

HYPERSENSPECTRAL SPECTRAL LIBRARIES FOR CROP GENOTYPE AND VARIETY DISCRIMINATION: LINKING SPECTRAL PHENOTYPING WITH GENETIC DIVERSITY

Devasri Moovendran¹, Pazhanivelan Sellaperumal^{2*}, Ragunath Kaliaperumal³, Vanitha G⁴, Kalpana R⁵

¹Department of Remote Sensing and Geographic Information System, Tamil Nadu Agricultural University, Coimbatore – 641003, India. (devasri.pgrsg2024@tnau.ac.in)

²Directorate of Crop Management, Tamil Nadu Agricultural University, Coimbatore – 641003, India. (pazhanivelans@gmail.com)

³Centre for Water and Geospatial Studies, Tamil Nadu Agricultural University, Coimbatore – 641003, India. (ragunathkp@tnau.ac.in)

⁴School of Post Graduate Studies, Tamil Nadu Agricultural University, Coimbatore – 641003, India. (vanithag@tnau.ac.in)

⁵Department of Agronomy, Tamil Nadu Agricultural University, Coimbatore – 641003, India. (kalpana@tnau.ac.in)

*Corresponding author: Pazhanivelan S (pazhanivelans@gmail.com)

ABSTRACT

Crop breeding depends on characterizing genetic diversity, but phenotyping remains a bottleneck relative to genotyping, motivating interest in hyperspectral imaging, which captures continuous reflectance curves shaped by leaf pigments, water, nitrogen and structure, together forming a "spectral phenotype." Genotype-oriented spectral libraries linking this spectral phenotype rigorously to genetic diversity remain scarce, and remote sensing and genetics literatures have developed largely in isolation. This review addresses that gap by synthesizing evidence on the biological pathway from genotype to spectrum, spectral library design and metadata standards, preprocessing and feature-selection methods, machine learning and validation practices, and cross-scale sensing from field to UAV to satellite, drawing on studies across rice, wheat, maize, soybean, cotton and other crops. Findings show genotype and variety discrimination is achievable, particularly at leaf and seed level under controlled conditions, but reported accuracies are frequently confounded by growth stage and genotype-by-environment effects and inflated by scan-level data leakage rather than genuine genetic signal. Discriminating wavelengths cluster in red-edge and pigment-sensitive visible regions for canopy studies and extend into shortwave infrared for composition-driven seed studies, while standardized, multi-environment libraries with grouped validation remain rare and genomic integration is concentrated in few groups and crops. A standardized, metadata-rich library framework coupling spectral phenotypes with genomic data could enable cost-effective germplasm screening, genomic prediction and improved breeding decisions, provided rigorous external validation and explicit biological interpretation accompany future work before spectral discrimination is trusted as a genetic proxy.

KEYWORDS: Crop breeding; Genetic diversity; Hyperspectral imaging; Plant phenotyping; Spectral libraries; Spectral phenotype

INTRODUCTION:

Feeding a growing population depends on how well breeders can find, keep and use variation within crop species. That variation sits in genebanks and breeding nurseries as thousands of accessions and lines, and only a fraction of it is ever characterised in detail. The bottleneck is rarely genotyping any more. Sequencing costs have fallen far enough that a programme can score tens of thousands of markers on a routine basis, while the measurement of plant traits in the field still leans on hand scoring, destructive sampling and a small number of plots per season. Araus and Cairns (2014) described this imbalance as the defining constraint on modern breeding, and the gap has widened since.

High-throughput phenotyping was developed to close it. Sensors mounted on carts, gantries, aircraft and satellites now record crop condition repeatedly and without contact (Fahlgren et al., 2015; Araus et al., 2018). The premise is straightforward, if the link between genotype and phenotype is to be resolved at scale, phenotype measurement has to become as cheap and repeatable as genotyping already is (Sozzani & Benfey, 2011). Among these, imaging spectroscopy is distinctive. A hyperspectral instrument samples reflected radiation in many narrow, contiguous bands, so it records not a few broad colours but the shape of the reflectance curve itself. That shape responds to leaf pigments, water, nitrogen and protein, cell wall constituents, leaf internal structure and the arrangement of leaves in the canopy (Jacquemoud & Baret, 1990; Ustin et al., 2009). A hyperspectral measurement is therefore an indirect, integrated record of the plant's biochemical and structural condition at the moment of measurement. We call this record the spectral phenotype.

If spectral phenotypes are to say anything useful about genetic diversity, single experiments are not enough. Reflectance depends on the sensor, the illumination, the growth stage and the growing environment as much as on the plant's genetic background, so a measurement made in one trial cannot be compared with another unless both were collected and documented in a consistent way. This is the argument for spectral libraries: organised, calibrated, richly described collections of spectra, each tied to a known genotype, plot, stage and environment. Reference libraries of this kind are well established for minerals and land cover (Baldrige et al., 2009; Meerdink et al., 2019), and crop-level versions have been built at cultivar resolution (Rao et al., 2007), but genotype-oriented libraries designed for genetic work remain scarce.

The gap this review addresses is the space between two literatures that rarely meet. Remote sensing papers report that varieties can be separated spectrally, often with high accuracy, but seldom ask what the discriminating wavelengths mean biologically or whether the separation would survive a new season. Genetics papers increasingly use spectral data as a phenotype, yet treat the sensor as a black box. Our central question is how hyperspectral spectral libraries can bridge the two: how an observable spectral phenotype can be connected, carefully and with appropriate caution, to the genetic diversity beneath it.

The review has three objectives,

- To set out the biological pathway from genotype to spectral phenotype, define the terms precisely, and describe what a genotype-oriented spectral library requires in design, acquisition and metadata.
- To synthesise the evidence on hyperspectral genotype and variety discrimination across crops, giving particular attention to growth stage and genotype-by-environment effects, and to assess the analytical practice that produced it, including preprocessing, band selection, machine learning and validation.
- To examine scaling from field to satellite and integration with genomics, and to identify the gaps that current evidence actually supports.

From Genotype to Spectral Phenotype:

The link between a genotype and its spectrum is real but indirect, and it helps to follow it one step at a time. Alleles differ; those differences alter gene expression; expression shapes the concentrations of pigments, water, nitrogen compounds, proteins, cellulose and lignin in leaf tissue, and shapes leaf thickness, mesophyll arrangement and cuticle properties, with downstream effects on canopy architecture. Each of these influences how light is absorbed, transmitted and scattered. What a spectrometer records **is** the end of that chain, not the beginning. Chen et al. (2025) recently showed that leaf hyperspectral reflectance can be related to patterns of gene expression, which supports the logic of the chain while also underlining how many steps lie between a nucleotide and a reflectance value.

Radiative transfer modelling makes the middle of the chain explicit. The PROSPECT family of models reproduces leaf reflectance and transmittance from a small set of parameters describing pigment content, water, dry matter and internal structure (Jacquemoud & Baret, 1990; Jacquemoud et al., 1996), and later work has examined how variable the structural parameter itself is across species and conditions (Boren et al., 2019; Yu et al., 2026). Inversion of such models has been used to retrieve leaf traits in crops (Kong et al., 2025), and radiative transfer descriptions have been embedded in functional-structural plant models so that canopy-level spectra follow from simulated plant architecture (Coussement et al., 2018). The practical message is that the spectrum encodes traits, and genotype acts on the spectrum through those traits.

