ICPA
17th INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE
ConBAP 2026
CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO E DIGITAL
13-16 July 2026 | Porto Alegre, Brazil

A MultiModal Spectral-Robustness-LLM Pipeline for Non-Destructive Identification of *Loropetalum Chinese* Cultivars

Shekhar Suman Borah[1], Harshitha Manujnatha[2], Theoren Stroud[1], Prabha Sundaravadivel[1], Lakshman Tamil[2], Patricia Knight[3], Siva P. Kumpatla[4]

[1]Center for Robotics and Intelligent Systems, The University of Texas at Tyler, TX, USA
[2] The University of Texas at Dallas, Richardson, TX, USA.
[3]Coastal Research and Extension Center, Mississippi State University, Poplarville, MS, USA.
[4]Thad Cochran Southern Horticultural Laboratory, USDA-ARS, Poplarville, MS, USA.
Contact Email: psundaravadivel@uttyler.edu

A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil

Abstract.

High-throughput phenotyping of ornamental crops increasingly depends on spectral sensing methods that can capture subtle physiological and pigment-related differences among genotypes. This study presents an integrated hyperspectral imaging (HSI) pipeline for distinguishing three morphologically similar Loropetalum chinense cultivars: Cerise Charm, Purple Daybreak, and Red Diamond. The workflow combines preprocessing, feature engineering, and multi-method band selection with conventional machine-learning classifiers. Hyperspectral reflectance data from a Resonon Pika L system covering 367-1004 nm were processed using Savitzky-Golay smoothing, Standard Normal Variate normalization, derivative-based exploratory analysis, vegetation index calculation, and Principal Component Analysis. Band importance was assessed using four complementary methods: Analysis of Variance, Pearson correlation, Mutual Information, and Random Forest feature importance. A consensus set of 15 informative wavelengths was then selected. Classification models, including Support Vector Machines, Random Forest, XGBoost, and a shallow 1D convolutional neural network, were evaluated using full spectra, the reduced 15-band subset, and engineered spectral features derived from vegetation indices and regional reflectance means. Whole-plant similarity analysis based on the Spectral Angle Mapper showed high overall similarity, above 90%, among cultivars, but also consistent leaf-level differences that supported classification. Random Forest and XGBoost achieved accuracy of about 0.80-0.83 with full spectra and retained solid performance, about 0.73-0.77, with the compact 15-band subset. The engineered feature set performed best, with XGBoost exceeding 90% accuracy. The resulting compact and physiologically interpretable spectral features also support future integration with multimodal AI and LLM-based phenotyping systems. These findings demonstrate the potential of hyperspectral sensing for reliable non-destructive cultivar identification and low-cost ornamental phenotyping.

Keywords.
Hyperspectral Imaging (HSI), Loropetalum chinense, Large Language Model (LLM), Cultivar Classification, Machine Learning (ML), Spectral Phenotyping, Ornamental Plants.

A MultiModal Spectral-Robustness-LLM Pipeline for Non-Destructive Identification of *Loropetalum* Chinese Cultivars

1 Introduction

Modern horticulture increasingly relies on rapid, non-destructive tools for cultivar identification, plant health assessment, and breeding support. Hyperspectral imaging (HSI) is especially useful because it captures reflectance across hundreds of narrow wavelengths in the visible and near-infrared range, revealing information about pigments, water content, and leaf structure (Lu et al. 2022). Previous studies have shown that hyperspectral data can distinguish cultivars and growth stages in several crops, while feature-selection methods can reduce dimensionality and sensor cost (Amankwah 2015). *Loropetalum chinense*, a popular ornamental shrub with colorful foliage and ribbon-like flowers, is difficult to identify reliably by appearance alone. Many cultivars have been developed, but the visible differences are often subtle, especially at early growth stages. Leaf color ranges from green to deep burgundy, while flowers vary from white to pink depending on anthocyanin and flavonoid content. A spectral approach could support nurseries, breeders, and researchers by improving labeling accuracy and quality control. Leaves and flowers are also expected to show different spectral signatures because of their distinct pigment compositions. Floral pigments such as anthocyanins, carotenoids, and betalains absorb strongly in the blue-green and red regions (Angel et al. 2025), whereas leaves typically show low visible reflectance, a red-edge rise near 700 nm, and high near-infrared scattering.

HSI provides much finer spectral detail than RGB or multispectral imaging. By measuring narrow and contiguous wavebands, it can detect subtle biochemical and structural differences that are otherwise difficult to observe. This capability has been applied to disease detection, nutrient assessment, and cultivar discrimination across a range of crops (Gutiérrez et al. 2018) and has become an important tool in high-throughput phenotyping. Extending measurements into the shortwave-infrared range can further reveal absorption features associated with lignin, cellulose, and water (Gill et al. 2022). Machine-learning (ML) methods such as Support Vector Machine (SVM), Random Forest, and XGBoost are commonly used for hyperspectral analysis because they can model nonlinear relationships and feature interactions. These approaches have been effective for detecting stress, nutrient deficiency, and disease in many plant species, while dimensionality reduction techniques and deep-learning models can improve performance when enough training data are available (Patiluna et al. 2025; Liu et al. 2022).

Despite these advances, ornamental shrubs such as *Loropetalum chinense* remain relatively understudied. Recent computer-vision research has demonstrated the feasibility of UAV-based analysis and large-scale phenotyping of *Loropetalum* canopies under challenging field conditions (Manjunatha et al., 2026), but systematic hyperspectral characterization and cultivar-level discrimination have not yet been explored. This work addresses that gap by analyzing the spectral signatures of three widely grown *L. chinense* cultivars: Cerise Charm, Purple Daybreak, and Red Diamond, using leaf and flower spectra collected with a Resonon Pika L hyperspectral camera. Although these cultivars appear visually similar, they differ in pigment levels and leaf structure, which may produce measurable differences in their spectral profiles. Capturing these patterns can improve cultivar identification and support high-throughput phenotyping in ornamental horticulture. To this end, hyperspectral reflectance data (367–1004 nm) were processed using Savitzky-Golay smoothing, SNV normalization, derivative analysis, vegetation indices, and PCA. Informative wavelengths were identified through ANOVA, Pearson correlation, Mutual Information, and Random Forest feature importance, resulting in a compact 15-band subset. SVM, Random Forest, XGBoost, and a shallow 1D CNN were evaluated using full spectra, reduced-band data, and engineered features. Cultivar differences were further analyzed using SAM, PCA, derivative spectra, and mean reflectance profiles to support non-destructive cultivar discrimination and reduced-band sensor development.

2 Literature survey

HSI combines spectroscopy and imaging to measure absorption, scattering, and reflectance across a broad wavelength range. In the visible to near-infrared region, these measurements capture information related to pigments such as chlorophyll and other biochemical components (Slaton et al. 2001; Tang et al. 2024). Hyperspectral data are typically stored as a three-dimensional cube, where each pixel contains a full spectrum, allowing chemical and structural variation to be mapped across leaves or canopy surfaces. For example, Sun et al. 2019 used pixel-level hyperspectral visualization to assess moisture patterns in tea plants.

