealed clear separation among the treatments. The first two principal components explained 69.8% (PC1) and 18.2% (PC2) of the total variance, respectively. In the biplot (Figure 3A), three main clusters were identified, each corresponding to specific groupings of the water regime treatments.

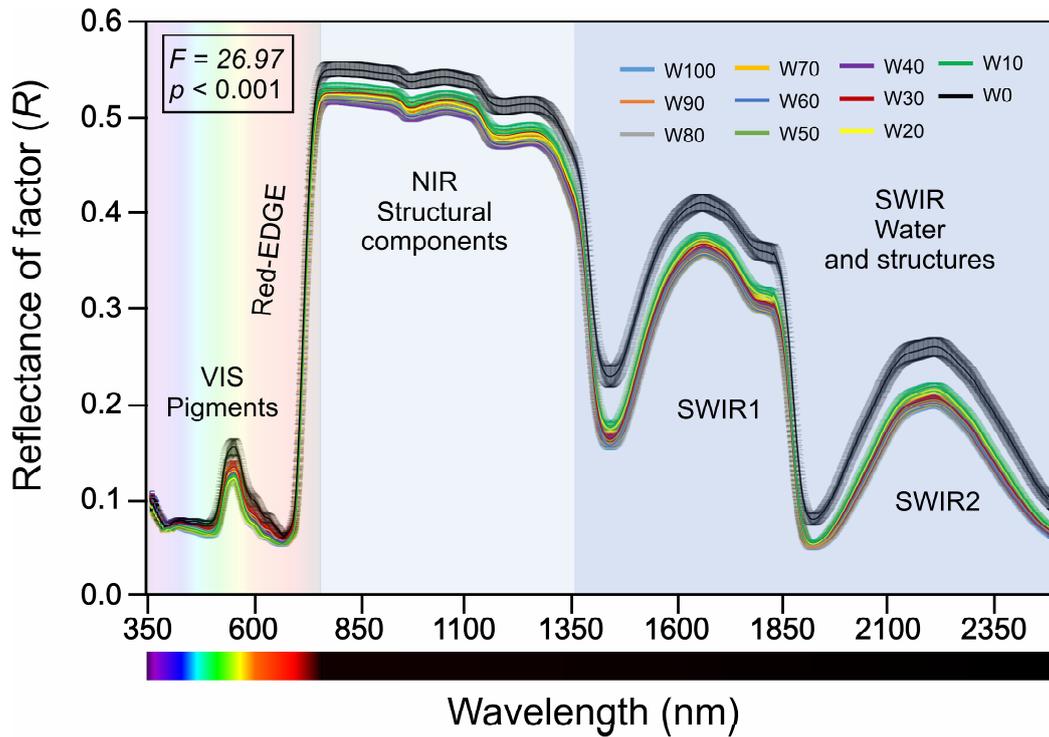


Figure 2. UV–VIS–NIR–SWIR reflectance profiles (350–2500 nm) of fully expanded *Glycine max* leaves under distinct water regimes. Reflectance spectra are shown for all the treatments, ranging from W100 to W0. The spectral domains are segmented as follows: VIS (350–700 nm; pigment absorption), NIR (700–1350 nm; structural leaf properties), and SWIR (1350–2500 nm; water and additional structural components). Significant differences between treatments were detected via one-way ANOVA ($F = 26.97, p < 0.001$). (Mean \pm SE). ($n = 24$).

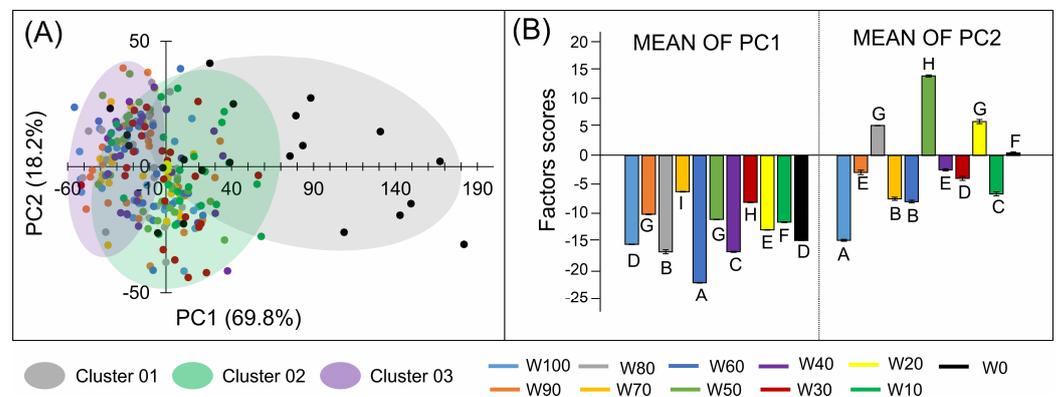


Figure 3. Principal component analysis (PCA) of leaf reflectance data from *Glycine max* under contrasting water regimes. **(A)** Biplot of the first two principal components (PC1 and PC2), explaining 69.8% and 18.2% of the total variance, respectively. Individual samples are coloured according to dot treatments, and ellipses highlight the main clusters identified by the analysis. **(B)** Mean factor scores (\pm SE) for each treatment along PC1 and PC2. Different uppercase letters indicate significant differences among treatments ($p < 0.05$). The treatments ranged from well-watered (W100) to severe water deficit (W0) ($n = 24$). The colours for each treatment are represented in the legend with the corresponding line styles.

The distribution of individual samples within the PCA space revealed that well-watered treatments (W100, W90, W80) were predominantly grouped within Cluster 1. Intermediate water regimes (W70, W60, W50, W40, W30, W20) were associated mainly with Cluster 2, whereas the most severe deficit treatments (W10, W0) were clearly separated into Cluster 03.

The mean factor scores for each treatment along PC1 and PC2 (Figure 3B) further supported these groupings. W100 and W80 presented the most negative mean values on PC1, whereas W0 and W10 presented the highest positive scores.

Principal component analysis of the hyperspectral reflectance data revealed that the first three principal components accounted for the majority of the spectral variability in the soybean leaves. PC1, PC2, and PC3 explained 69.8%, 18.2%, and 5.2% of the total variance, respectively, with the cumulative variance reaching over 93% by the third component (Figure 4A).

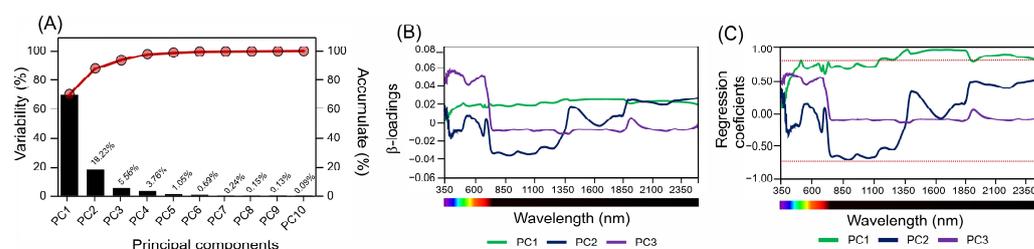


Figure 4. Principal component analysis (PCA) of hyperspectral reflectance data from soybean leaves. (A) Percentage of variance explained by the first ten principal components, with the cumulative variance indicated by the red line. (B) Standardized β -loadings for the first three principal components (PC1, PC2, PC3) as a function of wavelength, highlighting the main spectral regions contributing to variability. (C) Regression coefficients for the first three principal components across the spectral range (350–2500 nm), showing the wavelengths most strongly associated with each PC. The red dotted lines represent the limits of -0.75 and $+0.75$ limited in regression coefficients.

The standardized β -loadings for PC1, PC2, and PC3 (Figure 4B) revealed that the major spectral regions contributing to variability were distributed across the VIS, NIR, and SWIR domains. PC1 was associated primarily with broad features spanning the entire spectrum, whereas PC2 and PC3 exhibited more distinct peaks and troughs, especially in regions at approximately 600 nm, 1400 nm, and 1900 nm.

The regression coefficients for the first three principal components (Figure 4C) highlighted the wavelengths most strongly linked to each PC. Notably, the PC1 coefficients remained high across the VIS and NIR regions, whereas the PC2 and PC3 coefficients varied more markedly in the SWIR region, identifying specific wavelengths that contributed to the separation of water regimes in the PCA. All analyses were conducted using reflectance data in the 350–2500 nm range.

3.4. Variable Importance of Vegetation Indices for Leaf Trait Prediction

The evaluation of 25 spectral vegetation indices revealed substantial variation in their relative importance for predicting physiological and biochemical responses in soybean leaves across water regimes. Among all indices, CRI2 exhibited the highest variable importance, reaching approximately 35%. ARI2 and ARI1 followed, with relative importance values of approximately 16% and 12%, respectively. Other indices, such as CRI1 (8%), VOG1 (5%), and PSSRc (2%), also contributed to the predictive modelling but with a lower impact.

Conversely, the classic indices related to greenness, moisture, and general vegetation status (e.g., the NDVI, GNDVI, EVI, SAVI, OSAVI, MSI, NDMI, and PRI) all presented relative importance values below 2%. The majority of indices were below this threshold, indicating limited relevance for predicting physiological and biochemical variation under the tested water regimes (Figure 5).

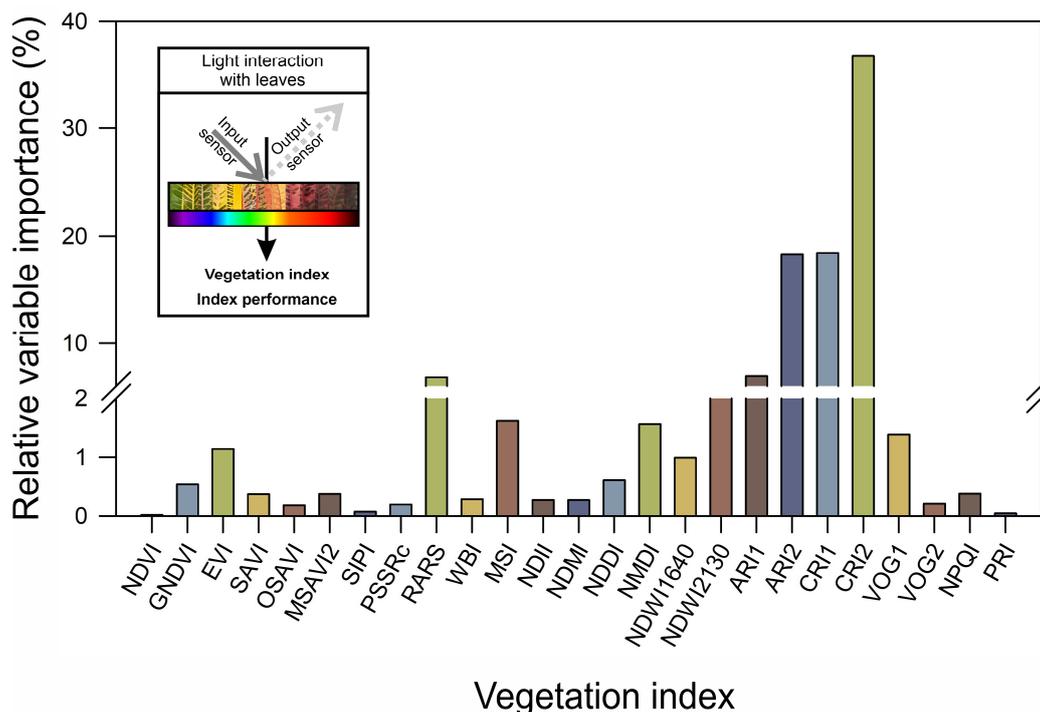


Figure 5. Relative importance of spectral vegetation indices for predicting physiological and biochemical responses in soybean leaves under different water regimes. The bar chart displays the relative variable importance (%) of 25 vegetation indices derived from hyperspectral data. The inset illustrates the principle of light interaction with leaves and the vegetation index calculation. Index abbreviations: NDVI, normalized difference vegetation index; GNDVI, green normalized difference vegetation index; EVI, enhanced vegetation index; SAVI, soil-adjusted vegetation index; OSAVI, optimized soil-adjusted vegetation index; MSAVI2, modified soil-adjusted vegetation index 2; SIPI, structure insensitive pigment index; PSSRc, pigment specific simple ratio—carotenoids; RARS, red-edge anthocyanin reflectance signal; WBI, water band index; MSI, moisture stress index; NDII, normalized difference infrared index; NDMI, normalized difference moisture index; NDDI, normalized difference drought index; NMDI, normalized multi-band drought index; NDWI1640, normalized difference water index with 1640 nm band; NDWI2130, normalized difference water index with 2130 nm; ARI1, anthocyanin reflectance 1; ARI2, anthocyanin reflectance 2; CRI1, carotenoid reflectance 1; CRI2, carotenoid reflectance 2; VOG1, Vogelmann red edge 1; VOG2, Vogelmann red edge 2; NPQI, normalized phaeophytinization index; PRI, photochemical reflectance index, shows greater contribution to the predictive modelling of leaf responses.