Terminology matters here, because loose usage has caused real confusion in the literature. A genotype is the genetic constitution of an individual or line. A phenotype is a measurable characteristic, such as chlorophyll concentration or specific leaf area. A spectral phenotype is the reflectance signature measured under defined conditions, treated as a trait in its own right. A spectral biomarker is a spectral feature that has been shown to track a specific biological quantity. None of these is a genetic marker. A wavelength that separates two varieties in a classifier is a statistical discriminator; it becomes biologically meaningful only when tied to a measured trait, and genetically meaningful only when tied to genomic variation through explicit analysis. Keeping these categories separate is the single most important discipline in this field.

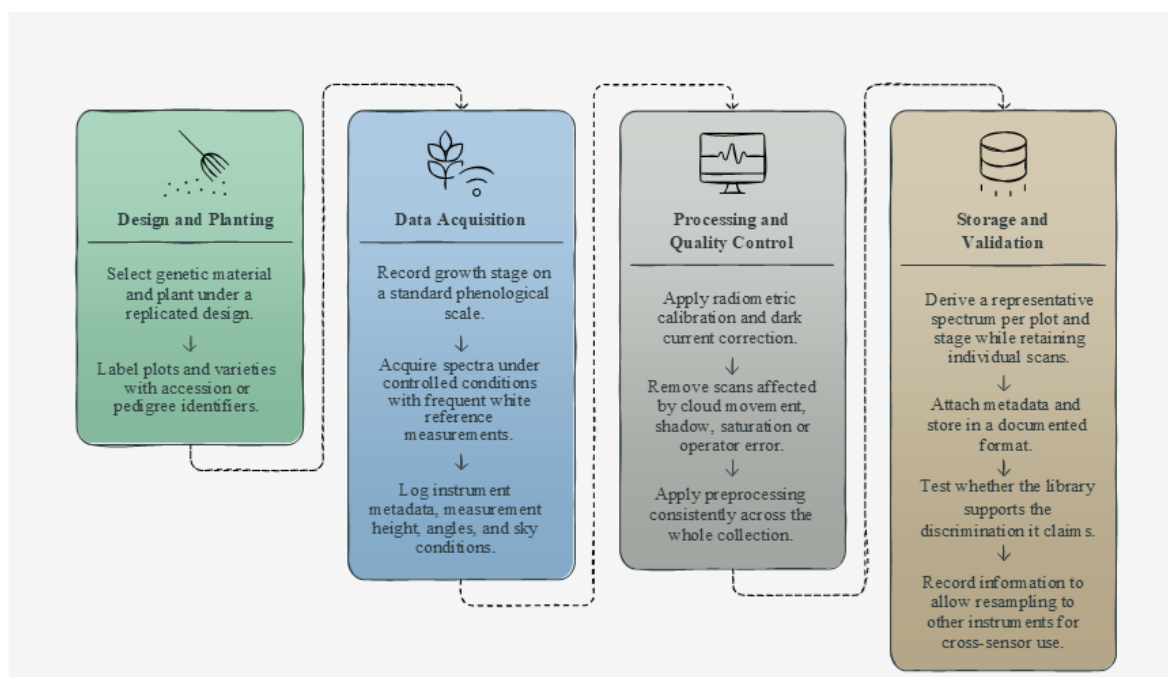


Fig. 1. Conceptual pathway from genotype to spectral phenotype through gene expression, biochemical traits, structural traits, and light interaction.

Fundamentals of Crop Spectral Signatures:

Vegetation reflectance across the optical range has a characteristic shape, and each part of it has a physical cause. In the visible region, chlorophyll absorbs strongly in the blue and red, with carotenoids contributing in the blue, which produces the familiar green peak near 550 nm. Absorption weakens sharply between roughly 680 and 750 nm, and reflectance rises steeply. This transition is the red edge. Its steepness and its position along the wavelength axis both shift with chlorophyll concentration and leaf structure, which is why the red edge has been studied for so long as a diagnostic feature (Horler et al., 1983; Curran et al., 1990; Gitelson et al., 1996). Beyond it, the near infrared plateau is governed mainly by scattering at cell wall and air space boundaries, so it responds to leaf internal structure, leaf layering and canopy density rather than to pigments. Further into the shortwave infrared, water dominates, with strong absorption near 1450 and 1940 nm, and weaker features attributable to dry matter constituents such as cellulose, lignin, starch and protein.

Several derived quantities are used to summarise these features. Red edge position, usually estimated from the maximum of the first derivative or by fitting techniques, condenses the transition into a single number (Baranowski & Rokne, 2005; Ju et al., 2010), and wavelet-based estimators have been proposed to make the estimate more stable in canopy data (Li et al., 2017). Derivative spectra sharpen absorption shoulders and suppress broad baseline effects. Continuum removal isolates the depth and width of individual absorption features. Indices of this kind can also be derived automatically rather than chosen by hand, by searching wavelength combinations under a hyperparameter optimisation framework (Koh et al., 2022). Narrowband vegetation indices combine a small number of wavelengths into ratios that track pigments, water or structure, and hyperspectral versions of these indices were shown early on to relate more closely to crop biophysical variables than their broadband equivalents (Thenkabail et al., 2000). Canopy-level derivative features have also been used to detect fluorescence-related effects near the red edge (Zarco-Tejada et al., 2003).

For genotype work, the relevant point is that these regions carry different kinds of information and are not equally useful. Pigment-driven regions respond quickly to nutrition, stress and development, so they can be highly discriminating within a single measurement date and highly unstable across dates. Structural regions in the near infrared may reflect more persistent architectural differences between varieties, but they are also strongly affected by canopy closure and viewing geometry. Shortwave infrared features tied to dry matter are attractive because they relate to composition, but they demand instruments with a wide spectral range and good signal quality. Table 1 summarises the main regions, the constituents that drive them, and how cautiously each can be linked to genetic differences.

Table 1. Spectral regions, their dominant absorbers and how cautiously each can be linked to genetic differences.