Compared with RGB imaging, hyperspectral cameras provide much richer spectral detail. By capturing reflectance in many narrow bands, they can detect early physiological changes linked to stress or disease before visible symptoms appear (de Castro et al. 2021). This has supported applications in phenotyping, plant-health monitoring, and water-stress detection (Lowe et al. 2017). Because spectral changes often appear before visible decline, HSI is widely used for early detection of pests, diseases, and abiotic stress (Behmann et al. 2018; Duarte-Carvajalino et al. 2021). It also helps researchers study natural variation in stress tolerance and supports more sustainable crop-management strategies (Sytar et al. 2017). Water-stress detection is one of the most studied applications. Hyperspectral data combined with ML models have shown strong performance in estimating crop water status (Duarte-Carvajalino et al. 2021). Spectral indices also provide simple quantitative indicators of water stress and can support irrigation decisions (Osco et al. 2019). Some studies have shown that hyperspectral signatures can distinguish drought from disease and other stress types, enabling more targeted management (Susič et al. 2018). When paired with deep learning, these data can also improve pathogen detection and track disease progression over time (Nagasubramanian et al. 2019).

ML plays a central role in extracting useful information from hyperspectral data. These methods can handle large spectral datasets and identify subtle patterns related to stress or physiological change. Studies have reported improved water-stress assessment with machine learning (Virnodkar et al. 2020) and have shown strong relationships between hyperspectral reflectance and drought indicators such as stomatal conductance and leaf water potential (Wong et al. 2023). At the same time, UAV-borne hyperspectral systems have expanded access to high-resolution data across large fields (Liu et al. 2021), while improved sensor technology and new stress-sensitive indices continue to strengthen field-based measurements (Vairavan 2024). Pigments strongly indicate plant function, and the red-edge region is especially useful because it reflects changes in pigments, leaf structure, and stress. Floral tissues also have distinct spectral signatures, making them useful for classification when leaf spectra are not enough. Most hyperspectral studies follow the same basic pipeline of preprocessing, feature extraction, and classification (Goetz et al. 1985). The Common preprocessing steps include Savitzky-Golay smoothing and SNV normalization to stabilize reflectance curves (Geladi and Kowalski 1986). Vegetation indices provide interpretable features, but high-dimensional data often require additional methods such as ANOVA, Mutual Information, or ReliefF to identify informative wavelengths and avoid overfitting (Aishwarya et al. 2023). ML models are widely used for classification. Support Vector Machines perform well in high-dimensional spectral spaces (Pal and Foody 2010), while Random Forest and XGBoost offer strong accuracy and feature-importance estimates (Breiman 2001; Chen and Guestrin 2016). Deep learning methods, including convolutional neural networks (CNNs), can automatically learn complex spectral and spatial features directly from hyperspectral and image data. However, their successful deployment typically requires large, well-annotated datasets for training and validation (Li et al., 2017; Mohanty et al., 2016). Despite these advances, HSI is still underused in ornamental horticulture. No structured study has evaluated hyperspectral classification of *Loropetalum chinense* cultivars using both leaf and flower signatures. This work addresses that gap by applying a complete hyperspectral analysis pipeline to an ornamental species and assessing its potential for reliable cultivar identification.

3 Materials and Methods

3.1 Plant material and hyperspectral acquisition

Three commercially available *Loropetalum chinense* cultivars were used in this study: Cerise Charm, which has pink flowers and burgundy foliage; Purple Daybreak, with light purple flowers and green-purple foliage; and Red Diamond, with red flowers and dark burgundy foliage. Potted plants were grown under uniform greenhouse conditions. At peak bloom, representative flowers and fully expanded leaves were detached and gently pressed against a matte black background to reduce specular reflection. Hyperspectral images were collected with a Resonon Pika L camera covering 367-1004 nm with approximately 2 nm sampling and 300 bands under stabilized halogen illumination. White-reference and dark-current corrections were applied in Spectronon to obtain calibrated reflectance. For each organ and cultivar, multiple regions of interest were manually selected to capture within-organ variability. In total, 204 ROI spectra were exported: 32 flower and 38 leaf spectra for Cerise Charm, 27 flower and 45 leaf spectra for Purple Daybreak, and 31 flower and 31 leaf spectra for Red Diamond. Representative RGB images of flowers and leaves from the three cultivars are shown in Figure 1.

Figure 1: Representative flower (top row) and leaf (bottom row) samples from the three *Loropetalum chinense* cultivars used in this study: (a) Cerise Charm, (b) Purple Daybreak, and (c) Red Diamond.

3.2 Spectral data handling and preprocessing

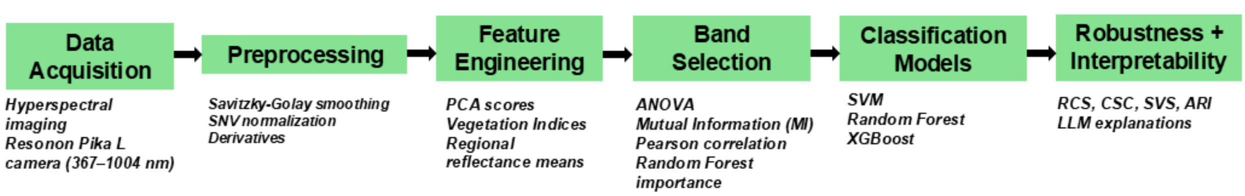


Figure 2: Overview of the hyperspectral cultivar-classification pipeline.

The ROI-level spectra were exported as CSV files and analyzed in Python using pandas. Since all spectra shared a common wavelength range (367.08-1003.84 nm, ~2 nm intervals), no resampling was required. Raw spectra were smoothed using an 11-point, second-order Savitzky-Golay filter, followed by Standard Normal Variate (SNV) normalization to reduce scattering effects. First and second derivatives were calculated to highlight subtle spectral features and were used for visualization and exploratory analysis only. The overall workflow is illustrated in Figure 2.

3.3 Feature extraction

To reduce dimensionality while retaining important spectral information, several feature representations were extracted from the SNV-normalized spectra. PCA was first applied to the 300-band data, and the first ten principal components, explaining approximately 99% of the variance, were retained. Mean reflectance values were also calculated for five spectral regions: blue (400-500 nm), green (500-600 nm), red (600-700 nm), red-edge (700-750 nm), and near-infrared (750–1000 nm). In addition, vegetation indices including Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Red-Edge Chlorophyll Index (CI_{re}), and the Anthocyanin Reflectance Indices (ARI1 and ARI2), were computed to capture chlorophyll and anthocyanin-related information. The resulting feature pool consisted of the full spectra, PCA scores, regional mean reflectance values, and vegetation indices. The vegetation indices used in this study are defined as follows:

$$NDVI = \frac{R_{NIR} - R_{Red}}{R_{NIR} + R_{Red}} \tag{1}$$

$$GNDVI = \frac{R_{NIR} - R_{Green}}{R_{NIR} + R_{Green}} \tag{2}$$

$$CI_{re} = \frac{R_{NIR}}{R_{re}} - 1 \tag{3}$$

$$ARI1 = \frac{1}{R_{550}} - \frac{1}{R_{700}} \tag{4}$$

$$ARI2 = R_{NIR}\left(\frac{1}{R_{550}} - \frac{1}{R_{700}}\right) \tag{5}$$

3.4 Band selection

To identify the most informative wavelengths for distinguishing the three *Loropetalum chinense* cultivars, we used four band-selection methods: ANOVA, Mutual Information, Pearson correlation, and Random Forest importance. These methods highlighted the blue-green region, around 600 nm, the red-edge region, and parts of the near-infrared range. The top wavelengths from each method were merged and overlapping bands were removed. From this consensus set, 15 wavelengths were selected for the reduced-band experiments. This subset preserved the most important pigment- and structure-related information while greatly reducing dimensionality.