These results demonstrate that carotenoid and anthocyanin indices are the most informative spectral metrics for the prediction of leaf responses to water deficit in soybean (Figure 5).

3.5. Hierarchical Clustering of Spectral Profiles Under Water Regimes

Hierarchical clustering analysis based on spectral data revealed clear separation among the *Glycine max* treatments according to water availability. The Euclidean distance matrix (Figure 6A) revealed low dissimilarity between the well-watered treatments (W100, W90, W80, W70, W60), with pairwise distances ranging from 0.01–0.34. The moderate water deficit treatments (W50, W40, W30) also clustered closely together (distances < 0.5) but presented increasing divergence from both the well-watered and extreme deficit groups.

The most pronounced spectral dissimilarity was observed between W0 and all the other treatments, with Euclidean distances ranging from 1.4–2.0, and between W0 and W10, with a distance of 0.66. The dendrogram (Figure 6B) highlighted the formation of distinct clusters, with W0 forming a separate branch and W10 and W20 grouping together, clearly

separated from the well-watered (W100–W60) and moderate-deficit (W50–W30) treatments. This structure reflects the spectral divergence resulting from the gradient of water restriction, with severe deficit and extreme restriction regimes forming distinct clusters.

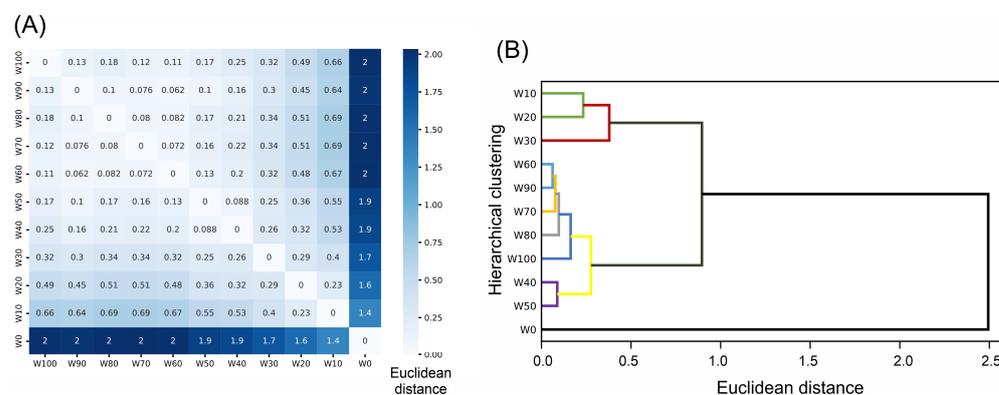


Figure 6. Hierarchical clustering of *Glycine max* treatments under different water regimes on the basis of spectral data. (A) Euclidean distance matrix showing pairwise distances among all water regimes (W100 to W0), with the intensity of blue shading indicating increasing dissimilarity. (B) Dendrogram from hierarchical clustering analysis, highlighting the formation of distinct groups among treatments. The separation reflects the spectral divergence imposed by the water availability gradient, with severe deficit (W0) and extreme restriction (W10, W20) forming distinct clusters from the well-watered and moderate-deficit groups according to the Euclidean distance. The colours in the hierarchical clustering indicate the proximity among similar treatments.

The confusion matrices in Figure 7 summarize the classification results of the eleven water regimes via eight machine learning models on the basis of hyperspectral reflectance data. The random forest, decision tree, gradient boosting, and MLP classifiers produced perfect classifications, with all samples correctly assigned to their original classes (values of 1.00 along the diagonal and zeros elsewhere). Logistic regression also achieved high accuracy, with a few misclassifications, mostly between intermediate treatments such as W80, W70, and W60.

In contrast, SVM, KNN, and naive Bayes showed increased rates of confusion, particularly among adjacent or similar water regimes. SVM and KNN presented off-diagonal values up to 0.22, indicating some overlap in prediction between regimes such as W10, W20, and W0. Naive Bayes resulted in the highest number of misclassifications, with proportions up to 0.22 for certain deficit treatments, reflecting reduced discriminatory power in more challenging scenarios.

Overall, tree-based ensemble models and neural network approaches demonstrated the highest predictive precision for distinguishing water regimes in soybean leaves, whereas distance- and probability-based models struggled to separate spectrally similar classes.

3.6. Correlation Between Spectral Data and Leaf Biochemical/Biophysical Traits

The Pearson correlation heatmap (Figure 8A) demonstrated distinct association patterns between hyperspectral reflectance (350–2500 nm) and key biochemical and biophysical leaf traits in *Glycine max*. Photosynthetic pigments, both per area and mass (Chl a, Chl b, Chl a + b, Car), exhibited strong positive correlations with reflectance in the near-infrared region (NIR; >750 nm), with r values approaching +1. Conversely, negative correlations were observed in the visible region (VIS; 350–700 nm), particularly between 600–700 nm (r values near -1). Flavonoids, proline, and total phenolics displayed the opposite trend, being positively correlated with VIS reflectance and negatively correlated with the NIR bands.

Lignin, cellulose, and some stress markers (RSA, ELK, RWC) were generally negatively correlated with reflectance across much of the spectrum. The most pronounced negative associations were found in the shortwave infrared (SWIR; >1400 nm) band, where the r

values reached below -0.5 , especially for RSA, ELK, and RWC, indicating a strong inverse relationship between reflectance and these parameters under varying water regimes.

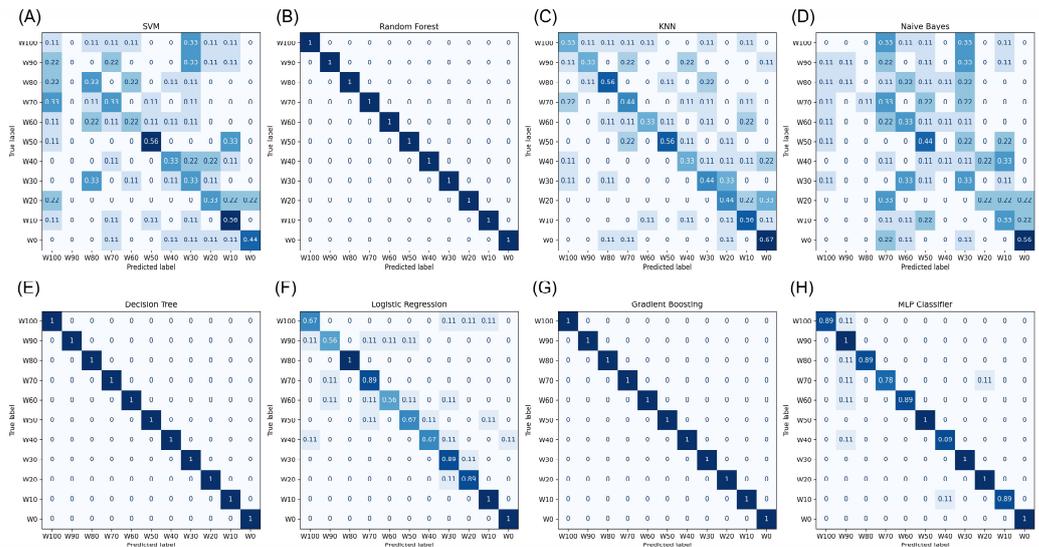


Figure 7. Confusion matrices for eight machine learning classifiers for predicting water regimes in soybean plants on the basis of hyperspectral reflectance data. The classification performance of the following models is shown: (A) support vector machine (SVM), (B) random forest, (C) k-nearest neighbors (KNN), (D) naive Bayes, (E) decision tree, (F) logistic regression, (G) gradient boosting, and (H) multilayer perceptron (MLP classifier). The predicted classes (x-axis) and true classes (y-axis) correspond to the eleven water regimes (W100 to W0). Each cell shows the proportion of correctly or incorrectly classified samples per class in increasing light to dark blue. The values in the boxes indicate correct classifications, whereas the off-diagonal entries denote misclassifications.

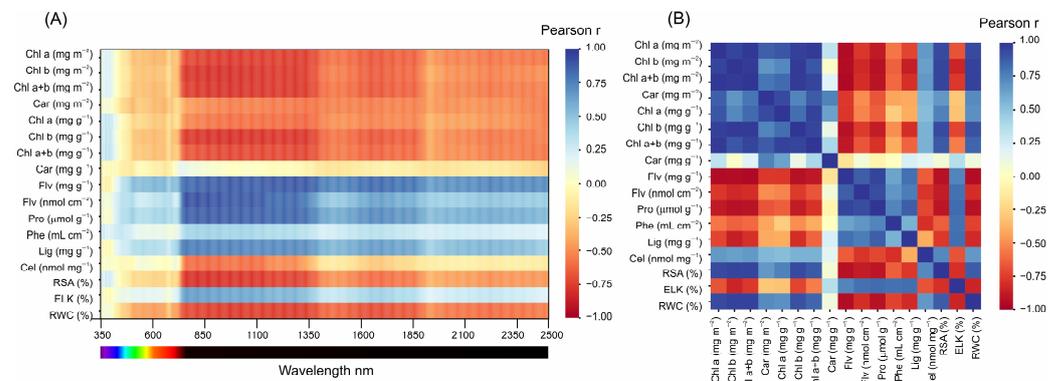


Figure 8. Pearson correlation heatmaps between leaf spectral and biochemical/biophysical traits in *Glycine max*. (A) Correlation matrix showing the Pearson correlation coefficient (r) between hyperspectral reflectance (350–2500 nm) and selected biochemical and biophysical variables: chlorophyll a (Chl a), chlorophyll b (Chl b), total chlorophyll (Chl a + b), carotenoids (Car), flavonoids (Flv), proline (Pro), total phenolics (Phe), lignin (Lig), cellulose (Cel), radical scavenging activity (RSA), electrolyte leakage (ELK), and relative water content (RWC), all expressed per area or mass as indicated. (B) Pearson correlation matrix among the same leaf traits. The color scale ranges from -1 (strong negative correlation, blue) to $+1$ (strong positive correlation, red), with yellow indicating near-zero correlations.