Region	Wavelength range (nm)	Dominant absorber or driver	Biological significance	Relevance to genotype work	References
Blue	400-500	Chlorophyll a/b and carotenoids	Strong pigment absorption; sensitive to	Useful but saturates quickly at high pigment	Sims & Gamon (2002); Ustin et al. (2009)

			pigment pool size and stress	content; strongly stage dependent	
Green peak	500-600	Reduced pigment absorption	Reflectance maximum near 550 nm; tracks chlorophyll concentration	Discriminates varieties differing in leaf greenness, but confounded with nitrogen supply	Gitelson et al. (1996); Sims & Gamon (2002)
Red	600-680	Chlorophyll a	Deep absorption trough; saturates at moderate chlorophyll levels	Limited standalone discriminating power in closed canopies	Sims & Gamon (2002); Thenkabail et al. (2000)
Red edge	680-750	Transition from pigment absorption to structural scattering	Position and slope shift with chlorophyll content and leaf structure	Most frequently reported discriminating region; requires stage control	Horler et al. (1983); Curran et al. (1990); Gitelson et al. (1996); Li et al. (2017)
Near infrared plateau	750-1300	Scattering at cell wall / air space interfaces	Responds to leaf internal structure, leaf layering, canopy density	Can capture persistent architectural differences; strongly affected by canopy closure and geometry	Jacquemoud & Baret (1990); Boren et al. (2019)
Water absorption features	1400-1500; 1900-2000	Leaf and canopy water	Depth tracks water content and relative water status	Genotype differences in water relations may be visible, but environment dominates	Jacquemoud et al. (1996); Ustin et al. (2009)
Shortwave infrared dry-matter features	2000-2500	Cellulose, lignin, starch, protein	Composition-related absorption; weak in fresh leaves	Promising for compositional varietal differences; needs wide-range, high-quality instruments	Jacquemoud et al. (1996); Yu et al. (2026)

4. Hyperspectral Spectral Libraries and a Standardized Workflow

A spectral library is a curated set of reflectance measurements stored together with the information needed to interpret and reuse them. Established libraries for Earth surface materials, such as the ASTER and ECOSTRESS collections, set the pattern: consistent measurement protocols, documented instruments, and descriptions detailed enough that a user who was not present can judge whether a spectrum suits their purpose (Baldrige et al., 2009; Meerdink et al., 2019). Crop-oriented versions exist but are fewer. Rao et al. (2007) built an agricultural crop library and used it to classify crops at cultivar level, which remains one of the clearest demonstrations that library-based discrimination can reach below the species boundary. Library-matching approaches have also been examined more generally for material identification (Nidamanuri & Ramiya, 2013), and object-based libraries have been proposed for transferring knowledge into drone imagery (Chandra & Nidamanuri, 2024).

It is worth separating two kinds of library. A general library aims to cover many surface types at coarse taxonomic resolution and is optimised for breadth. A genotype-specific library aims to represent variation within a crop species, across named varieties, stages and environments, and is optimised for depth and internal consistency. The second is much harder to build, because the differences it must capture are small relative to the noise introduced by illumination, geometry and phenology.

Design decisions follow from that difficulty. Genetic material must be identified unambiguously, with variety names, accession or pedigree identifiers and seed source recorded. The field layout should be a proper experimental design with replication, so that genotype effects can be separated from position effects. Every spectrum needs a plot identity, not merely a plot-average value, because grouping information is essential later for honest validation. Growth stage should be recorded on a standard phenological scale at the time of measurement rather than inferred afterwards. Instrument make and model, spectral range, sampling interval and field of view should be logged, along with measurement height, viewing and solar angles, sky conditions, time,

and the white reference panel used. Radiometric calibration and dark current correction should be documented, and position recorded by GPS. Pacheco-Labrador and Martín (2015) show how sensor characteristics alone can shift retrieved reflectance, which is a reminder that instrument metadata is not administrative detail but part of the measurement. The plant phenotyping community has already formalised much of this thinking in minimum information standards for phenotypic data (Ćwiek-Kupczyńska et al., 2016), and those standards translate directly to spectral work. Table 2 lists the metadata we consider essential.

The build workflow follows a consistent order. Genetic material is selected and planted under a replicated design. Plots and varieties are labelled. Growth stage is recorded. Spectra are acquired under controlled conditions with frequent white reference measurements. Radiometric calibration is applied. Quality control removes scans affected by cloud movement, shadow, saturation or operator error. Preprocessing is applied consistently across the whole collection. A representative spectrum is derived per plot and stage, with the individual scans retained rather than discarded. Metadata is attached, and the result is stored in a documented format. Validation then tests whether the library supports the discrimination it claims. A final step, often neglected, is preparing the library for cross-sensor use by recording enough information to allow resampling to other instruments.

Errors enter at every stage. Illumination changes during a measurement session, panel degradation, inconsistent sensor height, canopy movement in wind, soil background in incompletely closed canopies, and inconsistent stage assignment all contribute. None of these is exotic; all of them are capable of producing apparent genotype differences that are really measurement differences.

5. Preprocessing and Feature Selection:

Raw reflectance is rarely used directly. Instrument noise, scattering caused by leaf surface properties and canopy geometry, and baseline drift all obscure the chemical information of interest, and preprocessing exists to reduce those effects. Smoothing by the Savitzky–Golay filter fits a low-order polynomial within a moving window and remains the standard first step because it reduces noise without flattening absorption features (Savitzky & Golay, 1964). First and second derivatives suppress additive and multiplicative baseline effects and sharpen shoulders, which is why derivative spectra are common in red edge work. Standard normal variate transformation centres and scales each spectrum individually, and detrending removes curvature, both introduced to handle scatter in diffuse reflectance measurements (Barnes et al., 1989). Multiplicative scatter correction addresses the same problem by regressing each spectrum on a reference spectrum (Geladi et al., 1985). Continuum removal normalises absorption features against a fitted hull so that depth can be compared across spectra. Rinnan et al. (2009) reviewed these methods and reached a conclusion that still holds: no single preprocessing chain is best in general, and the appropriate choice depends on the instrument, the sample presentation and the property being predicted. Table 3 sets out the main options and what each is for.

Feature selection matters even more in genotype work than in trait estimation, for a simple reason. A hyperspectral scan may contain several hundred to more than a thousand bands, while a variety trial may contain a few dozen plots. Neighbouring bands are highly correlated, so the effective dimensionality is far lower than the nominal one, and a flexible classifier given all bands and few samples will find separations that do not generalise. Principal component analysis reduces dimension but produces components that are hard to interpret biologically. Variable importance in projection scores rank wavelengths within a partial least squares model. Competitive adaptive reweighted sampling selects informative wavelengths iteratively and has been widely adopted in chemometrics (Li et al., 2009). Boruta compares the importance of each wavelength against randomised shadow features, which gives a defensible rule for inclusion (Kursa & Rudnicki, 2010). Regularised regression, recursive feature elimination, mutual information ranking and permutation importance are all in use. Narrowband indices and derivative features can also be treated as engineered inputs.

Two cautions apply. First, selected wavelengths are often unstable: repeat the selection on a resampled dataset and a different set emerges, particularly when samples are few. Reporting selection stability, rather than a single winning set, is more informative. Second, feature selection performed on the whole dataset before splitting leaks information into the test set and inflates accuracy. Selection must be performed inside the cross-validation loop, on the training fold alone. Table 4 compares the main approaches.