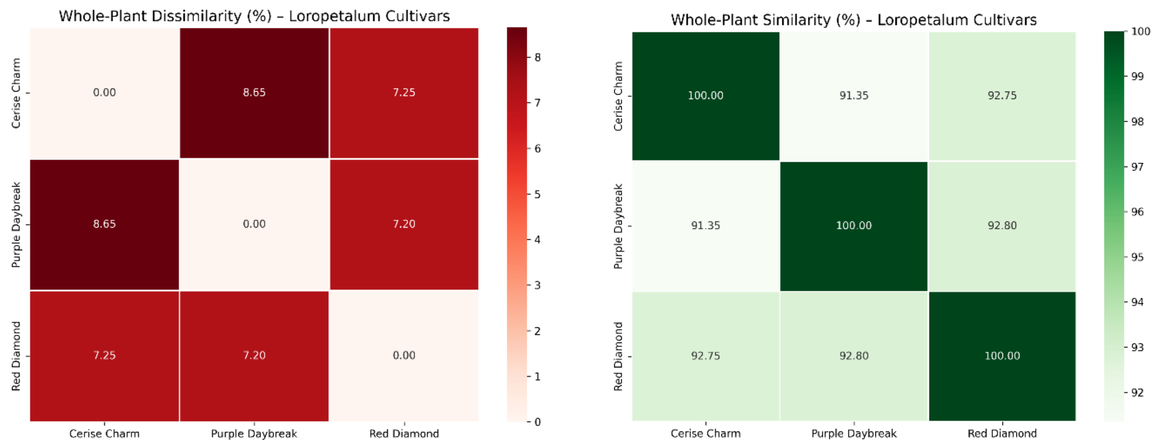
3.5 Machine-learning models

The hyperspectral features were evaluated using a three-class classification task for the three *Loropetalum* cultivars. The data were split into training and testing sets using an 80/20 stratified split. SVM with an RBF kernel, Random Forest, XGBoost, and a shallow 1D CNN were trained using the full spectra, the 15-band subset, and the engineered features. Performance was measured using 5-fold cross-validation, overall accuracy, macro F1-score, weighted F1-score, and confusion matrices.

4. Results

4.1 Spectral Similarity and Dissimilarity Analysis (SAM-Based)

Spectral similarity among the three *Loropetalum chinense* cultivars was evaluated using Spectral Angle Mapper (SAM) on the 300-band SNV-normalized spectra. Average spectral angles between cultivar pairs were converted to percentage similarity and dissimilarity values for easier interpretation. The whole-plant dissimilarity heatmap (Figure 3(a)) shows clear but modest separation among cultivars, with dissimilarity values ranging from 7.2% to 8.7%. The largest separation was observed between Cerise Charm and Purple Daybreak, while Red Diamond exhibited intermediate distances. Corresponding similarity values (Figure 3(b)) ranged from 91.3% to 92.8%, indicating high overall spectral resemblance with consistent cultivar-specific differences. Organ-wise analysis (Table 1) showed that leaves provided greater discrimination (86.8-90.8% similarity) than flowers (94-96% similarity), suggesting that leaf spectra contain more distinctive information. Despite the high overall similarity, these measurable differences support reliable cultivar discrimination and are consistent with the strong classification performance obtained in subsequent analyses.



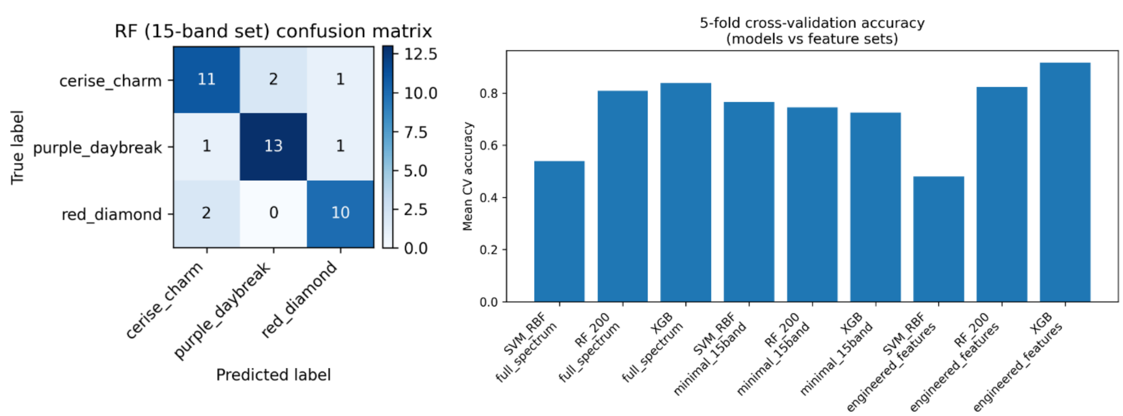
(a) (b)

Figure 3. Whole-plant spectral similarity and dissimilarity among *Loropetalum chinense* cultivars based on SAM analysis.
(a) Dissimilarity (%) heatmap showing spectral separation between cultivars.
(b) Similarity (%) heatmap illustrating the high but non-identical reflectance patterns shared among cultivars.

Table 1. Organ-wise spectral similarity and dissimilarity between Loropetalum cultivars based on SAM applied to full-spectrum (350-1000 nm) leaf and flower reflectance.

Organ	Cultivar pair	Similarity (%)	Dissimilarity (%)
Leaves	Cerise Charm vs Purple Daybreak	86.8	13.2
Leaves	Cerise Charm vs Red Diamond	90.8	9.2
Leaves	Purple Daybreak vs Red Diamond	90.3	9.7
Flowers	Cerise Charm vs Purple Daybreak	95.9	4.1
Flowers	Cerise Charm vs Red Diamond	94.7	5.3
Flowers	Purple Daybreak vs Red Diamond	95.3	4.7

4.2 Classification Performance Using Minimal-Band and Feature Sets



(a) (b)

Figure 4. Model performance results for minimal-band and full-spectrum classification.
(a) Confusion matrix for the Random Forest classifier trained on the 15-band minimal wavelength set.
(b) Mean 5-fold cross-validation accuracy across SVM, Random Forest, and XGBoost using the full spectrum, 15-band subset, and engineered spectral features.

The classification results for the reduced and engineered feature sets are presented in Figure 4. The Random Forest classifier trained on the selected 15-band subset (Figure 4(a)) achieved high classification accuracy, correctly identifying 11 of 14 Cerise Charm samples, 13 of 15 Purple Daybreak samples, and 10 of 12 Red Diamond samples. Most classification errors occurred between Cerise Charm and Red Diamond, whereas Purple Daybreak showed the highest class-level accuracy. Figure 4(b) compares the mean 5-fold cross-validation accuracies of SVM, Random Forest, and XGBoost under different feature configurations. Using the full 300-band spectra, Random Forest and XGBoost achieved accuracies of approximately 0.80-0.83. Performance remained strong when using only the 15-band subset (≈0.73-0.77), indicating that the selected wavelengths retained most of the relevant discriminatory information. The highest accuracy was obtained using engineered features derived from vegetation indices and regional reflectance means, with XGBoost exceeding 0.90 accuracy. These results demonstrate that compact, physiologically meaningful feature sets can effectively distinguish the cultivars while reducing data dimensionality and computational requirements.