The pairwise trait correlation matrix (Figure 8B) revealed clear groupings. Photosynthetic pigments strongly positively correlated with each other ($r > 0.8$). Flavonoids, proline, and phenolics were also positively correlated ($r > 0.5$) but strongly negatively correlated with pigments ($r < -0.5$). Lignin and cellulose were positively correlated ($r \approx 0.6$) and

negatively associated with RWC ($r \approx -0.6$). RSA and ELK were positively correlated, but both were negatively related to RWC.

These results highlight that the VIS and NIR spectral bands are decisive for distinguishing major biochemical components, whereas the SWIR bands are more strongly associated with stress-related traits that align structures and cell compounds. This highlights the value of full-spectrum reflectance profiling for the functional and physiological characterization of soybean leaves under contrasting water availability conditions.

3.7. Wavelength Selection for the Prediction and Classification of Leaf Traits

The selection of informative wavelengths by multiple multivariate algorithms revealed both common and variable-specific patterns across the 350–2500 nm range for all foliar traits in *Glycine max* (Figure 9). For all the parameters evaluated, the majority of the algorithms consistently identified wavelengths within the visible (VIS, 350–700 nm) and near-infrared (NIR, 700–1350 nm) regions as the most relevant for predictive modelling.

For pigments expressed per area (Chl a, Chl b, Chl a + b, Car; Figure 9A–D), the most frequently selected wavelengths were concentrated between 550 and 750 nm, spanning the red-edge transition. Pigments expressed per mass (Figure 9E–H) similarly presented selection peaks at the VIS–NIR boundary, but several algorithms also highlighted informative bands in the SWIR region (1400–1900 nm).

For secondary metabolites (Flv, Pro, Phe; Figure 9I–L) and cell wall components (Lig, Cel; Figure 9M,N), the algorithms predominantly selected wavelengths in both the VIS and SWIR regions. Notably, the SWIR region (between 1450–1900 nm and >2100 nm) was highly represented in the selection for phenolics, lignin, and cellulose.

For the stress and physiological indices (RSA, ELK, RWC; Figure 9O–Q), informative wavelengths were distributed across the spectrum, with mutual importance placed on the SWIR region (particularly around water absorption features at 1450 nm, 1940 nm, and 2200 nm). The random forest, Boruta, and genetic algorithm approaches often select broader ranges, whereas the LDA and VIP methods consistently highlight more discrete bands.

The repeated selection of red-edge (between 700–740 nm) and SWIR bands (especially 1450, 1940, and 2200 nm) by distinct algorithms reinforces their mechanistic link to pigment absorption and water status, providing strong candidates for compact multispectral sensor design and trait-specific remote phenotyping (Figure 9A–Q).

3.8. Predictive Modelling Using Hyperspectral Reflectance

The partial least squares regression (PLSR) models demonstrated robust predictive performance for multiple foliar traits in *Glycine max* on the basis of the hyperspectral reflectance data (Figures 10 and 11).

For the prediction of foliar pigment concentrations (Figure 10), high coefficients of determination (R^2) were observed for both the training and test sets. The models for chlorophyll a, chlorophyll b, total chlorophyll, and carotenoids (base area) achieved R^2 values ranging from 0.74–0.88, with low bias and a root mean square error (RMSE) typically less than 100. Pigments expressed per mass also showed satisfactory predictive capacity, with R^2 values between 0.44 and 0.67 for chlorophyll a, chlorophyll b, total chlorophyll, and carotenoids. The regression slopes for all pigment models were close to unity, and the MAE values remained low across the range of observed concentrations.

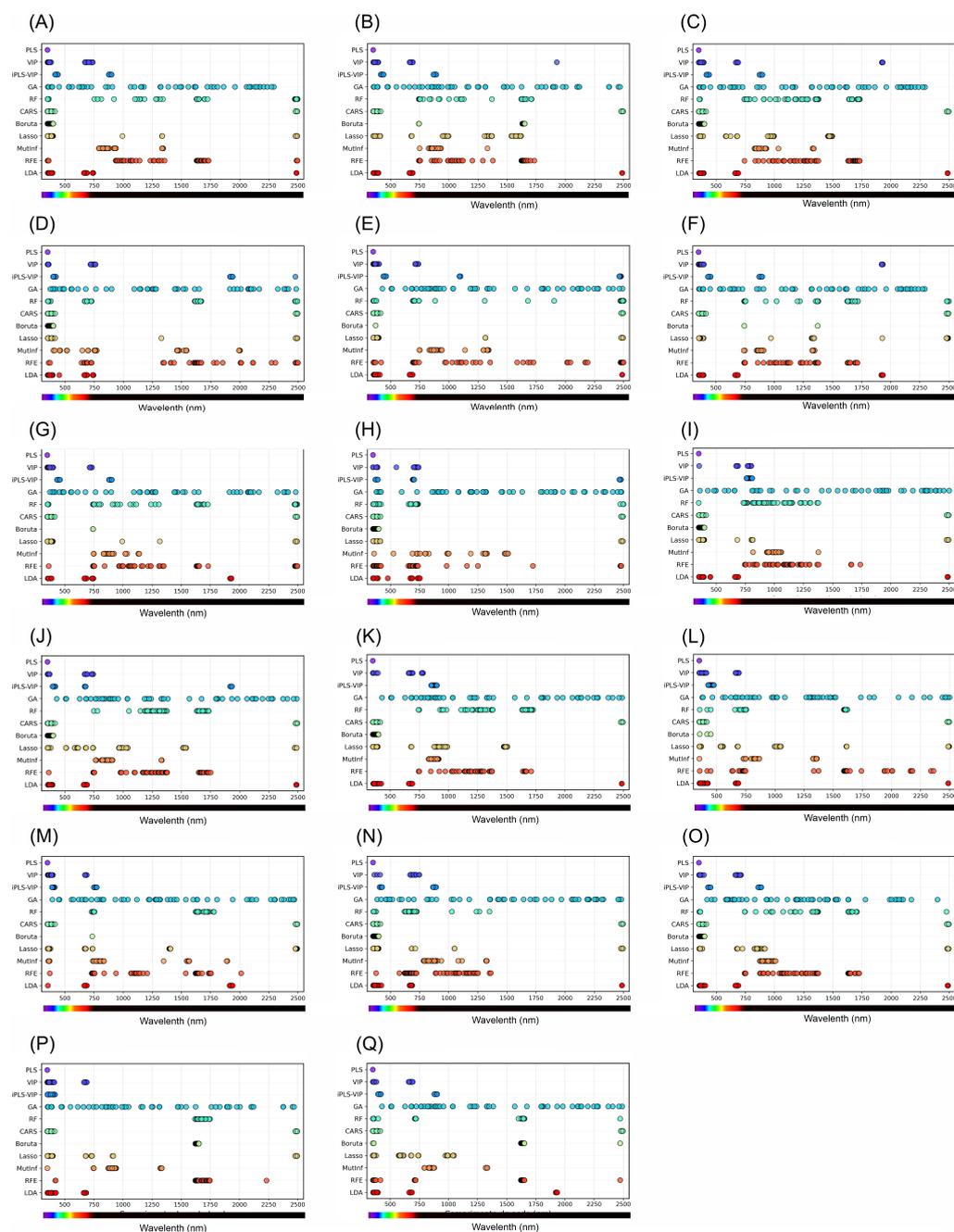


Figure 9. Wavelengths selected by different algorithms for the prediction and classification of foliar traits in *Glycine max*. (A–Q) represent the most informative wavelengths identified for each variable (in order): (A) chlorophyll a (Chl a, mg m^{-2}), (B) chlorophyll b (Chl b, mg m^{-2}), (C) total chlorophyll (Chl a + b, mg m^{-2}), (D) carotenoids (Car, mg m^{-2}), (E) chlorophyll a (mg g^{-1}), (F) chlorophyll b (mg g^{-1}), (G) total chlorophyll (mg g^{-1}), (H) carotenoids (mg g^{-1}), (I) flavonoids (Flv, mg g^{-1}), (J) flavonoids (nmol cm^{-2}), (K) proline (Pro, $\mu\text{mol g}^{-1}$), (L) phenolics (Phe, mL cm^{-2}), (M) lignin (Lig, mg g^{-1}), (N) cellulose (Cel, nmol mg^{-1}), (O) radical scavenging activity (RSA, %), (P) electrolyte leakage (ELK, %), (Q) relative water content (RWC, %). Each dot color indicates a selected wavelength for the respective algorithm. The y-axis lists the algorithms used by each color: PLS (Partial Least Squares), VIP (Variable Importance in Projection), iPLS-VIP (Interval Partial Least Squares), GA (Genetic Algorithm), RF (Random Forest), CARS (Competitive Adaptive Repeated Sampling), Boruta, Lasso, MutInf (Mutual Information), RFE (Recursive Feature Elimination), LDA (Linear Discriminant Analysis).

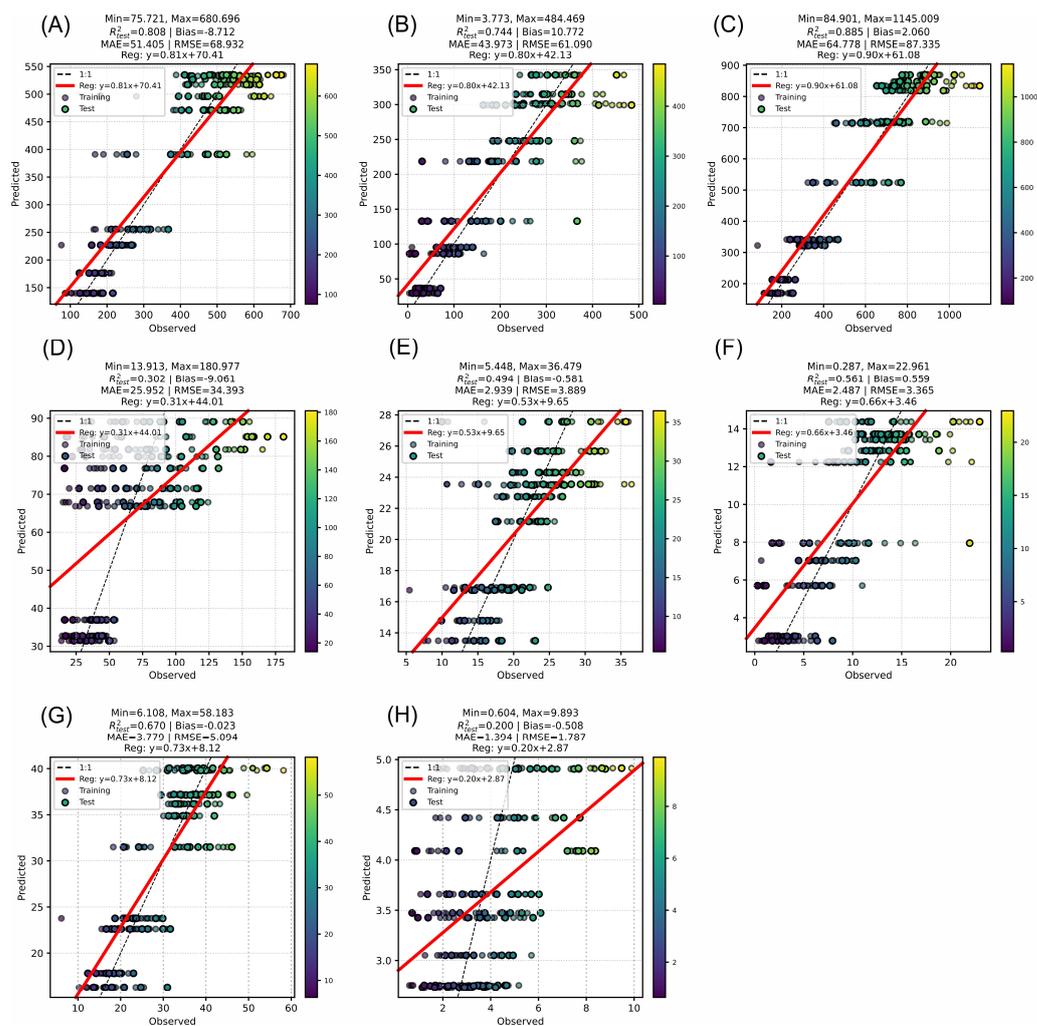


Figure 10. Performance of PLSR models for predicting foliar pigment concentrations in *Glycine max* using hyperspectral reflectance data. Scatter plots of observed versus predicted values for (A) chlorophyll a (Chl a, mg m^{-2}), (B) chlorophyll b (Chl b, mg m^{-2}), (C) total chlorophyll (Chl a + b, mg m^{-2}), (D) carotenoids (Car, mg m^{-2}), (E) chlorophyll a (mg g^{-1}), (F) chlorophyll b (mg g^{-1}), (G) total chlorophyll (mg g^{-1}), and (H) carotenoids (mg g^{-1}). The training and test sets are indicated, with the red line representing the 1:1 relationship. Each image shows the regression equation, coefficient of determination (R^2), bias, mean absolute error (MAE), and root mean square error (RMSE) for both sets. The color scale denotes the sample density within each observed range.