Method	Principle	Strengths	Limitations	References
Principal component analysis	Orthogonal projection maximising variance	Efficient dimension reduction; stable	Components lack direct biological interpretation; variance is not always discriminative	Liu et al. (2010)
PLS variable importance in projection	Ranks wavelengths by contribution to PLS components	Interpretable; supervised; widely used	Threshold choice is arbitrary; unstable in small samples	Burnett et al. (2021)
Competitive adaptive reweighted sampling	Iterative Monte Carlo sampling retaining informative wavelengths	Effective at reducing to a small band set	Results vary between runs; requires repetition to assess stability	Li et al. (2009)

Boruta	Compares importance against randomised shadow features	Provides a principled inclusion rule; handles correlated features	Computationally heavy; tends to retain many bands	Kursa & Rudnicki (2010)
Regularised regression (LASSO, elastic net)	Shrinks coefficients, driving some to zero	Selection and fitting in one step	Arbitrary selection among correlated bands	Roberts et al. (2017)
Recursive feature elimination	Iteratively removes least important features	Simple; works with any base model	Prone to leakage if run outside the validation loop	Király & Tóth (2025)
Permutation importance	Measures accuracy loss when a feature is shuffled	Model agnostic; intuitive	Misleading with highly correlated bands	Breiman (2001)
Narrowband vegetation indices	Fixed ratios of selected wavelengths	Interpretable; transferable between sensors	Discards information; index choice constrains what can be detected	Thenkabail et al. (2000)
Derivative and red edge features	Engineered features from spectral shape	Directly tied to a known biological transition	Sensitive to noise and to red edge estimation method	Baranoski & Rokne (2005); Ju et al. (2010)

Hyperspectral Discrimination Across Crops:

Work on genotype and variety discrimination has grown around a handful of crops, and the pattern of findings is more instructive than any individual result. Table 5 compiles the studies discussed here.

Table 5. Hyperspectral studies relevant to genotype and variety discrimination across crops.

Crop	Measurement level	Sensor or data type	Analytical approach	Reported outcome	Key limitation for genetic inference	Reference
Rice	Leaf, laboratory	Hyperspectral leaf reflectance	Wavelength selection and spectral discrimination	Small subsets of narrow bands separated paddy rice samples	Laboratory conditions; leaf level only	Song et al. (2011)
Rice	Panicle, canopy	Hyperspectral reflectance	PCA followed by support vector classification	Panicles discriminated from other canopy components	Single stage; discrimination is of organ, not genotype	Liu et al. (2010)
Rice	Grain	Visible/near infrared reflectance	Chemometric classification	Variety and authenticity distinguished	Grain, not standing crop; single laboratory setting	Liang et al. (2009)
Rice	Grain	Hyperspectral imaging	Chemometric models	Variety and quality discriminated together	No field validation	Wang et al. (2014)
Rice	Seed	Near infrared hyperspectral imaging	Deep learning	High reported classification accuracy for seed varieties	Controlled imaging; validation splits not group based	Jin et al. (2022)
Rice	Seed	Hyperspectral imaging	Deep learning classifier	Variety classification reported	As above	Wang et al. (2025)
Rice	Whole plant	Automated hyperspectral imaging	Trait quantification	Physiological trait variation quantified across materials	Controlled environment	Ting et al. (2023)
Wheat	Canopy, field	Hyperspectral imaging	Machine learning	High-throughput phenotyping of breeding material	Trait prediction rather than variety labelling	Kaur et al. (2024)

Wheat	Canopy, breeding trials	Canopy hyperspectral reflectance	Regression and prediction models	Grain yield predicted across trials	Yield prediction, not genotype identification	Montesinos-López et al. (2017)
Maize	Leaf, field	Hyperspectral reflectance	PLS regression	Physiological and biochemical traits predicted across genotypes	Traits, not direct genotype classification	Yendrek et al. (2016)
Maize	Seed	Hyperspectral imaging	Chemometric calibration	Seed varieties discriminated	Laboratory; small variety panel	Zhang et al. (2012)
Maize	Seed	Hyperspectral imaging	CNN-LSTM framework	Seed variety classification	Deep model with limited external validation	Zhu et al. (2025)
Soybean	Seed	Hyperspectral imaging	Ensemble machine learning	Seed varieties classified non-destructively	Laboratory setting	Wei et al. (2020)
Soybean	Canopy	Hyperspectral reflectance bands	Machine learning plus GWAS	Yield-related bands associated with genomic loci	Single programme and environment set	Yoosefzadeh-Najafabadi et al. (2021)
Cotton	Canopy, field	Spectral reflectance through season	Regression on lint yield	Within-season reflectance related to yield	Cultivar discrimination not the objective	Attia & Rajan (2016)
Cotton	Leaf, field	In situ hyperspectral reflectance	Indices and machine learning	Leaf chlorophyll phenotyped	Trait level; environment not separated	Thorp et al. (2024)
Turfgrass	Canopy	Broadband spectral reflectance	Species and cultivar models under drought	Species and cultivars separated by drought response	Broadband; response depends on stress level	Jiang & Carrow (2007)
Woody species	Canopy	Field spectroradiometer and simulated bands	Seasonal separability analysis	Separability varied strongly with season	Demonstrates date dependence of discrimination	Soubry & Guo (2021)

Rice has attracted sustained attention, partly because varietal identity carries commercial weight. At leaf level under laboratory conditions, Song et al. (2011) selected a small number of wavelengths that separated paddy rice samples, showing that discriminating information can be concentrated in relatively few narrow bands. Canopy-level work has been less about variety and more about condition, but panicle-stage measurements have been separated using principal components followed by support vector classification (Liu et al., 2010). A large body of laboratory imaging work addresses grain and seed rather than the standing crop: visible and near infrared spectra have been used to judge variety and authenticity (Liang et al., 2009), hyperspectral imaging to discriminate variety and quality together (Wang et al., 2014), and deep networks applied to near infrared hyperspectral images of seed have reported high classification accuracy (Jin et al., 2022; Wang et al., 2025). Hyperspectral imaging has also been used to quantify physiological trait variation among rice materials in a controlled system (Ting et al., 2023). Wheat work has leaned towards trait prediction and breeding application rather than variety labelling. Hyperspectral imaging combined with machine learning has been used for high-throughput phenotyping in winter wheat (Kaur et al., 2024), and canopy hyperspectral reflectance has been used to predict grain yield across breeding trials (Montesinos-López et al., 2017). Maize has been approached from both ends: leaf-level hyperspectral reflectance predicts physiological and biochemical traits across genotypes (Yendrek et al., 2016), while seed-level imaging with chemometric or deep learning models separates varieties (Zhang et al., 2012; Zhu et al., 2025). Soybean shows the same split, with seed variety classification by ensemble learning (Wei et al., 2020) alongside canopy work aimed at yield-related traits and their genetic basis (Yoosefzadeh-Najafabadi et al., 2021). Cotton studies have concentrated on chlorophyll and nitrogen status and on the relationship between within-season reflectance and lint yield rather than on cultivar identification (Attia & Rajan, 2016; Thorp et al., 2024). Outside the major grains, turfgrass species and cultivars have been separated by their reflectance responses to drought (Jiang & Carrow, 2007), and seasonal spectral separation of woody species has been shown to depend strongly on date of measurement (Soubry & Guo, 2021).