4.3 Spectral Band Importance Analysis

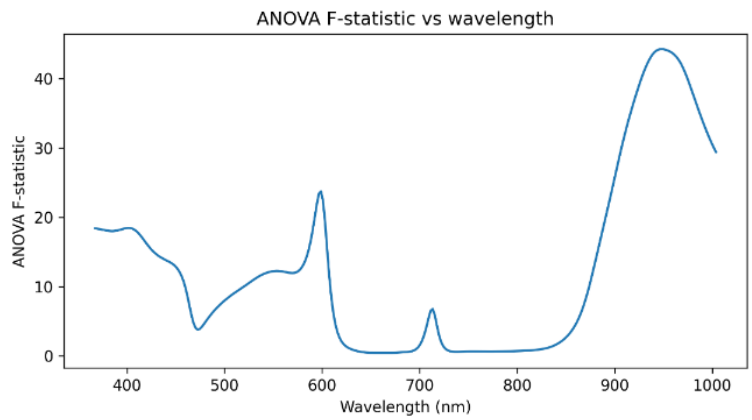


Figure 5. Band-selection curves across the VNIR hyperspectral range. ANOVA F-statistics identifying wavelengths with significant cultivar variation.

To identify the most informative wavelengths for cultivar discrimination, four complementary band-selection methods were evaluated: ANOVA F-statistics, Pearson correlation, Mutual Information (MI), and Random Forest (RF) feature importance (Figures 5-8). These methods revealed that the visible, red-edge, and near-infrared (NIR) regions contain the most discriminative spectral information. The ANOVA results (Figure 5) showed moderate separation in the blue-green region (400–550 nm), a distinct peak near 600 nm associated with chlorophyll absorption, and strong responses in the red-edge (705-715 nm) and NIR regions (900-960 nm). These findings suggest that both pigment-related and structural characteristics contribute to cultivar differentiation. Pearson correlation analysis (Figure 6) identified important wavelengths near 460-480 nm, 600 nm, and around 710 nm, corresponding to anthocyanin absorption, chlorophyll variability, and red-edge shifts, respectively. Correlation values also increased in the NIR region, further highlighting structural differences among cultivars. The MI analysis (Figure 7) captured additional nonlinear relationships, with notable peaks in the blue region (450-480 nm), red-edge region (700-720 nm), and early NIR wavelengths (770-800 nm). These results indicate that cultivar-specific information is distributed across multiple spectral regions and includes nonlinear spectral interactions. RF feature importance (Figure 8) emphasized wavelengths in the blue (448-470 nm), red (655-670 nm), and red-edge (707-711 nm) regions as the most influential for classification. While NIR wavelengths also contributed, their importance was generally lower than that of pigment-sensitive visible and red-edge bands.

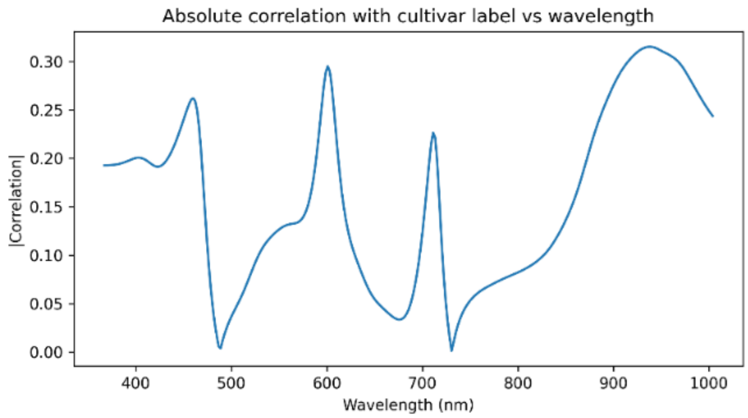


Figure 6. Absolute Pearson correlation between reflectance and cultivar label.

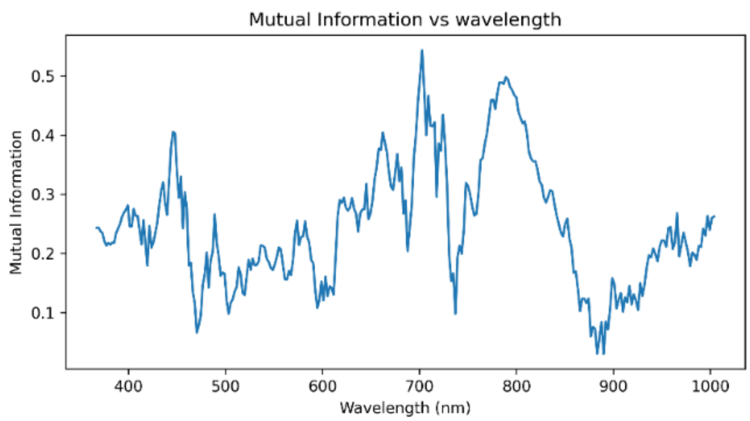


Figure 7. Mutual Information showing nonlinear wavelength relevance.

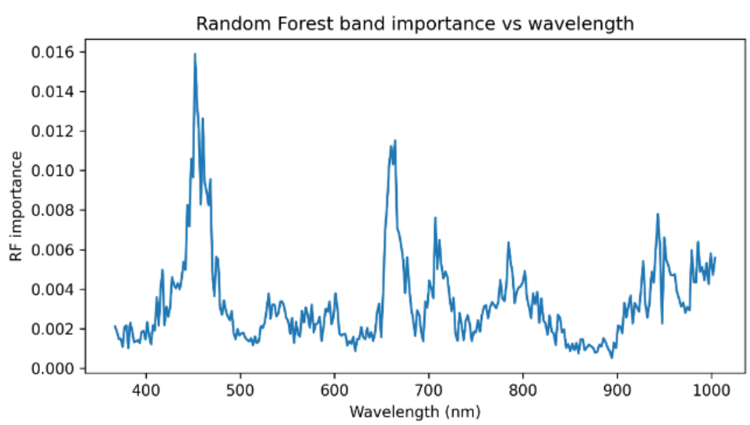


Figure 8. Random Forest feature importance reflecting model-driven discriminative power.

4.4 Derivative-Based Spectral Analysis of Leaves and Flowers

The Derivative spectra were used to enhance subtle spectral differences that were not readily visible in the original reflectance data. The first-derivative flower spectra (Figure 9) exhibited similar overall patterns across cultivars, with minor differences in the magnitude of the positive peak and negative lobe near the red-edge region. Cerise Charm showed a slightly stronger slope transition, indicating greater changes in reflectance across pigment-sensitive wavelengths. Greater cultivar separation was observed in the first-derivative leaf spectra (Figure 10), particularly within the green (500-600 nm) and red-edge (~700 nm) regions. Differences in peak height and width suggest variations in chlorophyll concentration and leaf structural properties among cultivars. The second-derivative flower spectra (Figure 11) revealed subtle shifts in curvature around the red-edge region, highlighting small cultivar-dependent differences in floral pigment absorption. In contrast, the second-derivative leaf spectra (Figure 12) displayed more pronounced separation, with noticeable differences in the amplitude and sharpness of the red-edge peak and subsequent trough. These features indicate variation in both pigment composition and internal leaf structure.