For the biochemical and physiological parameters (Figure 11), the PLSR models yielded moderate to high R^2 values for most variables. The prediction of flavonoids (mg g^{-1}) reached an R^2 of 0.72, with a low bias and RMSE of 4.52. Flavonoids (nmol cm^{-2}) and proline ($\mu\text{mol g}^{-1}$) achieved R^2 values of 0.36 and 0.15, respectively, reflecting a lower predictive strength for these variables. The models for phenolics, lignin, and cellulose presented R^2 values between 0.31 and 0.96, with the best performance obtained for cellulose ($R^2 = 0.96$, $\text{RMSE} = 4.83$). The radical scavenging activity (RSA) was also well predicted, with $R^2 = 0.81$ and $\text{RMSE} = 3.01$.

For all the traits, scatter plots of the observed versus predicted values indicated close clustering along the 1:1 line in the best-performing models, particularly for pigments and cellulose. The color density scale confirmed good model generalizability across the observed trait ranges in both the training and test sets.

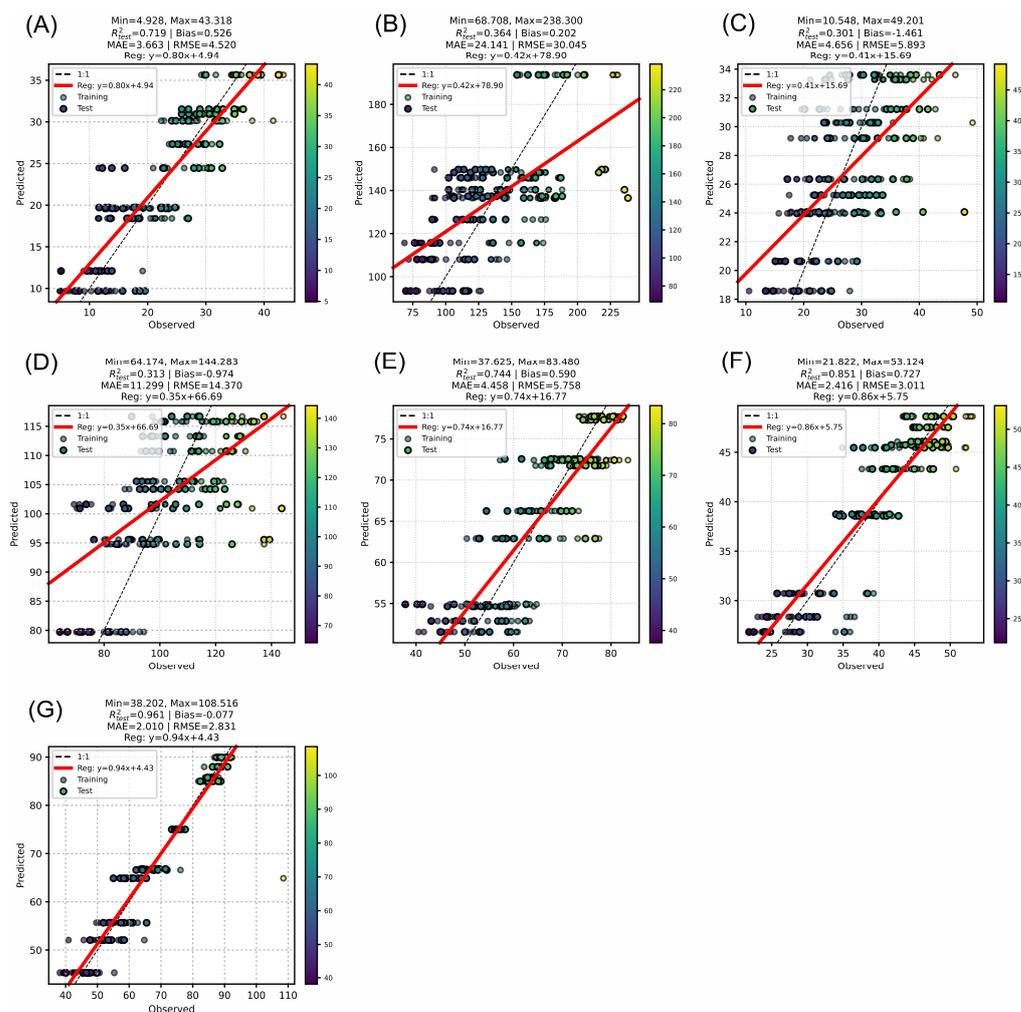


Figure 11. Predictive performance of PLSR models for biochemical and physiological traits in *Glycine max* leaves using hyperspectral reflectance data. Scatter plots of observed versus predicted values for (A) flavonoids (Flv, mg g^{-1}), (B) flavonoids (nmol cm^{-2}), (C) proline (Pro, $\mu\text{mol g}^{-1}$), (D) total phenolics (Phe, mL cm^{-2}), (E) lignin (Lig, mg g^{-1}), (F) cellulose (Cel, nmol mg^{-1}), and (G) radical scavenging activity (RSA, %). The training and test sets are indicated, with the red line representing the 1:1 relationship. For each trait, the regression equation, coefficient of determination (R^2), bias, mean absolute error (MAE), and root mean square error (RMSE) are reported. The color bar represents the sample density across the observed range.

3.9. Identification of the Most Responsive Wavelength Pairs via Spectral Correlation Analysis

The spectral correlation (R^2) maps generated via the hyperspectral vegetation index (HVI) method revealed the wavelength combinations most strongly associated with the prediction of each foliar biochemical and physiological trait in *Glycine max* (Figure 12).

Across all the traits, the majority of the wavelength pairs produced low R^2 values (dark blue), indicating weak or negligible predictive power for most combinations within the 350–2500 nm range. However, each map displays discrete clusters of yellow to red points, representing pairs of wavelengths with notably high predictive capacity (R^2 approaching or exceeding 0.75).

For chlorophyll-related traits (Figure 12A–C,E–G), the highest R^2 values were observed in combinations involving the red-edge (approximately 700–750 nm) and NIR bands (750–900 nm), often in association with SWIR wavelengths near 1400 nm or 1900 nm. These hotspots indicate regions where the synergy between pigment absorption features and water/structural absorption results in optimal sensitivity for model prediction.

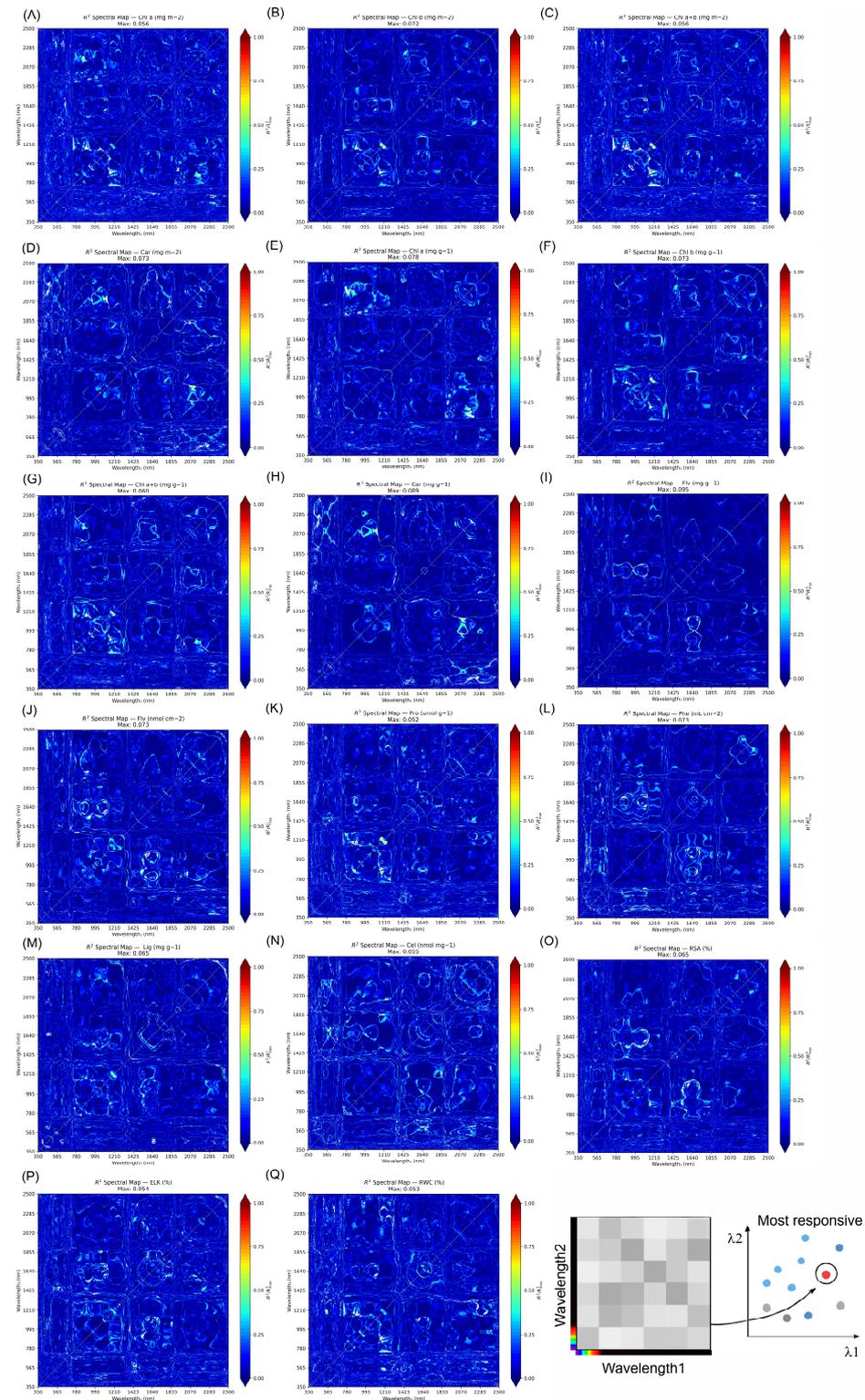


Figure 12. Spectral correlation (R^2) maps for all pairwise wavelength combinations via the hyperspectral vegetation index (HVI) method for the prediction of foliar biochemical and physiological traits in *Glycine max*. (A–Q) R^2 maps for (A) chlorophyll a (Chl a, mg m^{-2}), (B) chlorophyll b (Chl b, mg m^{-2}), (C) total chlorophyll (Chl a + b, mg m^{-2}), (D) carotenoids (Car, mg m^{-2}), (E) chlorophyll a (mg g^{-1}), (F) chlorophyll b (mg g^{-1}), (G) total chlorophyll (mg g^{-1}), (H) carotenoids (mg g^{-1}), (I) flavonoids (Flv, mg g^{-1}), (J) flavonoids (nmol cm^{-2}), (K) proline (Pro, $\mu\text{mol g}^{-1}$), (L) phenolics (Phe, mL cm^{-2}), (M) lignin (Lig, mg g^{-1}), (N) cellulose (Cel, nmol mg^{-1}), (O) radical scavenging activity (RSA, %), (P) electrolyte leakage (ELK, %), (Q) relative water content (RWC, %). The bottom right schematic illustrates the approach used to identify the most responsive wavelength pairs via the HVI algorithm (red points).