A parallel literature on stress phenotyping is worth reading alongside these results, because it establishes how much of the spectral signal belongs to condition rather than identity. Hyperspectral imaging with machine learning has been used to phenotype salt stress in tomato and related material at high throughput (Feng et al., 2019), and to screen salt tolerance when combined with chlorophyll fluorescence measurements (Weng et al., 2023). Reviews of heavy metal stress phenotyping cover the same ground for contamination effects (Zhai et al., 2023), and integrated platforms now combine reflectance, fluorescence and RGB imaging in one controlled system (Hassan

& Chang, 2025). The relevance to genotype work is direct: every stress response documented in this literature is a source of spectral variation that a variety classifier can absorb and mislabel as genetic difference. Reading across this work, four patterns stand out. Reported accuracies are consistently higher for seed and grain measured in the laboratory than for canopies measured in the field, which is unsurprising given the difference in control. Studies that use few genotypes report better separation than those with many, and comparisons across papers with different genotype counts are close to meaningless. The wavelengths identified as important cluster in the red edge and the pigment-sensitive visible region in most canopy studies, and extend into shortwave infrared regions in seed studies where composition drives the signal. And the strongest results almost always come from a single site and season, which leaves the durability of the discrimination untested.

Growth Stage and Genotype-by-Environment Effects:

A crop canopy changes more between tillering and maturity than most varieties differ from each other at any single moment. This is the central difficulty in spectral genotype work. As a rice canopy develops, leaf area increases and soil background disappears from the sensor's field of view, chlorophyll concentration rises and then falls, panicles emerge and alter both colour and structure, and senescence changes water and pigment content together. Every one of these shifts the reflectance curve, often by more than the difference between two varieties. Studies of rice growth and nitrogen status document how strongly stage governs the relationship between spectra and plant properties (Yang & Chen, 2004; Xue et al., 2004; Bajwa et al., 2009), and stage-specific models are common precisely because pooled models perform poorly (Chang et al., 2005; Ryu et al., 2011). Where growth stage itself is the classification target, the same spectral changes become the signal rather than the nuisance.

The practical consequence is that a classifier trained on spectra pooled across stages may be separating development rather than genotype, especially when varieties differ in maturity duration. A model that appears to identify a variety may in fact be identifying that the variety was three days further into heading. This is a confounding problem, not a modelling problem, and no amount of algorithm choice fixes it. The remedy is design: measure at defined phenological stages, record the stage for every scan, and validate within stage before validating across stages.

Environment introduces the same difficulty on another axis. Temperature, soil type, water availability, nutrition, disease pressure, radiation and management all modify the traits that drive reflectance. Nitrogen supply changes pigment concentration and canopy density (Nguyen & Lee, 2006; Xue & Yang, 2008; Yi et al., 2010; Tan et al., 2018). Disease alters both colour and structure (Cao et al., 2015). Nutrient imbalance produces characteristic spectral changes that a classifier can easily mistake for varietal identity (Wang et al., 2024), and contamination stress affects which spectral resolution is adequate at all (Zhang et al., 2019). It is useful to think of any measured spectrum as a sum of contributions:

Observed spectrum = genetic signal + environmental signal + developmental signal + measurement effects

Most published studies estimate the total and attribute it to the first term. Separating the terms requires the same materials grown in several environments and measured at matched stages, which is exactly what multi-environment breeding trials already provide. Spectral data collected across such trials is the most promising route to isolating the genetic component, and work using multi-environment spectral and genomic data points that way (Krause et al., 2019; Mróz et al., 2024). Table 9 summarises stage and environment factors, the spectral changes they produce, and how each can be controlled.

Table 9. Growth stage and genotype-by-environment effects on spectral discrimination.

Stage or environmental factor	Spectral change observed	Affected region(s)	Risk of confounding genotype signal	Recommended control or validation	References
Vegetative growth and canopy closure	Rising NIR reflectance; declining soil contribution	NIR plateau; visible	High: varieties differing in vigour appear separable for structural reasons	Measure at matched canopy cover; record LAI or cover	Yang & Chen (2004); Bajwa et al. (2009)
Tillering to stem elongation	Increasing chlorophyll; red edge shifts to longer wavelengths	Red edge; green peak	High	Stage-specific models; record stage on a standard scale	Xue et al. (2004); Ju et al. (2010)
Heading and flowering	Panicle emergence alters colour and canopy structure	Visible; NIR	Very high: varieties differing in heading date are separated by phenology	Match by developmental stage, not calendar date	Ryu et al. (2011); Liu et al. (2010)
Grain filling	Pigment decline begins; water content changes	Red edge; water bands	High	Repeat measurement across the window; model trajectory rather than a point	Chang et al. (2005)
Maturity and senescence	Chlorophyll loss; increasing	Visible; SWIR	High	Exclude or model separately;	Attia & Rajan (2016)

	dry matter signal			senescence rate is itself a trait	
Nitrogen supply	Pigment concentration and canopy density change together	Green peak; red edge	Very high: mimics varietal greenness differences	Common nutrition across genotypes; record fertiliser regime	Nguyen & Lee (2006); Xue & Yang (2008); Tan et al. (2018)
Water availability and drought	Water absorption deepens or weakens; structure changes	Water bands; NIR	High	Irrigation treatment recorded; test genotype signal within treatment	Jiang & Carrow (2007)
Nutrient imbalance (N, P, K)	Distinct spectral responses by deficiency type	Visible; red edge	High	Record nutrient status; avoid pooling across regimes	Wang et al. (2024)
Disease and pest pressure	Lesions and canopy loss alter colour and structure	Visible; NIR	High	Score incidence; exclude affected plots or model as covariate	Cao et al. (2015)
Soil type and background	Exposed soil contributes to plot reflectance	Visible; NIR	Moderate to high in open canopies	Record soil; restrict to closed-canopy stages where possible	Yang & Chen (2004)
Illumination and viewing geometry	Amplitude and shape shift with sun and view angles	All regions	Moderate; systematic if plots measured at different times	Fixed geometry; frequent white reference; record angles and time	Pacheco-Labrador & Martín (2015)
Sensor and spectral resolution	Band placement and FWHM alter derived features	All regions	Moderate; severe when comparing across studies	Record response functions; resample explicitly for cross-sensor work	Cundill et al. (2015); Zhang et al. (2019)

Machine Learning, Deep Learning and Validation

The analytical toolkit is broad and, for this problem, fairly settled. Partial least squares discriminant analysis remains a sensible default: it handles collinear inputs, is stable with modest sample sizes, and yields loadings that can be inspected. Linear discriminant analysis serves similar purposes when dimension has already been reduced. Support vector machines handle non-linear boundaries efficiently in high dimensions (Cortes & Vapnik, 1995), and random forests offer robustness and a usable importance measure at the cost of a tendency to favour correlated features (Breiman, 2001). Gradient boosting often performs marginally better on tabular spectral features but adds tuning burden. Neural networks have been applied throughout, from shallow architectures for nitrogen retrieval (Yi et al., 2010) to one-dimensional convolutional networks over the spectral axis and two- or three-dimensional networks over hyperspectral image cubes (Jiang & Li, 2020; Jin et al., 2022; Zhu et al., 2025). Attention-based architectures are appearing. Table 6 compares these methods.