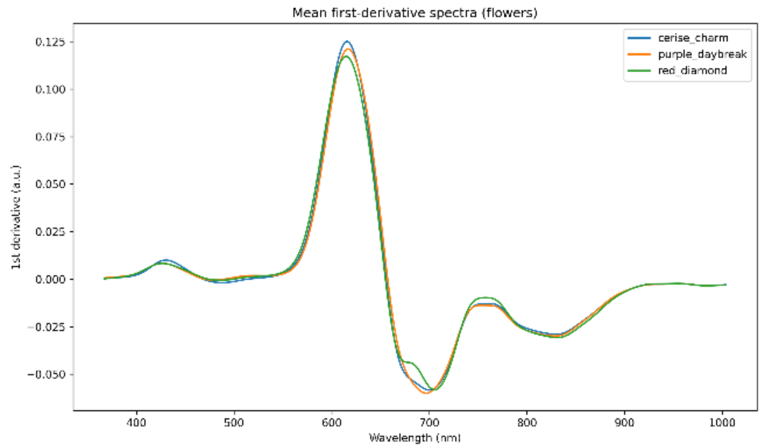


Figure 9. Mean first-derivative spectra of flowers for the three cultivars.

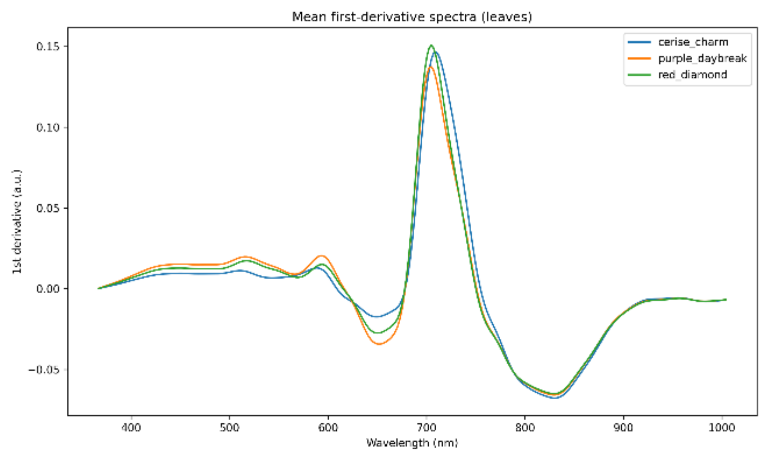


Figure 10. Mean first-derivative spectra of leaves for the three cultivars.

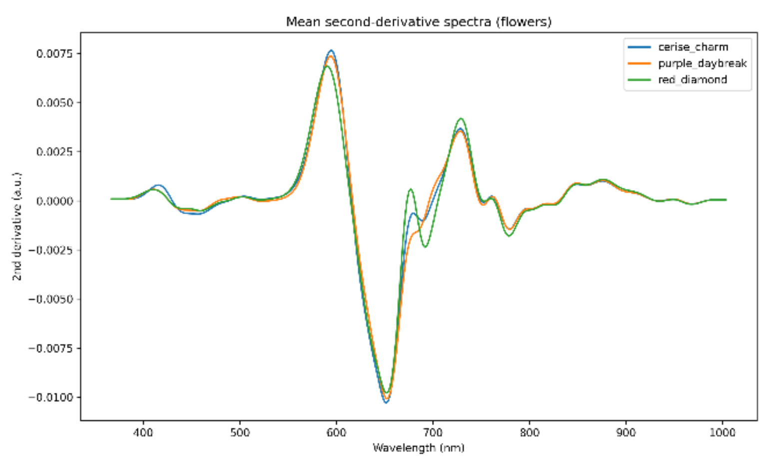


Figure 11. Mean second-derivative spectra of flowers.

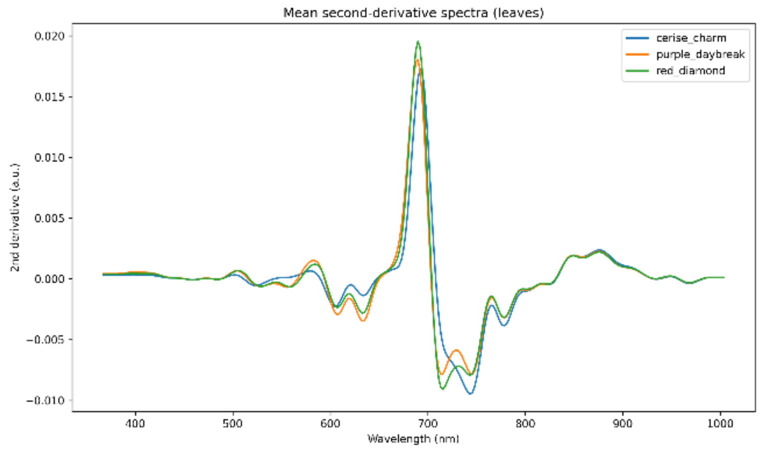


Figure 12. Mean second-derivative spectra of leaves.

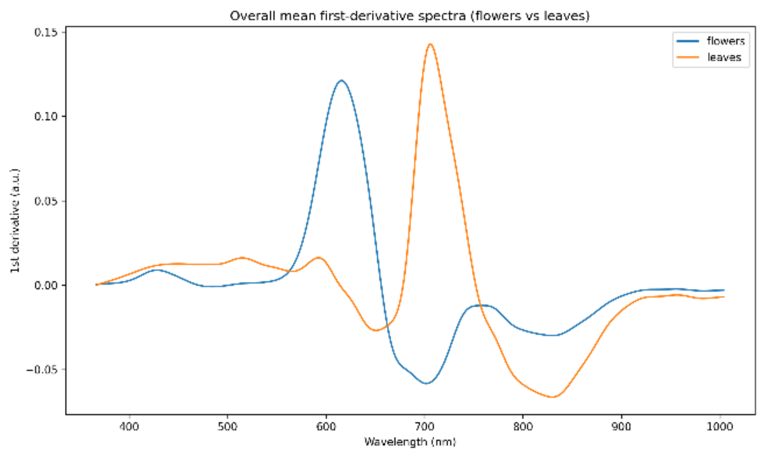


Figure 13. Overall mean first-derivative spectra comparing flowers and leaves (pooled across cultivars).