Carotenoids (Figure 12D,H) and secondary metabolites such as flavonoids, proline, and phenolics (Figure 12I–L) also presented the most responsive points between the VIS–NIR boundary and specific SWIR intervals (especially at approximately 1450, 1940, and 2200 nm), with the highest R^2 clusters scattered but generally located along these axes.

For cell wall components (lignin and cellulose; Figure 12M,N) and physiological stress markers (RSA, ELK, RWC; Figure 12O–Q), the strongest correlations were detected between pairs of SWIR wavelengths. Particularly those spanning the water absorption regions near 1450, 1940, and 2200 nm.

These responsive points, where R^2 approaches 1.0, represent the wavelength pairs with the greatest sensitivity for the nondestructive prediction of each respective trait. The schematic in the lower right of Figure 12 illustrates the HVI approach, highlighting how only a minority of wavelength combinations yield maximal prediction power, reflected as isolated yellow/red points within the correlation landscape (Figure 12).

In summary HVI, shows the R^2 correlation maps not only confirm the known associations between certain spectral regions and foliar biochemistry but also visually reveal the unique, trait-specific wavelength pairs driving predictive performance. This approach highlights the potential of hyperspectral analysis not just for nondestructive trait estimation, but also for advancing our understanding of the optical signatures underlying plant physiological status. Such findings pave the way for the development of optimized indices and the rational design of next-generation remote sensing tools tailored to key functional traits in soybean and other crops.

4. Discussion

4.1. Overview of Key Findings

This study provides compelling evidence of the significant impacts of various water regimes on soybean (*Glycine max*) leaf physiology, biochemical composition, and hyperspectral reflectance. As anticipated, water stress led to a progressive decrease in photosynthetic pigments, specifically chlorophyll (Chl a and Chl b) and carotenoids, and an increase in protective compounds such as flavonoids (Flv) and proline (Pro), which are commonly associated with stress tolerance mechanisms [8,55]. Furthermore, spectral reflectance data revealed distinct patterns associated with different water treatments, with notable shifts in the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) regions, reflecting physiological changes in response to water availability [8,30,56].

Recent studies corroborate our findings, highlighting that drought stress significantly reduces the contents of photosynthetic pigments such as chlorophyll and carotenoids in soybean plants. For example, a study by Nehra et al., (2025) [57] reported a decrease in chlorophyll content under drought stress, attributed to increased reactive oxygen species (ROS) production leading to oxidative damage and chlorophyll degradation. Similarly, carotenoids, which play crucial roles in photoprotection and antioxidant defense, are also affected by water stress. Their reduction under stress limits nonphotochemical quenching and impairs their ability to scavenge ROS [58,59]. In our study, the reduction in chlorophyll and carotenoid contents under water stress conditions aligns with these reports, indicating a compromised photosynthetic capacity and increased susceptibility to oxidative damage [60,61].

The accumulation of protective compounds such as flavonoids and proline under water stress is a well-documented response in plants [62,63]. Flavonoids, known for their antioxidant properties, help mitigate oxidative stress by scavenging free radicals. Proline acts as an osmoprotectant, stabilizing cellular structures and maintaining cellular functions under stress conditions [64,65]. The elevated levels of these compounds under water stress coincide with SWIR reflectance changes, highlighting their mechanistic role in

osmotic adjustment and antioxidative defense, which together increase stress tolerance in soybean (Figures 1–12).

Hyperspectral reflectance analysis has emerged as a powerful tool for monitoring plant stress responses [32]. In our study, distinct shifts in reflectance patterns were observed across the VIS, NIR, and SWIR regions under different water regimes. These spectral changes are indicative of physiological alterations in plants, such as changes in pigment content and water status caused by changes in the thylakoid membrane. For example, a study by Almakas et al., (2025) [66] revealed that cold stress (similarly water or drought stress) in soybean led to significant changes in spectral reflectance, reflecting alterations in chlorophyll fluorescence and pigment composition. Similarly, our findings of spectral shifts under water stress conditions suggest that hyperspectral reflectance can serve as a noninvasive and efficient method for the early detection of stress in soybean plants [67].

The integration of physiological, biochemical, and hyperspectral data provides a comprehensive approach for understanding plant responses to water stress [2,66]. By combining these datasets, we can develop predictive models to assess plant health and stress levels more accurately. For example, machine learning algorithms can be trained on hyperspectral data to classify stress levels on the basis of spectral features, enabling real-time monitoring of crop health [20,21]. Moreover, the identification of specific spectral bands associated with stress indicators can aid in the development of targeted remote sensing tools for precision agriculture.

Therefore, spectral reflectance shifts are a direct consequence of pigment degradation and metabolic adjustment under water stress, highlighting how hyperspectral sensors capture underlying physiological mechanisms noninvasively [8,68,69]. This mechanistic understanding supports both breeding for drought tolerance and improved management of water-limited systems.

4.2. Principal Component Analysis and Model Performance

Principal component analysis (PCA) of the hyperspectral reflectance data effectively revealed spectral variability across eleven water treatments in soybean (*Glycine max*). The first two principal components accounted for 88% of the variance, clearly distinguishing treatments on the basis of water availability. This aligns with findings by Furlanetto et al. (2024) [70], who utilized PCA to classify potassium deficiency in soybean plants, achieving 100% variance across various developmental stages and seasons.

The separation of the well-watered treatments from those under intermediate and severe water deficit conditions underscores the utility of PCA in capturing water stress-induced spectral changes. This approach has been corroborated by studies [28,47], which applied PCA to hyperspectral data to monitor plant stress responses, demonstrating its efficacy in classifying varying stress levels.

With respect to model performance, machine learning algorithms such as random forest (RF), gradient boosting, and multilayer perceptron (MLP) have achieved perfect classification accuracy in predicting water regimes on the basis of hyperspectral data. These ensemble methods effectively handle complex, high-dimensional data, as supported by research by Furlanetto et al. (2020) [70], who reported high classification accuracies when RF was used to identify potassium deficiency in soybean.

Conversely, simpler models such as support vector machines (SVMs) and k-nearest neighbors (KNNs) struggled with misclassifications, particularly between regimes with similar spectral signatures. This highlights a fundamental issue in the use of traditional models. For example, while they are computationally efficient, they lack the capacity to capture the complex interactions between multiple spectral features. These models are

limited by their inherent assumption of linearity or proximity-based classification, which often fails when the spectral data are highly variable and multidimensional [71–74].

Therefore, while PCA serves as a robust tool for dimensionality reduction and visualization of spectral data, the integration of advanced machine learning algorithms enhances the predictive accuracy for water stress classification in soybean. The combination of PCA and ensemble learning models offers a powerful framework for the nondestructive assessment of plant physiological responses to varying water regimes, facilitating informed decision-making in precision agriculture to reduce water stress in plants [68,75,76].

4.3. Correlation Analysis and Implications for Spectral Indices

The selection of the most responsive wavelengths identified specific wavelengths within the VIS–NIR and SWIR regions as pivotal for predicting soybean leaf physiological traits under varying water regimes. Notably, wavelengths of approximately 1450 nm, 1940 nm, and 2200 nm have emerged as critical for assessing water status and stress markers. This aligns with findings of Wijewardana et al. (2019) [77], as SWIR reflectance specifically tracks changes in leaf water content and cell wall plasticity, reflecting the initial molecular events that precede irreversible tissue damage.

The convergence of multiple variable selection algorithms, partial least squares regression (PLSR), random forest (RF), and variable importance in projection (VIP), on these spectral regions underscores their robustness for the nondestructive assessment of soybean leaf physiology [20,21]. For example, PLSR has been effectively utilized to model leaf physiological responses to water stress, as demonstrated in studies by Wijewardana et al. (2019) [77].

Interestingly, our study also revealed correlations between spectral reflectance and biochemical markers of stress, such as proline and flavonoids [78]. The accumulation of these compounds, known for their roles in osmotic regulation and antioxidative defense, respectively, was reflected in spectral changes, suggesting that hyperspectral imaging can serve as a noninvasive tool to monitor plant physiological responses to water stress.

Furthermore, the identification of optimal spectral bands for predicting key physiological traits can enhance the development of spectral indices tailored for water stress detection. For example, indices incorporating wavelengths of approximately 1450 nm, 1940 nm, and 2200 nm could be particularly effective in monitoring the water status of soybean. This approach aligns with the findings of Wong et al. (2023) [79], who emphasized the importance of specific spectral bands in developing effective vegetation indices for drought monitoring.

Collectively, these results validate that hyperspectral reflectance, combined with robust chemometric modelling, enables reliable and simultaneous prediction of a wide array of physiological and biochemical traits in soybean leaves. However, they also highlight the importance of variable selection and model calibration for traits with weaker spectral signatures or higher biological noise, emphasizing the need for continuous refinement and validation across diverse genotypes and environments.

4.4. Wavelength Selection for Predictive Modelling

Hyperspectral datasets comprising thousands of contiguous bands offer detailed plant physiological insights, yet their practical application hinges on identifying minimal but informative spectral subsets [80]. The hybrid feature-selection strategy, which integrates partial least squares regression (PLSR) with the ensemble (random forest), kernel-based (support vector machine) and Boruta algorithms, demonstrates that a strategic dozen wavelengths capture over 95 % of the predictive variance for key leaf traits. Figure 9A–Q map algorithm-specific band importance scores, consistently showing chlorophyll a and

b absorption peaks ($\approx 550\text{--}750\text{ nm}$) as primary predictors of photosynthetic capacity and stress onset, corroborating established physiological absorption profiles.

Shortwave-infrared (SWIR) absorption windows at $\sim 1450\text{ nm}$ and $\sim 1940\text{ nm}$ are crucial for estimating water content and turgor, with independent validation yielding R^2 values > 0.74 (Figure 10A–C). These bands coincide with first and second water vibrational overtones, underscoring their mechanistic link to leaf hydration. Furthermore, mid-infrared overtone bands at $\sim 1700\text{ nm}$ and $\sim 2230\text{ nm}$ correlate strongly (R^2 test > 0.65 ; Figure 12) with cell-wall constituents such as cellulose and lignin, reflecting their absorption by C–H and O–H bond vibrations. This spectral pinpointing aligns with molecular absorption theory and provides direct insight into cell-wall maturity and structural rigidity [81,82].