Table 6. Classification and regression methods applied to hyperspectral crop data.

Method	Handles collinearity	Typical sample size need	Interpretability	Main risk	References
PLS-DA	Yes, by design	Modest	Good; loadings inspectable	Component number chosen by unguarded cross-validation	Burnett et al. (2021)
Linear discriminant analysis	Only after dimension reduction	Modest	Good	Fails when bands exceed samples	Liu et al. (2010)
Support vector machine	Yes, via kernels	Modest to large	Low	Kernel and cost tuning inside the validation loop is often neglected	Cortes & Vapnik (1995); Liu et al. (2010)
Random forest	Partially	Moderate	Moderate; importance available	Importance biased towards correlated bands	Breiman (2001)

Gradient boosting	Partially	Moderate to large	Moderate	Heavy tuning; easy to overfit small spectral datasets	Wei et al. (2020)
Artificial neural network	Yes	Large	Low	Overfitting with few plots; unstable band attribution	Yi et al. (2010)
1D CNN over wavelengths	Yes	Large	Low	Learns dataset-specific artefacts when plots are few	Jin et al. (2022)
2D/3D CNN over image cubes	Yes	Large	Low	Spatial context can encode plot identity, inflating accuracy	Jiang & Li (2020); Zhu et al. (2025)
Attention and transformer models	Yes	Very large	Low to moderate	Data requirements exceed most breeding-trial datasets	Bourriz et al. (2025)

Model choice, however, is rarely what determines whether a published accuracy means anything. Validation design is. The single most damaging practice in this literature is splitting data at the level of the individual scan. When ten or twenty scans are taken from each plot and those scans are randomly assigned to training and test sets, nearly every plot appears in both. The model then needs only to recognise the plot, and plot recognition is trivial because scans from one plot share illumination, canopy structure, soil background and operator technique. A concrete example makes the difference clear. Suppose a trial contains twelve plots per variety and fifteen scans per plot. A random split across all 180 scans per variety will typically report accuracy above ninety per cent. The same data split so that whole plots are assigned to either training or testing, with no plot spanning both, will usually report substantially less, and the difference is not model failure but the removal of leakage.

Grouped validation is the correction. Cross-validation folds should be defined by plot, so that all scans from a plot move together; where the question concerns unseen material, folds should be defined by genotype, so that a variety appearing in training never appears in testing. Roberts et al. (2017) set out the general principle for data with hierarchical or spatial structure, and the argument transfers directly. Király and Tóth (2025) document the same failure in chemometric practice, including the related error of scaling or preprocessing across the whole dataset before splitting. Beyond grouping, external validation matters: a model tested on a different season, site or sensor tells the reader something a resampled internal estimate cannot. Best-practice guidance for spectroscopic trait prediction now makes these expectations explicit (Burnett et al., 2021; Kothari et al., 2023), and we would apply the same expectations to discrimination studies.

Spectral Biomarkers and Biological Interpretation

It is worth restating a distinction that is easy to blur. A statistical feature is a wavelength or combination that improves a model. A spectral indicator is a feature that behaves consistently across datasets. A physiological or biochemical indicator is a feature with an established mechanistic link to a measurable plant property, such as chlorophyll concentration or leaf water content. A genetic marker is a variant in the genome. These are four different things, and a wavelength can be the first without being any of the others.

The point is not pedantic. If a model reports that reflectance near 720 nm discriminates two rice varieties, the honest interpretation is that the varieties differ in something that affects the red edge, most plausibly chlorophyll concentration or leaf structure, at that stage and in that environment. Whether the difference is genetically controlled, and through which loci, is a separate question requiring genomic data. The strongest work in this area builds the bridge explicitly rather than asserting it: measuring the biochemical trait alongside the spectrum, showing that the spectral feature predicts the trait, and then relating the trait or the spectral value to genomic variation (Yendrek et al., 2016; Herritt et al., 2016; Yassue et al., 2023). Recent work linking leaf reflectance to gene expression pushes in the same direction (Chen et al., 2025).

Interpretation is also where disagreement in the literature becomes visible, and the disagreements are worth naming. Preprocessing choices that improve results in one study degrade them in another, which Rinnan et al. (2009) attribute to differences in sample presentation rather than to any intrinsic superiority of a method. The wavelengths identified as most important vary considerably between studies of the same crop, partly because selection methods differ and partly because selection is unstable in small samples. Accuracies reported for controlled laboratory measurements are not comparable to those from field canopies, and leaf-level results do not transfer cleanly to canopy level, where structure and background dominate. Satellite results sit further still from laboratory ones. Treating these as one body of evidence, rather than as separate measurement regimes, is a recurring source of overstatement.

From Field to UAV to Satellite

Measurements of the same crop at different scales are not interchangeable. A leaf clip on a spectrometer isolates a single leaf under a controlled artificial source. A field spectroradiometer held above a plot integrates leaves, stems, shadow, panicles and exposed soil under natural illumination that changes through the session. An unmanned aerial platform adds viewing geometry effects and platform motion, and trades spectral detail for spatial coverage. A satellite adds atmosphere, coarse pixels that mix plots with roads and bunds, and a fixed overpass time that may not coincide with the stage of interest. Each step upward loses control and gains extent, and a

spectral feature that discriminates at one scale need not survive to the next. This progression from controlled field measurements to increasingly extensive UAV and satellite observations is illustrated in Fig. 2.

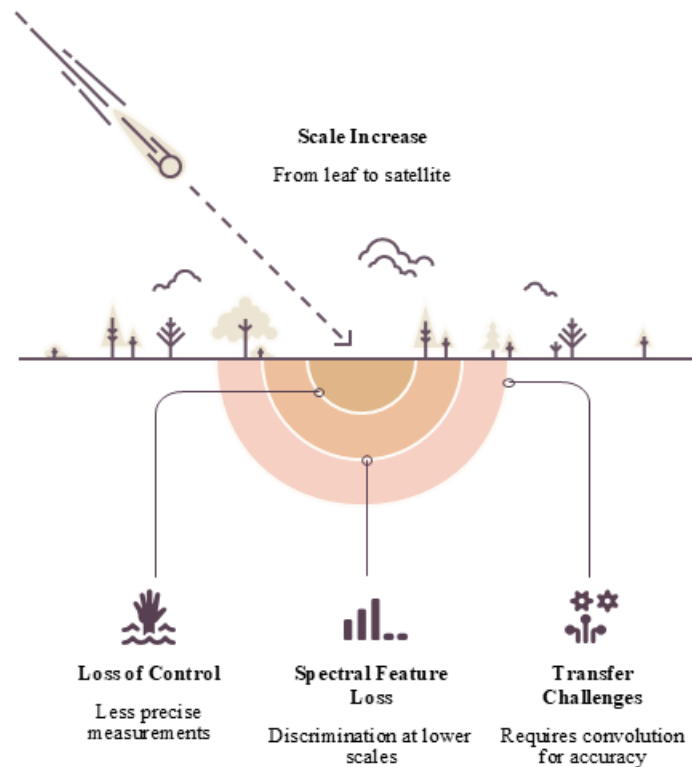


Fig. 2. Scale-dependent impacts on spectral measurements from field-level observations to UAV and satellite sensing, illustrating the increase in spatial extent and corresponding loss of measurement control, spectral detail, and transferability.