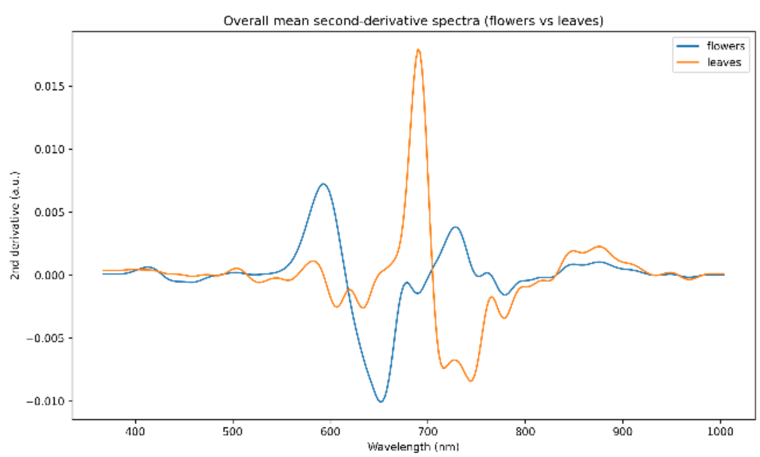


Figure 14. Overall mean second-derivative spectra of flowers versus leaves.

Organ-level comparisons further emphasized these distinctions. The mean first-derivative spectra (Figure 13) showed that leaves possessed a stronger red-edge slope and deeper decline into the NIR region than flowers, reflecting higher chlorophyll content and greater internal scattering. Similarly, the mean second-derivative spectra (Figure 14) exhibited a sharp, narrow red-edge feature in leaves compared to the broader and lower-amplitude response observed in flowers. The derivative analysis enhanced both cultivar- and organ-level spectral differences, particularly in the visible and red-edge regions. These results support the wavelength selection findings and help explain the strong classification performance obtained using red-edge and pigment-sensitive spectral features.

4.5 Principal Component Analysis (PCA) of Hyperspectral Signatures

PCA was used to examine spectral variation among cultivars and between plant organs. The PCA plot of all samples (Figure 15) shows clear separation between leaves and flowers along PC1, indicating substantial differences in their spectral characteristics. Leaves clustered at negative PC1 values, whereas flowers clustered at positive PC1 values, reflecting differences in pigment composition, red-edge behavior, and near-infrared reflectance. Although some overlap among cultivars was observed within each organ group, separation along PC2 suggests the presence of cultivar-specific variation. The PCA of flower spectra alone (Figure 16) revealed relatively compact clustering with modest cultivar separation. Purple Daybreak tended to occupy higher PC2 values, while Cerise Charm and Red Diamond showed slight shifts along PC1. These differences likely reflect variations in floral pigment composition and concentration. In contrast, the PCA of leaf spectra (Figure 17) showed stronger cultivar discrimination. Cerise Charm, Purple Daybreak, and Red Diamond formed more distinct clusters within the PC1-PC2 space, indicating greater spectral variability among leaf samples. This separation is consistent with differences in chlorophyll content, anthocyanin accumulation, leaf structure, and red-edge characteristics.

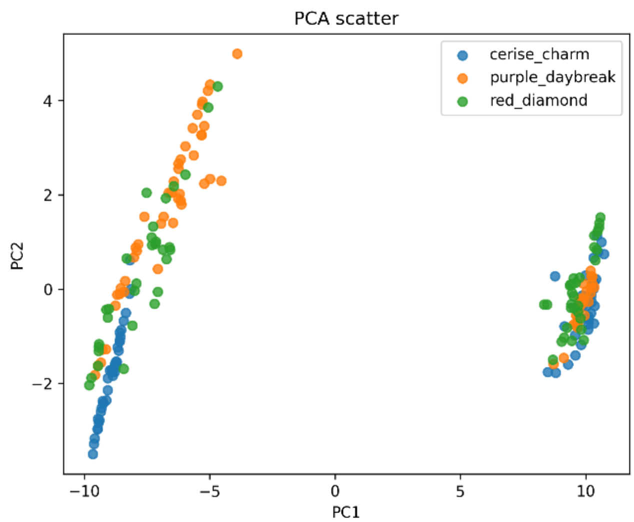


Figure 15. PCA of all flower and leaf samples showing clear organ-level separation along PC1.

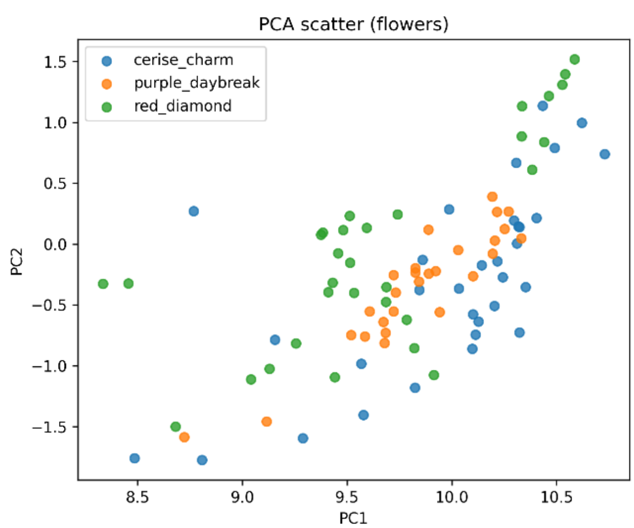


Figure 16. PCA of flower samples only, showing subtle cultivar-specific shifts within a compact cluster.

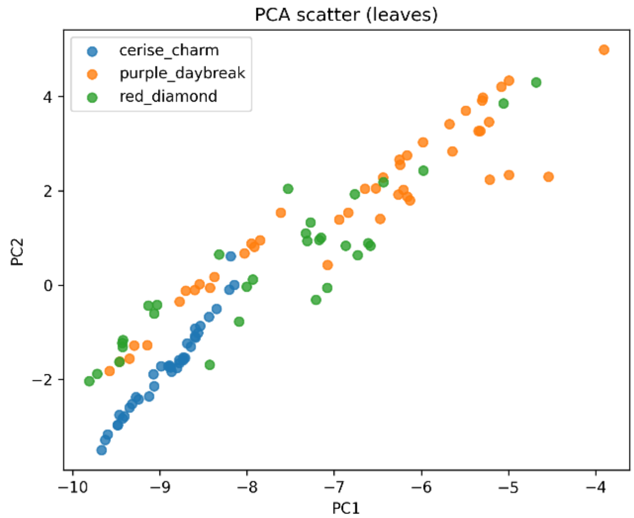


Figure 17. PCA of leaf samples only, revealing strong cultivar clustering driven by pigment and structural differences.

4.6 Mean Spectral Reflectance of Flowers and Leaves

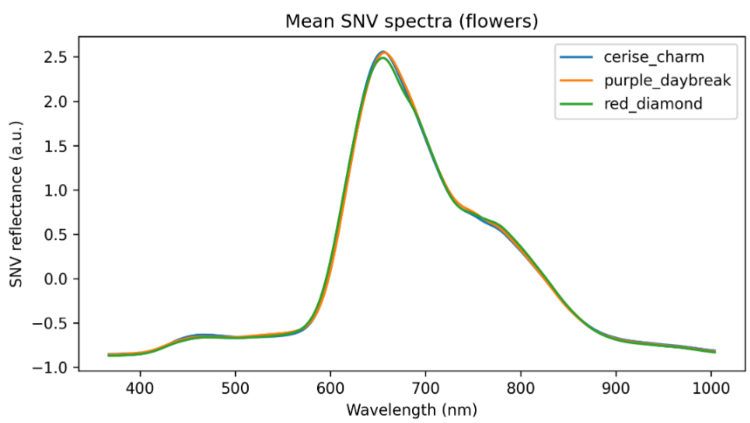


Figure 18. Mean SNV-normalized reflectance spectra of flowers.