The transition from full-spectrum analyses to a targeted twelve-band model reduces the data volume and processing time by $>80\%$, while maintaining the model error within $\pm 5\%$. This efficiency gain is critical for real-time, in-field monitoring, such as UAV platforms and autonomous greenhouse systems can incorporate tunable laser-line filters at these wavelengths [68,69], enabling continuous mapping of hydration, nutrient status and biomechanical properties. Decision-support algorithms can then trigger precision interventions, dynamic irrigation adjustments, targeted nutrient delivery or the application of growth regulators well before conventional stress symptoms manifest.

Re-evaluation of conventional vegetation indices (e.g., NDVI and PRI) via the existing full-spectrum models against trait-tuned hyperspectral indices (“Cell Wall Index” blending red-edge and $\sim 2230\text{ nm}$; “Turgor Stress Index” combining red-edge and $\sim 1450\text{ nm}$) demonstrated improvements of approximately 10–15% in R^2 during both calibration and independent validation (Figure 10D–F). These indices were derived from the same dataset and model parameters, confirming that strategic wavelength selection alone enhances predictive performance without requiring retraining of new models [41,69,83]. Trait-tuned HVIs thus distil multi-band complexity into single metrics directly correlated with agronomic endpoints [10,22], facilitating seamless integration into precision-agriculture dashboards for both water-limited systems and biomass-focused bioenergy crops.

The spectral correlation matrices show patchwork patterns of high-correlation hotspots interspersed with low-information regions, highlighting the unequal distribution of explanatory power across wavelengths, similar to [84,85]. This nonuniformity validates that a tiered modelling approach by PLSR offers rapid calibration and interpretable loadings for operational sensor deployment; random forest excels at detecting spectral anomalies in time-series data; and Boruta maintains ongoing marker discovery, adapting to novel genotypic and environmental conditions.

In summary, while most algorithms converge on the red-edge region (particularly $\sim 700\text{ nm}$) and classical water absorption bands (notably 1450, 1940, and 2200 nm) [41,69,83], some models highlight additional diagnostic bands (e.g., 675, 1200, and $>2100\text{ nm}$) depending on the target trait (Figures 1–12). This convergence across multiple feature selection methods underscores the physiological relevance and robustness of these specific wavelengths as predictors of drought-related responses in soybean [10,22]. The identification of these consistent “spectral hot spots” suggests that future proximal or UAV-based sensors can be engineered to target a minimal set of key wavelengths, drastically reducing sensor complexity and data redundancy without compromising prediction accuracy.

Expanding upon these findings, future work should explore the robustness of the twelve-band model under varying illumination angles, leaf orientations and canopy structures, as well as its transferability across species and developmental stages. Integrating these discrete wavelengths into miniaturized spectrometers or hyperspectral snapshot cameras could revolutionize plant phenotyping workflows, enabling high-throughput screening in breeding programmes and real-time crop monitoring at the field scale.

5. Conclusions

This study demonstrates that a minimal set of 12 strategically selected wavelengths—primarily in the red-edge (550–750 nm) and shortwave infrared (1450, 1940, 2200 nm) regions—enables highly accurate, nondestructive predictions of physiological and biochemical drought responses in soybean under controlled conditions. The ensemble machine learning models (random forest, gradient boosting, and multilayer perceptron) achieved over 95% accuracy in classifying eleven water regimes, and the PLSR models predicted key leaf traits with R^2 values of 0.96 (cellulose) and 0.88 (chlorophyll a). This compact-band approach reduced the data dimensionality by more than 80% with minimal loss of predictive power, providing a cost-efficient basis for real-time drought phenotyping and sensor development.

Our hybrid feature selection and modelling framework is novel compared with previous studies that relied on full-spectrum data or generic indices, offering practical advances for precision agriculture. Future studies should validate these findings under diverse field conditions and further integrate hyperspectral and other sensing modalities for robust crop monitoring.

Author Contributions: Conceptualization, C.A.d.O., N.G.V., W.C.A., M.R.N. and R.F.; Data curation, C.A.d.O., J.A.M.D., M.R.N. and R.F.; Formal analysis, C.A.d.O., N.G.V., W.A.M., D.H.S.d.M., R.H.F., L.G.T.C., A.S.R., R.B.d.O., J.A.M.D., M.R.N. and R.F.; Funding acquisition, R.B.d.O., M.L.C., J.A.M.D., M.R.N. and R.F.; Investigation, C.A.d.O., N.G.V., W.A.M., J.V.F.G., D.H.S.d.M., R.H.F., L.G.T.C., R.B.d.O., M.L.C., J.A.M.D., M.R.N. and R.F.; Methodology, C.A.d.O., N.G.V., W.A.M., J.V.F.G., R.H.F., L.G.T.C., A.S.R., W.C.A., R.B.d.O., M.L.C., J.A.M.D., M.R.N. and R.F.; Project administration, M.R.N. and R.F.; Resources, M.R.N. and R.F.; Software, C.A.d.O., N.G.V., W.A.M., J.V.F.G., R.H.F., L.G.T.C., A.S.R., W.C.A., R.B.d.O., M.L.C., J.A.M.D., M.R.N. and R.F.; Supervision, M.R.N. and R.F.; Validation, C.A.d.O., W.A.M., D.H.S.d.M., R.H.F., L.G.T.C., A.S.R., R.B.d.O., M.L.C., J.A.M.D., M.R.N. and R.F.; Visualization, M.R.N. and R.F.; Writing—original draft, C.A.d.O., N.G.V., W.A.M., J.V.F.G., D.H.S.d.M., R.H.F., L.G.T.C., A.S.R., W.C.A., R.B.d.O., M.L.C., J.A.M.D., M.R.N. and R.F.; Writing—review & editing, C.A.d.O., N.G.V., W.A.M., D.H.S.d.M., R.H.F., L.G.T.C., W.C.A., R.B.d.O., M.L.C., J.A.M.D., M.R.N. and R.F. All authors have read and agreed to the published version of the manuscript.

Funding: Programa de Apoio à Fixação de Jovens Doutores no Brasil (CNPq 168180/2022–7), Fundação Araucária (CP 19/2022—Jovens Doutores), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We would like to thank the editors and reviewers for their comments and suggestions to improve our work and the State University of Maringá for supporting the analyses.

Conflicts of Interest: The authors declare that they have no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. dos Santos, G.L.A.A.; Reis, A.S.; Besen, M.R.; Furlanetto, R.H.; Rodrigues, M.; Crusiol, L.G.T.; de Oliveira, K.M.; Falcioni, R.; de Oliveira, R.B.; Batista, M.A.; et al. Spectral Method for Macro and Micronutrient Prediction in Soybean Leaves Using Interval Partial Least Squares Regression. *Eur. J. Agron.* **2023**, *143*, 126717. [[CrossRef](#)]
2. Crusiol, L.G.T.; Nanni, M.R.; Furlanetto, R.H.; Sibaldelli, R.N.R.; Sun, L.; Gonçalves, S.L.; Foloni, J.S.S.; Mertz-Henning, L.M.; Nepomuceno, A.L.; Neumaier, N.; et al. Assessing the Sensitive Spectral Bands for Soybean Water Status Monitoring and Soil Moisture Prediction Using Leaf-Based Hyperspectral Reflectance. *Agric. Water Manag.* **2023**, *277*, 108089. [[CrossRef](#)]

3. Vargas-Almendra, A.; Ruiz-Medrano, R.; Núñez-Muñoz, L.A.; Ramírez-Pool, J.A.; Calderón-Pérez, B.; Xoconostle-Cázares, B. Advances in Soybean Genetic Improvement. *Plants* **2024**, *13*, 3073. [[CrossRef](#)]
4. Yoosefzadeh-Najafabadi, M.; Tulpan, D.; Eskandari, M. Using Hybrid Artificial Intelligence and Evolutionary Optimization Algorithms for Estimating Soybean Yield and Fresh Biomass Using Hyperspectral Vegetation Indices. *Remote Sens.* **2021**, *13*, 2555. [[CrossRef](#)]
5. Flexas, J.; Medrano, H. Drought-Inhibition of Photosynthesis in C₃plants: Stomatal and Non-Stomatal Limitations Revisited. *Ann. Bot.* **2002**, *89*, 183–189. [[CrossRef](#)]
6. Chaudhry, S.; Sidhu, G.P.S. Climate Change Regulated Abiotic Stress Mechanisms in Plants: A Comprehensive Review. *Plant Cell Rep.* **2022**, *41*, 1–31. [[CrossRef](#)]
7. Wang, Q.; Zuo, Z.; Wang, X.; Gu, L.; Yoshizumi, T.; Yang, Z.; Yang, L.; Liu, Q.; Liu, W.; Han, Y.-J.; et al. Photoactivation and Inactivation of in Arabidopsis Cryptochrome 2. *Science* **2016**, *354*, 343–347. [[CrossRef](#)] [[PubMed](#)]
8. Shurygin, B.; Chivkunova, O.; Solovchenko, O.; Solovchenko, A.; Dorokhov, A.; Smirnov, I.; Astashev, M.E.; Khort, D. Comparison of the Non-Invasive Monitoring of Fresh-Cut Lettuce Condition with Imaging Reflectance Hyperspectrometer and Imaging PAM-Fluorimeter. *Photonics* **2021**, *8*, 425. [[CrossRef](#)]
9. Nievola, C.C.; Carvalho, C.P.; Carvalho, V.; Rodrigues, E. Rapid Responses of Plants to Temperature Changes. *Temperature* **2017**, *4*, 371–405. [[CrossRef](#)]
10. Falcioni, R.; Gonçalves, J.V.F.; de Oliveira, K.M.; de Oliveira, C.A.; Reis, A.S.; Crusiol, L.G.T.; Furlanetto, R.H.; Antunes, W.C.; Cezar, E.; de Oliveira, R.B.; et al. Chemometric Analysis for the Prediction of Biochemical Compounds in Leaves Using UV-VIS-NIR-SWIR Hyperspectroscopy. *Plants* **2023**, *12*, 3424. [[CrossRef](#)]
11. Falcioni, R.; de Oliveira, C.A.; Vedana, N.G.; Mendonça, W.A.; Gonçalves, J.V.F.; da Silva Haubert, D.D.F.; de Matos, D.H.S.; Reis, A.S.; Antunes, W.C.; Crusiol, L.G.T.; et al. Progressive Water Deficit Impairs Soybean Growth, Alters Metabolic Profiles, and Decreases Photosynthetic Efficiency. *Plants* **2025**, *14*, 2615. [[CrossRef](#)] [[PubMed](#)]
12. Falcioni, R.; de Oliveira, R.B.; Chicati, M.L.; Antunes, W.C.; Demattê, J.A.M.; Nanni, M.R. Fluorescence and Hyperspectral Sensors for Nondestructive Analysis and Prediction of Biophysical Compounds in the Green and Purple Leaves of Tradescantia Plants. *Sensors* **2024**, *24*, 6490. [[CrossRef](#)] [[PubMed](#)]
13. Hassanzadeh, A.; Murphy, S.P.; Pethybridge, S.J.; van Aardt, J. Growth Stage Classification and Harvest Scheduling of Snap Bean Using Hyperspectral Sensing: A Greenhouse Study. *Remote Sens.* **2020**, *12*, 3809. [[CrossRef](#)]
14. Buchhorn, M.; Raynolds, M.K.; Walker, D.A. Influence of BRDF on NDVI and Biomass Estimations of Alaska Arctic Tundra. *Environ. Res. Lett.* **2016**, *11*, 1–13. [[CrossRef](#)]
15. Chaves, M.E.D.; De Carvalho Alves, M.; De Oliveira, M.S.; Sáfiadi, T. A Geostatistical Approach for Modeling Soybean Crop Area and Yield Based on Census and Remote Sensing Data. *Remote Sens.* **2018**, *10*, 680. [[CrossRef](#)]
16. Sims, D.A.; Gamon, J.A. Relationships between Leaf Pigment Content and Spectral Reflectance across a Wide Range of Species, Leaf Structures and Developmental Stages. *Remote Sens. Environ.* **2002**, *81*, 337–354. [[CrossRef](#)]
17. Jia, M.; Li, D.; Colombo, R.; Wang, Y.; Wang, X.; Cheng, T.; Zhu, Y.; Yao, X.; Xu, C.; Ouer, G.; et al. Quantifying Chlorophyll Fluorescence Parameters from Hyperspectral Reflectance at the Leaf Scale under Various Nitrogen Treatment Regimes in Winter Wheat. *Remote Sens.* **2019**, *11*, 2838.
18. Ryu, J.H.; Jeong, H.; Cho, J. Performances of Vegetation Indices on Paddy Rice at Elevated Air Temperature, Heat Stress, and Herbicide Damage. *Remote Sens.* **2020**, *12*, 2654. [[CrossRef](#)]
19. Li, K.; Wang, C.; Rong, G.; Wei, S.; Liu, C.; Yang, Y.; Sudu, B.; Guo, Y.; Sun, Q.; Zhang, J. Dynamic Evaluation of Agricultural Drought Hazard in Northeast China Based on Coupled Multi-Source Data. *Remote Sens.* **2023**, *15*, 57. [[CrossRef](#)]
20. Sobejano-Paz, V.; Mikkelsen, T.N.; Baum, A.; Mo, X.; Liu, S.; Köppl, C.J.; Johnson, M.S.; Gulyas, L.; García, M. Hyperspectral and Thermal Sensing of Stomatal Conductance, Transpiration, and Photosynthesis for Soybean and Maize under Drought. *Remote Sens.* **2020**, *12*, 3182. [[CrossRef](#)]
21. Osman, S.O.M.; Saad, A.S.I.; Tadano, S.; Takeda, Y.; Konaka, T.; Yamasaki, Y.; Tahir, I.S.A.; Tsujimoto, H.; Akashi, K. Chemical Fingerprinting of Heat Stress Responses in the Leaves of Common Wheat by Fourier Transform Infrared Spectroscopy. *Int. J. Mol. Sci.* **2022**, *23*, 2842. [[CrossRef](#)]
22. Crusiol, L.G.T.; Sun, L.; Sun, Z.; Chen, R.; Wu, Y.; Ma, J.; Song, C. In-Season Monitoring of Maize Leaf Water Content Using Ground-Based and UAV-Based Hyperspectral Data. *Sustainability* **2022**, *14*, 9039. [[CrossRef](#)]
23. da Silva Junior, C.A.; Nanni, M.R.; Shakir, M.; Teodoro, P.E.; de Oliveira-Júnior, J.F.; Cezar, E.; de Gois, G.; Lima, M.; Wojciechowski, J.C.; Shiratsuchi, L.S. Soybean Varieties Discrimination Using Non-Imaging Hyperspectral Sensor. *Infrared Phys. Technol.* **2018**, *89*, 338–350. [[CrossRef](#)]
24. Baio, F.H.R.; Santana, D.C.; Teodoro, L.P.R.; de Oliveira, I.C.; Gava, R.; de Oliveira, J.L.G.; da Silva Junior, C.A.; Teodoro, P.E.; Shiratsuchi, L.S. Maize Yield Prediction with Machine Learning, Spectral Variables and Irrigation Management. *Remote Sens.* **2023**, *15*, 79.