Transferring a field spectral library to another sensor is a defined operation rather than an approximation. Each target instrument has band centres, bandwidths expressed as full width at half maximum, and a spectral response function describing its sensitivity across each band. A field spectrum is convolved with those response functions to produce the reflectance that instrument would have recorded. Cundill et al. (2015) show that ignoring response function differences shifts index values enough to matter, which is a direct argument for storing full-resolution spectra in a library rather than only derived indices. Resampling alone is not sufficient for satellite work: atmospheric correction, bidirectional reflectance effects and spatial mixing all intervene between a plot spectrum and a pixel value.

Spaceborne imaging spectroscopy has changed what is feasible at this scale. PRISMA and EnMAP provide contiguous narrowband data at spatial resolutions of tens of metres (Guanter et al., 2015; Cogliati et al., 2021), and crop type mapping with hyperspectral satellite data combined with machine learning is now an active field (Bourriz et al., 2025). Unmanned platforms occupy the intermediate position and have become standard in breeding trials, where plot-level phenotyping across large nurseries is the goal (Dobbels & Lorenz, 2019; Herr et al., 2023).

Ambition should be calibrated here. Mapping where a known variety was planted, characterising genetic resources at landscape scale, monitoring breeding trials, and screening germplasm collections are all plausible applications. Identifying an unknown genotype directly from a satellite pixel is not supported by current evidence, and we are not aware of published work demonstrating it under field conditions with proper external validation. Pixel sizes remain large relative to plot sizes in most breeding contexts, revisit intervals rarely match phenological windows, and the environmental and developmental terms discussed in Section 7 are harder to control from orbit, not easier.

Integration with Genomics and Breeding

The reason a genetics readership should care about spectral libraries is that a spectral phenotype, once standardised, behaves like any other quantitative trait and can be analysed with the same tools. Treating reflectance at each wavelength, or a derived spectral index, as a phenotype allows association mapping against marker data. Herritt et al. (2016) took this approach for a reflectance index in soybean and identified associated genomic loci. Yoosefzadeh-Najafabadi et al. (2021) extended it to yield-related reflectance bands, and Yassue et al. (2023) used genome-wide association on hyperspectral reflectance to examine growth-related traits in maize.

Prediction is the second route, and arguably the more practical one. Krause et al. (2019) constructed relationship matrices from hyperspectral reflectance and used them for genomic prediction of grain yield in wheat, showing that spectral information can complement marker-based relationships rather than merely substituting for them.

Canopy reflectance has been used to predict yield directly in breeding data (Montesinos-López et al., 2017), and spectral indices have been used as secondary traits in indirect selection (Lozada et al., 2019). Multi-environment work suggests that imagery-derived similarity among genotypes can augment genomic selection (Mróz et al., 2024). Multi-trait and network approaches provide a framework for interpreting several correlated spectral phenotypes together (Momen et al., 2019).

Beyond mapping and prediction, standardised spectral libraries support genetic resource characterisation. Screening a germplasm collection spectrally is far cheaper than measuring its biochemistry, and while a spectrum cannot replace laboratory determination, it can prioritise which accessions merit it. The same data serves genotype-by-environment analysis, because spectral phenotypes are cheap enough to collect in every environment of a multi-location trial. Looking further, spectral phenotypes sit naturally alongside transcriptomic and metabolomic data, since all three describe molecular and biochemical state; the demonstration that leaf reflectance relates to gene expression makes that integration less speculative than it once was (Chen et al., 2025). Table 8 maps these applications to the evidence base.

Table 8. Applications linking hyperspectral phenotyping to genetics, genomics and breeding.

Application	How spectral data is used	Evidence status	References
Genome-wide association	Reflectance bands or indices treated as quantitative phenotypes	Demonstrated in soybean and maize	Herritt et al. (2016); Yoosefzadeh-Najafabadi et al. (2021); Yassue et al. (2023)
Genomic prediction	Spectral relationship matrices combined with marker relationships	Demonstrated in wheat	Krause et al. (2019)
Indirect selection	Spectral indices used as secondary traits for yield	Demonstrated in winter wheat	Lozada et al. (2019)
Multi-environment prediction	Imagery-derived genotype similarity augments genomic selection	Demonstrated in spring wheat	Mróz et al. (2024)
Yield prediction in breeding data	Canopy hyperspectral reflectance predicts grain yield	Demonstrated across breeding trials	Montesinos-López et al. (2017)
Multi-trait interpretation	Networks and structural equation models across correlated spectral traits	Methodological framework available	Momen et al. (2019)
Germplasm characterisation	Spectral screening prioritises accessions for laboratory analysis	Plausible; limited published demonstration at collection scale	Rao et al. (2007)
Linking spectra to expression	Leaf reflectance related to gene expression patterns	Early demonstration	Chen et al. (2025)
Spectral diversity as a diversity proxy	Spectral variation used to index biological variation	Demonstrated at community level; unresolved within crop species	Wang et al. (2018)

Standardization, Research Gaps and Future Perspectives

Standardization is the precondition for everything discussed above. Comparable work requires agreement on spectral range and sampling interval, on calibration protocols and reference panels, on metadata fields, on preprocessing conventions and reporting, on file formats, and on how spectral response functions are documented so that resampling is reproducible. Quality control criteria should be stated rather than assumed. Data belongs in public repositories with persistent identifiers, and benchmark datasets with fixed splits would allow methods to be compared honestly. The FAIR principles apply directly, and the phenotyping community's minimum information standards provide a starting point that needs only spectral extension (Ćwiek-Kupczyńska et al., 2016). Analysis code should be published alongside data, since preprocessing choices materially affect results.

We would propose a compact framework. A genotype-oriented spectral library should specify: the genetic material and its identifiers; the experimental design and replication; the phenological stages at which measurement occurs; the instrument and its full characteristics; the acquisition protocol including geometry and calibration; the preprocessing applied, with raw data retained; the plot and scan grouping structure; the environmental description; and the validation results that establish what the library can and cannot support. The standardized workflow shown in Fig. 3 provides the operational basis for building comparable genotype-oriented spectral libraries.