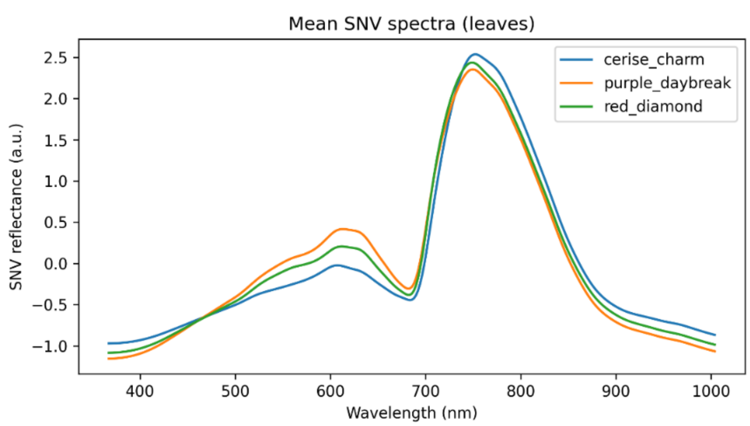


Figure 19. Mean SNV-normalized reflectance spectra of leaves.

The mean SNV-normalized reflectance spectra of the three *Loropetalum chinense* cultivars are shown in Figures 18 and 19. Flower spectra (Figure 18) exhibited similar overall shapes, with subtle differences primarily in the visible region (400-700 nm). All cultivars showed low reflectance in the blue and red regions due to pigment absorption, while reflectance remained relatively low throughout the NIR region. Purple Daybreak flowers displayed slightly higher visible reflectance, whereas Cerise Charm and Red Diamond exhibited more similar spectral profiles. Leaf spectra (Figure 19) displayed the typical vegetation signature, characterized by strong absorption in the blue and red regions, a pronounced red-edge transition near 700-750 nm, and high reflectance in the NIR region. Cultivar-level differences were more evident in leaves than in flowers. Red Diamond exhibited the strongest visible-region absorption and lowest reflectance, while Purple Daybreak showed relatively higher green reflectance and a slightly shifted red-edge position. Cerise Charm displayed intermediate spectral characteristics. The stronger separation observed among leaf spectra indicates that leaf tissues provide more robust cultivar-specific information than flowers. Differences in chlorophyll concentration, anthocyanin content, and internal leaf structure contribute to these spectral variations. These findings are consistent with the PCA, derivative analysis, and band-selection results, which collectively identify leaf and red-edge features as the most informative components for cultivar discrimination.

4.7 Integration Pathways into Multimodal Language Model Architectures

The spectral outputs produced across the preceding sections share a structural property that sets them apart from raw hypercube data: they are simultaneously compact and biochemically interpretable. The consensus 15-band wavelength subset, vegetation-index feature vectors, and organ-resolved SAM dissimilarity matrices encode cultivar-discriminative information in a form that requires no full-tensor storage and imposes minimal computational overhead. This combination numerical compactness paired with physiologically grounded semantics is precisely what makes these representations compatible with the emerging class of multimodal reasoning systems built around large language models. Current LLM architecture designed for tool-augmented or retrieval-assisted inference operate increasingly over heterogeneous input streams, pairing visual observations with structured numerical context to reach natural-language conclusions. In horticultural practice, this translates directly: a system receiving the anthocyanin reflectance index values and red-edge chlorophyll estimates derived here, alongside co-registered UAV canopy imagery of *Loropetalum* stands, could produce cultivar-specific phenotypic assessments without requiring the original 300-band hypercube per plant. The language model would not be processing spectra it would be reasoning over a structured biochemical vocabulary, one whose terms map onto knowledge already encoded in plant physiology and precision agriculture literature. Another important aspect of resource-limited devices is scalability. The results shown here indicate that the 15-band subset of data has been classified with an overall accuracy of 73 to 77%. Therefore, it is possible to use a smaller multispectral sensor rather than requiring a complete hyperspectral instrument to provide sufficient data for an edge-based LLM inference module. The ability to combine lightweight quantized models, which are now commonly utilized in the small language model (SLM) field, provides a practical path to achieve real-time and explainable cultivar identification within field nursery applications. The approach to spectral phenotyping developed in this study should therefore be viewed as an open framework, it is structured as a sensing front-end that was specifically developed to provide outputs that interface with the next generation of plant intelligent systems based on language.

5. Discussion and limitations

The spectral differences among the three *Loropetalum chinense* cultivars were consistent with their known pigment and structural traits. Cerise Charm showed higher visible reflectance and a sharper red-edge transition, while Purple Daybreak and Red Diamond showed lower visible reflectance, suggesting stronger anthocyanin-related absorption. Differences in near-infrared reflectance also indicated variation in leaf internal structure. These trends were supported by derivative spectra, vegetation indices, and band-selection results. The most informative wavelengths were concentrated in physiologically meaningful regions linked to anthocyanin absorption, chlorophyll absorption, the red-edge, and near-infrared scattering. This shows that a small set of bands can preserve most of the discriminatory information and supports the feasibility of compact reduced-band sensors for ornamental phenotyping. Engineered spectral features performed better than raw spectra, indicating that physiologically informed feature design improves classification, especially when the dataset is limited. The overall performance was comparable to that reported for other horticultural crops, showing that ornamental shrubs can also be identified reliably using hyperspectral methods. The combined use of derivative analysis, vegetation indices, and statistically selected wavelengths provides both strong performance and biological interpretability. Several limitations should be noted. The study was conducted under controlled greenhouse and laboratory conditions, so the results may not directly transfer to field environments. The dataset was relatively small and included only three cultivars, which limits generalization to broader genetic diversity. Cross-season, cross-location, and cross-sensor validation will be needed before practical deployment. Environmental variation such as lighting, temperature, humidity, and nutrient status may also influence spectral signatures and should be considered in future work.

6. Conclusion

This study developed a hyperspectral machine-learning pipeline for the non-destructive identification of three *Loropetalum chinense* cultivars using reflectance data from 367-1004 nm. The results showed that leaf spectra provided stronger cultivar discrimination than flower spectra, with key differences concentrated in the visible, red-edge, and near-infrared regions. Multi-method band selection identified a compact 15-band subset that preserved most of the classification performance of the full 300-band dataset. Among the evaluated models, Random Forest and XGBoost achieved strong performance, while engineered spectral features derived from vegetation indices and regional reflectance means produced the highest accuracy. These findings demonstrate that cultivar discrimination can be achieved using a small number of physiologically meaningful wavelengths, providing a foundation for the development of cost-effective multispectral sensing systems for ornamental plant phenotyping and cultivar verification. Compact feature representations may also facilitate future integration with multimodal AI and language-model-based decision support systems.