25. Vian, C.E.D.F.; Andrade Júnior, A.M.; Baricelo, L.G.; Da Silva, R.P. Origins, Evolução e Tendências Da Indústria de Máquinas Agrícolas. *Rev. Econ. Sociol. Rural* **2013**, *51*, 719–744.
26. Sonobe, R.; Wang, Q. Hyperspectral Indices for Quantifying Leaf Chlorophyll Concentrations Performed Differently with Different Leaf Types in Deciduous Forests. *Ecol. Inform.* **2017**, *37*, 1–9.
27. Braga, P.; Crusiol, L.G.T.; Nanni, M.R.; Caranhato, A.L.H.; Fuhrmann, M.B.; Nepomuceno, A.L.; Neumaier, N.; Farias, J.R.B.; Koltun, A.; Gonçalves, L.S.A.; et al. Vegetation Indices and NIR-SWIR Spectral Bands as a Phenotyping Tool for Water Status Determination in Soybean. *Precis. Agric.* **2021**, *22*, 249–266.
28. Mondal, S.; Karmakar, S.; Panda, D.; Pramanik, K.; Bose, B.; Singhal, R.K. Crucial Plant Processes under Heat Stress and Tolerance through Heat Shock Proteins. *Plant Stress* **2023**, *10*, 100227. [[CrossRef](#)]
29. Galvão, L.S.; Formaggio, A.R.; Tisot, D.A. Discrimination of Sugarcane Varieties in Southeastern Brazil with EO-1 Hyperion Data. *Remote Sens. Environ.* **2005**, *94*, 523–534. [[CrossRef](#)]
30. Kováč, D.; Veselovská, P.; Klem, K.; Večeřová, K.; Ač, A.; Peñuelas, J.; Urban, O. Potential of Photochemical Reflectance Index for Indicating Photochemistry and Light Use Efficiency in Leaves of European Beech and Norway Spruce Trees. *Remote Sens.* **2018**, *10*, 1202.
31. de Oliveira, K.M.; Ferreira Gonçalves, J.V.; Falcioni, R.; Almeida de Oliveira, C.; de Fatima da Silva Haubert, D.; Mendonça, W.A.; Teixeira Crusiol, L.G.; Berti de Oliveira, R.; Reis, A.S.; Cezar, E.; et al. Classification of Soil Horizons Based on VisNIR and SWIR Hyperespectral Images and Machine Learning Models. *Remote Sens. Appl. Soc. Environ.* **2024**, *36*, 101362. [[CrossRef](#)]
32. Sexton, T.; Sankaran, S.; Cousins, A.B. Predicting Photosynthetic Capacity in Tobacco Using Shortwave Infrared Spectral Reflectance. *J. Exp. Bot.* **2021**, *72*, 4373–4383. [[CrossRef](#)] [[PubMed](#)]
33. Falcioni, R.; de Oliveira, R.B.; Chicati, M.L.; Antunes, W.C.; Demattê, J.A.M.; Nanni, M.R. Estimation of Biochemical Compounds in Tradescantia Leaves Using VIS-NIR-SWIR Hyperspectral and Chlorophyll a Fluorescence Sensors. *Remote Sens.* **2024**, *16*, 1910.
34. Crusiol, L.G.T.; Sun, L.; Sibaldelli, R.N.R.; Junior, V.F.; Furlaneti, W.X.; Chen, R.; Sun, Z.; Wuyun, D.; Chen, Z.; Nanni, M.R.; et al. Strategies for Monitoring Within-Field Soybean Yield Using Sentinel-2 Vis-NIR-SWIR Spectral Bands and Machine Learning Regression Methods. *Precis. Agric.* **2022**, *23*, 1093–1123.
35. Falcioni, R.; Gonçalves, J.V.F.; de Oliveira, K.M.; Antunes, W.C.; Nanni, M.R. VIS-NIR-SWIR Hyperspectroscopy Combined with Data Mining and Machine Learning for Classification of Predicted Chemometrics of Green Lettuce. *Remote Sens.* **2022**, *14*, 6330. [[CrossRef](#)]
36. Gitelson, A.; Solovchenko, A. Non-Invasive Quantification of Foliar Pigments: Possibilities and Limitations of Reflectance- and Absorbance-Based Approaches. *J. Photochem. Photobiol. B Biol.* **2018**, *178*, 537–544.
37. Falcioni, R.; Moriwaki, T.; Gibin, M.S.; Vollmann, A.; Pattaro, M.C.; Giacomelli, M.E.; Sato, F.; Nanni, M.R.; Antunes, W.C. Classification and Prediction by Pigment Content in Lettuce (*Lactuca sativa* L.) Varieties Using Machine Learning and ATR-FTIR Spectroscopy. *Plants* **2022**, *11*, 3413. [[CrossRef](#)]
38. Ragaee, S. Antioxidant Activity and Nutrient Composition of Selected Cereals for Food Use. *Food Chem.* **2006**, *98*, 32–38. [[CrossRef](#)]
39. Izenman, A.J. *Modern Multivariate Statistical Techniques*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2008.
40. Carlson, T.N.; Ripley, D.A. On the Relation between NDVI, Fractional Vegetation Cover, and Leaf Area Index. *Remote Sens. Environ.* **1997**, *62*, 241–252. [[CrossRef](#)]
41. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298.
42. Ahamed, T.; Tian, L.; Zhang, Y.; Ting, K.C. A Review of Remote Sensing Methods for Biomass Feedstock Production. *Biomass and Bioenergy* **2011**, *35*, 2455–2469. [[CrossRef](#)]
43. Bannari, A.; Morin, D.; Bonn, F.; Huete, A.R. A Review of Vegetation Indices. *Remote Sens. Rev.* **1995**, *13*, 95–120. [[CrossRef](#)]
44. Blackburn, G.A. Spectral Indices for Estimating Photosynthetic Pigment Concentrations: A Test Using Senescent Tree Leaves. *Int. J. Remote Sens.* **1998**, *19*, 657–675. [[CrossRef](#)]
45. Falcioni, R.; Antunes, W.C.; Demattê, J.A.M.; Nanni, M.R. Biophysical, Biochemical, and Photochemical Analyses Using Reflectance Hyperspectroscopy and Chlorophyll a Fluorescence Kinetics in Variegated Leaves. *Biology* **2023**, *12*, 704. [[CrossRef](#)]
46. Peñuelas, J.; Gamon, J.A.; Fredeen, A.L.; Merino, J.; Field, C.B. Reflectance Indices Associated with Physiological Changes in Nitrogen- and Water-Limited Sunflower Leaves. *Remote Sens. Environ.* **1994**, *48*, 135–146. [[CrossRef](#)]
47. Hunt, E.R.; Rock, B.N. Detection of Changes in Leaf Water Content Using Near- and Middle-Infrared Reflectances. *Remote Sens. Environ.* **1989**, *30*, 43–54.
48. Cibula, W.G.; Zetka, E.F.; Rickman, D.L. Response of Thematic Mapper Bands to Plant Water Stress. *Int. J. Remote Sens.* **1992**, *13*, 1869–1880. [[CrossRef](#)]
49. Patil, P.P.; Jagtap, M.P.; Khatri, N.; Madan, H.; Vadduri, A.A.; Patodia, T. Exploration and Advancement of NDDI Leveraging NDVI and NDWI in Indian Semi-Arid Regions: A Remote Sensing-Based Study. *Case Stud. Chem. Environ. Eng.* **2024**, *9*, 100573.
50. Staszal, J.; Lupa, M.; Adamek, K.; Wilkosz, M.; Marcinkowska-Ochtyra, A.; Ochtyra, A. Spatial Insights into Drought Severity: Multi-Index Assessment in Małopolska, Poland, via Satellite Observations. *Remote Sens.* **2024**, *16*, 836. [[CrossRef](#)]

51. Gitelson, A.A.; Zur, Y.; Chivkunova, O.B.; Merzlyak, M.N. Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. *Photochem. Photobiol.* **2002**, *75*, 272. [[CrossRef](#)]
52. Vogelmann, J.E.; Rock, B.N.; Moss, D.M. Red Edge Spectral Measurements from Sugar Maple Leaves. *Int. J. Remote Sens.* **1993**, *14*, 1563–1575. [[CrossRef](#)]
53. Peñuelas, J.; Filella, I. Visible and Near-Infrared Reflectance Techniques for Diagnosing Plant Physiological Status. *Trends Plant Sci.* **1998**, *3*, 151–156. [[CrossRef](#)]
54. Peñuelas, J.; Filella, I.; Gamon, J.A. Assessment of Photosynthetic Radiation-Use Efficiency with Spectral Reflectance. *New Phytol.* **1995**, *131*, 291–296. [[CrossRef](#)]
55. Zheng, W.; Lu, X.; Li, Y.; Li, S.; Zhang, Y. Hyperspectral Identification of Chlorophyll Fluorescence Parameters of Suaeda Salsa in Coastal Wetlands. *Remote Sens.* **2021**, *13*, 2066. [[CrossRef](#)]
56. Barnes, M.L.; Breshears, D.D.; Law, D.J.; van Leeuwen, W.J.D.; Monson, R.K.; Fojtik, A.C.; Barron-Gafford, G.A.; Moore, D.J.P. Beyond Greenness: Detecting Temporal Changes in Photosynthetic Capacity with Hyperspectral Reflectance Data. *PLoS ONE* **2017**, *12*, e0189539. [[CrossRef](#)]
57. Nehra, A.; Kalwan, G.; Taneja, D.; Jangra, R.; Joshi, K.; Kumar, A.; Jain, P.K.; Nehra, K.; Ansari, M.W.; Singh, K.; et al. Comprehensive Structural, Evolutionary and Functional Analysis of Superoxide Dismutase Gene Family Revealed Critical Role in Salinity and Drought Stress Responses in Chickpea (*Cicer arietinum* L.). *Plant Physiol. Biochem.* **2025**, *226*, 110042. [[CrossRef](#)] [[PubMed](#)]
58. Valentini, R.; Epron, D.; De Angelis, P.; Matteucci, G.; Dreyer, E. In Situ Estimation of Net CO₂ Assimilation, Photosynthetic Electron Flow and Photorespiration in Turkey Oak (*Q. cerris* L.) Leaves: Diurnal Cycles under Different Levels of Water Supply. *Plant. Cell Environ.* **1995**, *18*, 631–640. [[CrossRef](#)]
59. Gill, S.S.; Tuteja, N. Reactive Oxygen Species and Antioxidant Machinery in Abiotic Stress Tolerance in Crop Plants. *Plant Physiol. Biochem.* **2010**, *48*, 909–930. [[CrossRef](#)]
60. Zhou, Y.H.; Zhang, Y.Y.; Zhao, X.; Yu, H.J.; Shi, K.; Yu, J.Q. Impact of Light Variation on Development of Photoprotection, Antioxidants, and Nutritional Value in *Lactuca sativa* L. *J. Agric. Food Chem.* **2009**, *57*, 5494–5500. [[CrossRef](#)]
61. Pinnola, A.; Bassi, R. Molecular Mechanisms Involved in Plant Photoprotection. *Biochem. Soc. Trans.* **2018**, *46*, 467–482. [[CrossRef](#)]
62. Gitelson, A.; Chivkunova, O.; Zhigalova, T.; Solovchenko, A. In Situ Optical Properties of Foliar Flavonoids: Implication for Non-Destructive Estimation of Flavonoid Content. *J. Plant Physiol.* **2017**, *218*, 258–264. [[CrossRef](#)]
63. Hichri, I.; Barrieu, F.; Bogs, J.; Kappel, C.; Delrot, S.; Lauvergeat, V. Recent Advances in the Transcriptional Regulation of the Flavonoid Biosynthetic Pathway. *J. Exp. Bot.* **2011**, *62*, 2465–2483. [[CrossRef](#)]
64. Zhou, Z.; Gao, H.; Ming, J.; Ding, Z.; Lin, X.; Zhan, R. Combined Transcriptome and Metabolome Analysis of Pitaya Fruit Unveiled the Mechanisms Underlying Peel and Pulp Color Formation. *BMC Genom.* **2020**, *21*, 734. [[CrossRef](#)]
65. Baranović, G.; Šegota, S. Infrared Spectroscopy of Flavones and Flavonols. Reexamination of the Hydroxyl and Carbonyl Vibrations in Relation to the Interactions of Flavonoids with Membrane Lipids. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2018**, *192*, 473–486. [[CrossRef](#)]
66. Almakas, A.; Elrys, A.S.; Desoky, E.-S.M.; Al-Shuraym, L.A.; Alhag, S.K.; Alshaharni, M.O.; Alnadari, F.; NanNan, Z.; Farooq, Z.; El-Tarabily, K.A.; et al. Enhancing Soybean Germination and Vigor under Water Stress: The Efficacy of Bio-Priming with Sodium Carboxymethyl Cellulose and Gum Arabic. *Front. Plant Sci.* **2025**, *15*, 1475148. [[CrossRef](#)] [[PubMed](#)]
67. Boshkovski, B.; Doupis, G.; Zapolska, A.; Kalaitzidis, C.; Koubouris, G. Hyperspectral Imagery Detects Water Deficit and Salinity Effects on Photosynthesis and Antioxidant Enzyme Activity of Three Greek Olive Varieties. *Sustainability* **2022**, *14*, 1432. [[CrossRef](#)]
68. Wang, D.; Cao, W.; Zhang, F.; Li, Z.; Xu, S.; Wu, X. A Review of Deep Learning in Multiscale Agricultural Sensing. *Remote Sens.* **2022**, *14*, 559.
69. Zhou, Q.; Yu, L.; Zhang, X.; Liu, Y.; Zhan, Z.; Ren, L.; Luo, Y. Fusion of UAV Hyperspectral Imaging and LiDAR for the Early Detection of EAB Stress in Ash and a New EAB Detection Index NDVI(776,678). *Remote Sens.* **2022**, *14*, 2428. [[CrossRef](#)]
70. Furlanetto, R.H.; Moriwaki, T.; Falcioni, R.; Pattaro, M.; Vollmann, A.; Sturion Junior, A.C.; Antunes, W.C.; Nanni, M.R. Hyperspectral Reflectance Imaging to Classify Lettuce Varieties by Optimum Selected Wavelengths and Linear Discriminant Analysis. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100400. [[CrossRef](#)]
71. Ge, Y.; Atefi, A.; Zhang, H.; Miao, C.; Ramamurthy, R.K.; Sigmon, B.; Yang, J.; Schnable, J.C. High-Throughput Analysis of Leaf Physiological and Chemical Traits with VIS–NIR–SWIR Spectroscopy: A Case Study with a Maize Diversity Panel. *Plant Methods* **2019**, *15*, 66.
72. Wang, L.; Chang, Q.; Li, F.; Yan, L.; Huang, Y.; Wang, Q.; Luo, L. Effects of Growth Stage Development on Paddy Rice Leaf Area Index Prediction Models. *Remote Sens.* **2019**, *11*, 361.
73. Hu, Y.; Wang, Z.; Li, X.; Li, L.; Wang, X.; Wei, Y. Nondestructive Classification of Maize Moldy Seeds by Hyperspectral Imaging and Optimal Machine Learning Algorithms. *Sensors* **2022**, *22*, 6064. [[CrossRef](#)]

74. Fu, P.; Meacham-Hensold, K.; Guan, K.; Bernacchi, C.J. Hyperspectral Leaf Reflectance as Proxy for Photosynthetic Capacities: An Ensemble Approach Based on Multiple Machine Learning Algorithms. *Front. Plant Sci.* **2019**, *10*, 730. [[CrossRef](#)] [[PubMed](#)]
75. Cotrozzi, L.; Lorenzini, G.; Nali, C.; Pellegrini, E.; Saponaro, V.; Hoshika, Y.; Arab, L.; Rennenberg, H.; Paoletti, E. Hyperspectral Reflectance of Light-Adapted Leaves Can Predict Both Dark- and Light-Adapted Chl Fluorescence Parameters, and the Effects of Chronic Ozone Exposure on Date Palm (*Phoenix dactylifera*). *Int. J. Mol. Sci.* **2020**, *21*, 6441. [[PubMed](#)]
76. Nalepa, J. Recent Advances in Multi- and Hyperspectral Image Analysis. *Sensors* **2021**, *21*, 6002. [[CrossRef](#)]
77. Wijewardana, C.; Reddy, K.R.; Krutz, L.J.; Gao, W.; Bellaloui, N. Drought Stress Has Transgenerational Effects on Soybean Seed Germination and Seedling Vigor. *PLoS ONE* **2019**, *14*, e0214977. [[CrossRef](#)]
78. Gururani, M.A.; Venkatesh, J.; Ganesan, M.; Strasser, R.J.; Han, Y.; Kim, J.-I.; Lee, H.Y.; Song, P.S. In Vivo Assessment of Cold Tolerance through Chlorophyll-a Fluorescence in Transgenic Zoysiagrass Expressing Mutant Phytochrome A. *PLoS ONE* **2015**, *10*, e0127200.
79. Wong, C.Y.S.; Gilbert, M.E.; Pierce, M.A.; Parker, T.A.; Palkovic, A.; Gepts, P.; Magney, T.S.; Buckley, T.N. Hyperspectral Remote Sensing for Phenotyping the Physiological Drought Response of Common and Tepary Bean. *Plant Phenomics* **2023**, *5*, 21. [[CrossRef](#)]
80. Sun, Y.; Liu, B.; Yu, X.; Yu, A.; Gao, K.; Ding, L. From Video to Hyperspectral: Hyperspectral Image-Level Feature Extraction with Transfer Learning. *Remote Sens.* **2022**, *14*, 5118.
81. Nogales-Bueno, J.; Baca-Bocanegra, B.; Rooney, A.; Miguel Hernández-Hierro, J.; José Heredia, F.; Byrne, H.J. Linking ATR-FTIR and Raman Features to Phenolic Extractability and Other Attributes in Grape Skin. *Talanta* **2017**, *167*, 44–50. [[CrossRef](#)]
82. Prats-Mateu, B.; Felhofer, M.; de Juan, A.; Gierlinger, N. Multivariate Unmixing Approaches on Raman Images of Plant Cell Walls: New Insights or Overinterpretation of Results? *Plant Methods* **2018**, *14*, 52. [[CrossRef](#)] [[PubMed](#)]
83. Zhang, Y.; Li, X.; Wang, C.; Zhang, R.; Jin, L.; He, Z.; Tian, S.; Wu, K.; Wang, F. PROSPECT-PMP+: Simultaneous Retrievals of Chlorophyll a and b, Carotenoids and Anthocyanins in the Leaf Optical Properties Model. *Sensors* **2022**, *22*, 3025. [[CrossRef](#)]
84. Chen, J.; de Hoogh, K.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzler, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U.A.; Katsouyanni, K.; et al. A Comparison of Linear Regression, Regularization, and Machine Learning Algorithms to Develop Europe-Wide Spatial Models of Fine Particles and Nitrogen Dioxide. *Environ. Int.* **2019**, *130*, 104934. [[CrossRef](#)] [[PubMed](#)]
85. Zhu, H.; Chu, B.; Zhang, C.; Liu, F.; Jiang, L.; He, Y. Hyperspectral Imaging for Presymptomatic Detection of Tobacco Disease with Successive Projections Algorithm and Machine-Learning Classifiers. *Sci. Rep.* **2017**, *7*, 4125. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.