References

Aishwarya, G., Kumar, B. L. N. P., Syamala, D., & Sreya, N. S. (2023). Dimensionality reduction technique for hyperspectral remote sensing image classification. 2023 8th International Conference on Communication and Electronics Systems (ICCES), 1694-1698, doi: https://doi.org/10.1109/ICCES57224.2023.10192744

Amankwah, A. (2015). Spatial mutual information based hyperspectral band selection for classification. The Scientific World Journal, 2015, Article 630918, doi: https://doi.org/10.1155/2015/630918

Angel, Y., Raiho, A., Kathuria, D., Chadwick, K. D., Brodrick, P. G., Lang, E., Ochoa, F., & Shiklomanov, A. N. (2025). Deciphering the spectra of flowers to map landscape-scale blooming dynamics. Ecosphere, 16(2), e70127, doi: https://doi.org/10.1002/ecs2.70127

Behmann, J., Acebron, K., Emin, D., Bennertz, S., Matsubara, S., Thomas, S., & Rascher, U. (2018). Specim IQ: Evaluation of a new, miniaturized handheld hyperspectral camera for plant phenotyping and disease detection. Sensors, 18, 441, doi: https://doi.org/10.3390/s18020441

Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32, doi: https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794, doi: https://doi.org/10.1145/2939672.2939785

de Castro, A. I., Shi, Y., Maja, J. M., & Peña, J. M. (2021). UAVs for vegetation monitoring: Overview and recent scientific contributions. Remote Sensing, 13, 2139, doi: https://doi.org/10.3390/rs13112139

Duarte-Carvajalino, J., Silva-Arero, E., Góez-Vinasco, G., Torres-Delgado, L., Ocampo-Paez, O., & Castaño-Marín, Á. (2021). Estimation of water stress in potato plants using hyperspectral imagery and machine-learning algorithms. Horticulturae, 7, 176, doi: https://doi.org/10.3390/horticulturae7070176

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. Analytica Chimica Acta, 185, 1-17, doi: https://doi.org/10.1016/0003-2670(86)80028-9

Gill, T., Gill, S.K., Saini, D.K. et al. (2022). A Comprehensive Review of High Throughput Phenotyping and Machine Learning for Plant Stress Phenotyping. Phenomics 2(3), 156–183, doi: https://doi.org/10.1007/s43657-022-00048-z

Goetz, A. F. H., et al. (1985). Imaging spectrometry for Earth remote sensing. Science, 228, 1147-1153, doi: https://doi.org/10.1126/science.228.4704.1147

Gutiérrez, S., Fernández-Novales, J., Diago, M. P., & Tardaguila, J. (2018). On-the-go hyperspectral imaging for classifying grapevine varieties. Frontiers in Plant Science, 9, 1102.

Li, Y., Zhang, H., & Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sensing, 9(1), 67, doi: 10.3390/rs9010067.

Liu, H., Yu, T., Hu, B., Hou, X., Zhang, Z., Liu, X., & Bao, Q. (2021). UAV-Borne Hyperspectral Imaging Remote Sensing System Based on Acousto-Optic Tunable Filter for Water Quality Monitoring, Remote Sensing, 13, 4069, doi: https://doi.org/10.3390/rs13204069

Liu, K.-H., Yang, M.-H., Huang, S.-T., & Lin, C. (2022). Plant Species Classification Based on Hyperspectral Imaging via a Lightweight Convolutional Neural Network Model. Frontiers in Plant Science, 13, 855660, doi: https://doi.org/10.3389/fpls.2022.855660

Lowe, A., Harrison, N., & French, A. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. Plant Methods, 13, 80, doi: 10.1186/s13007-017-0233-z

Lu, Y., Young, S., Linder, E., Whipker, B., & Suchoff, D. (2022). Hyperspectral Imaging With Machine Learning to Differentiate Cultivars, Growth Stages, Flowers, and Leaves of Industrial Hemp (Cannabis sativa L.). Frontiers in Plant Science, 12, 810113, doi: https://doi.org/10.3389/fpls.2021.810113

Manjunatha, H., Sundaravadivel, P., Borah, S. S., Tamil, L., Knight, P. R., Torbert, H. A., & Kumpatla, S. P. (2026). Beyond Green: Domain-Aware Self-Validated Instance Counting for *Loropetalum* and Non-Green Ornamental Species Using Annotation-Free Robustness Metrics. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 9777–9786.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419, doi: 10.3389/fpls.2016.01419.

Nagasubramanian, K., Jones, S., Singh, A., Sarkar, S., Singh, A. & Ganapathysubramanian, B. (2019). Plant disease identification using explainable 3D deep learning on hyperspectral images. Plant Methods, 15, 98, doi: https://doi.org/10.1186/s13007-019-0479-8

Osco, L., Ramos, A., Moriya, É., et al. (2019). Modeling Hyperspectral Response of Water-Stress Induced Lettuce Plants Using Artificial Neural Networks, Remote Sensing, 11, 2797, doi: https://doi.org/10.3390/rs11232797

Pal, M., & Foody, G. M. (2010). Feature selection for classification of hyperspectral data by SVM. IEEE Transactions on Geoscience and Remote Sensing, 48(5), 2297-2307, doi: https://doi.org/10.1109/TGRS.2009.2039484

Patiluna, V., Owen Jr., J., Maja, J. M., et al. (2025). Using Hyperspectral Imaging and Principal Component Analysis to Detect and Monitor Water Stress in Ornamental Plants. Remote Sensing, 17, 285, doi: https://doi.org/10.3390/rs17020285

Slaton, M. R., Hunt, E. R., & Smith, W. K. (2001). Estimating near-infrared leaf reflectance from leaf structural characteristics. American Journal of Botany, 88, 278-284, doi: https://doi.org/10.2307/2657019

Sun, J., Zhou, X., Hu, Y. G., Wu, X. H., Zhang, X. D., & Wang, P. (2019). Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging. Computers and Electronics in Agriculture, 160, 153-159, doi: https://doi.org/10.1016/j.compag.2019.03.004

Susič, N., Žibrat, U., Širca, S., Strajnar, P., Razinger, J., Knapič, M., Vončina, A., Urek, G., & Gerič Stare, B. (2018). Discrimination between abiotic and biotic drought stress in tomatoes using hyperspectral imaging. Sensors and Actuators B: Chemical, 273, 842-852, doi: https://doi.org/10.1016/j.snb.2018.06.121

Sytar, O., Brestič, M., Živčák, M., et al. (2017). Applying hyperspectral imaging to explore natural plant diversity towards improving salt stress tolerance. Science of the Total Environment, 578, 90-99, doi: https://doi.org/10.1016/j.scitotenv.2016.08.014

Tang, T., Luo, Q., Yang, L., Gao, C., Ling, C., & Wu, W. (2024). Research review on quality detection of fresh tea leaves based on spectral technology. *Foods*, 13(1), 25. doi:10.3390/foods13010025.

Virnodkar, S., Pachghare, V., Patil, V., & Jha, S. (2020). Remote sensing and machine learning for crop water stress determination in various crops: a critical review. Precision Agriculture, 21, 1121-1155, doi: https://doi.org/10.1007/s11119-020-09711-9

Vairavan, C., Kamble, B. M., Durgude, A. G., Ingle, S. R., & Pugazenthi, K. (2024). Hyperspectral Imaging of Soil and Crop: A Review. *Journal of Experimental Agriculture International*. https://doi.org/10.9734/jeai/2024/v46i12290

Wong, C. Y., Gilbert, M. E., Pierce, M. A., Parker, T. A., Palkovic, A., Gepts, P., Magney, T. S., & Buckley, T. N. (2023). Hyperspectral remote sensing for phenotyping the physiological drought response of common and tepary bean. *Plant Phenomics*, 5, 0021. doi:10.34133/plantphenomics.0021