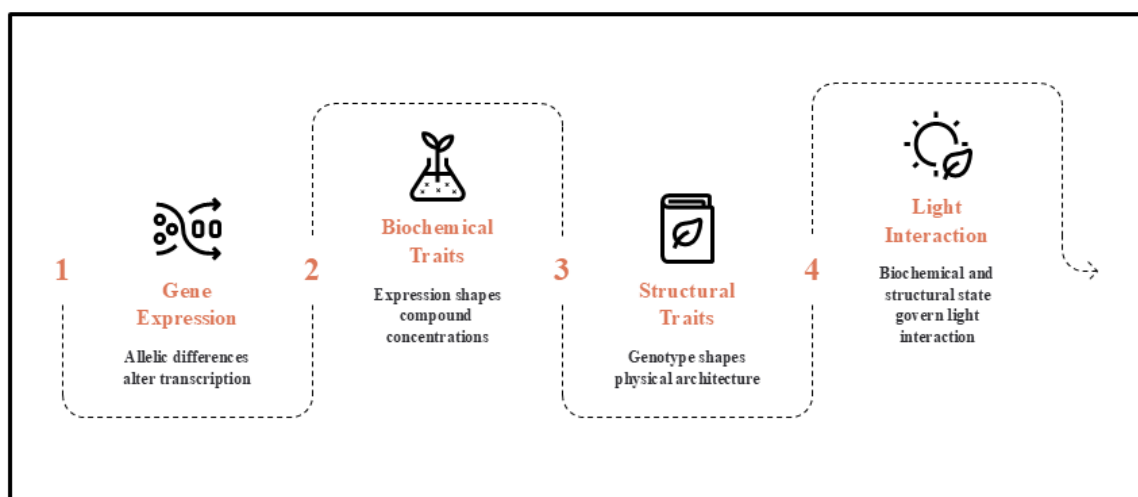


Fig. 3. Conceptual pathway linking genotype, gene expression, biochemical and structural traits, light interaction, and the resulting spectral phenotype.

The gaps that current literature supports are these. Standardised, openly available genotype-level spectral libraries are rare; most published work uses private datasets built for one experiment. Multi-environment spectral datasets on common genetic material are scarce, which is precisely why the genetic and environmental terms remain confounded. Cross-sensor transfer is documented for indices (Cundill et al., 2015) but rarely validated end to end from field library to satellite product. Genotype-by-environment analysis of spectral phenotypes is in its infancy. Biological interpretation of discriminating wavelengths is often omitted. Integration with genomic data exists but is concentrated in a small number of groups and crops (Herritt et al., 2016; Krause et al., 2019; Yoosefzadeh-Najafabadi et al., 2021; Yassue et al., 2023). External validation across seasons and sites is uncommon. Sample sizes are frequently small relative to the number of bands. Data leakage through scan-level splitting is widespread and largely unacknowledged (Roberts et al., 2017; Király & Tóth, 2025). Open benchmarks do not exist. Preprocessing and reporting are inconsistent enough to prevent meta-analysis. Longitudinal data following the same genotypes through development across years is almost absent.

Looking forward, we see a chain rather than a list. Genetic resources are phenotyped spectrally under standardised protocols; those measurements accumulate into libraries; libraries are resampled to unmanned aerial and satellite sensors for scaling; machine learning, including large pretrained models for spectral data, extracts features; those features are interpreted biologically to yield spectral biomarkers; biomarkers and spectral phenotypes are analysed jointly with genomic and other omics data; the resulting genotype–phenotype models support breeding decisions. Several components of this chain are established. Libraries, resampling, machine learning and association analysis all have published evidence behind them. Others are hypotheses at present: foundation models for crop spectra, digital twin representations of breeding trials, and routine multi-omics integration with spectral data are directions rather than demonstrated capabilities, and we flag them as such.

Table 7. Challenges identified in this review and the solutions supported by published evidence.

Challenge	Consequence	Supported solution	References
Growth stage confounded with genotype	Classifiers separate development rather than genetic background	Record stage for every scan; validate within stage before across stages	Yang & Chen (2004); Chang et al. (2005)
Environment confounded with genotype	Apparent varietal signal reflects nutrition, water or disease	Multi-environment trials with matched stages; record environmental covariates	Nguyen & Lee (2006); Wang et al. (2024); Mróz et al. (2024)
Scan-level data splitting	Reported accuracy inflated by plot recognition	Group folds by plot; leave-one-plot-out and leave-one-genotype-out	Roberts et al. (2017); Király & Tóth (2025)
Preprocessing applied before splitting	Information leaks from test to training data	Fit all preprocessing and scaling inside the training fold	Király & Tóth (2025); Burnett et al. (2021)
High dimensionality with few plots	Unstable models and unstable band selection	Band selection with stability assessment; regularisation	Li et al. (2009); Kursa & Rudnicki (2010)
Inconsistent preprocessing across studies	Results cannot be compared or pooled	Report full preprocessing chain; retain and publish raw spectra	Rinnan et al. (2009)
Sensor differences between studies	Indices and features are not comparable	Store full-resolution spectra with response function metadata; resample explicitly	Cundill et al. (2015); Pacheco-Labrador & Martín (2015)
Missing biological interpretation	Discriminating wavelengths treated as genetic markers	Measure biochemical traits alongside spectra; link spectrum to trait to genome	Yendrek et al. (2016); Chen et al. (2025)

Absent metadata standards	Datasets cannot be reused or combined	Adopt and extend minimum information standards for phenotyping	Ćwiek-Kupczyńska et al. (2016)
No open benchmarks	Method comparisons are not reproducible	Publish libraries with fixed, group-aware splits in open repositories	Baldrige et al. (2009); Meerdink et al. (2019)

CONCLUSION

Hyperspectral measurement gives plant scientists a fast, non-destructive way to observe the biochemical and structural state of a crop, and that state is shaped in part by the plant's genetic background. Turning this observation into something genetics can use depends less on sensors and algorithms than on organisation. Spectral libraries provide that organisation. By fixing how spectra are acquired, calibrated, described and stored, a library makes measurements from different plots, seasons and sites comparable, and comparability is what allows a spectral phenotype to be treated as a trait rather than as a curiosity.

The evidence reviewed here supports a measured position. Genotypes and varieties can be separated spectrally, sometimes with high accuracy, but the strength of that separation depends on when the measurement was taken, where the crop was grown, what instrument was used and how the model was tested. Growth stage and environment are not nuisances to be averaged away; they are large effects that can masquerade as genetic signal, and controlling them is a matter of experimental design rather than analysis. Equally, a wavelength that improves a classifier is not thereby a genetic marker. Interpretation should run through measured biological traits, and claims should stop where the evidence does.

Two priorities follow. The first is rigour in validation. Splitting data by plot and by genotype, keeping preprocessing and feature selection inside the validation loop, and testing models on unseen seasons and sites would change what many published accuracies mean. The second is integration. Spectral phenotypes analysed alongside marker data, in association studies and in prediction models, are where this field connects to breeding practice rather than sitting beside it.

The opportunity that follows is scale. If field libraries are built with enough care that they can be resampled to airborne and spaceborne instruments, phenotyping stops being limited to the plots a team can walk. Characterising genetic resources across whole regions, monitoring breeding trials from orbit, and screening germplasm at a cost that permits repetition all become reasonable ambitions. Reaching that point requires standardised libraries, honest validation and biological interpretation, in that order.

REFERENCES

1. Araus, J.L.; Cairns, J.E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science*, 19(1), 52-61. <https://doi.org/10.1016/j.tplants.2013.09.008>
2. Araus, J.L.; Kefauver, S.C.; Zaman-Allah, M.; Olsen, M.S.; Cairns, J.E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends in Plant Science*, 23(5), 451-466. <https://doi.org/10.1016/j.tplants.2018.02.001>
3. Attia, A.; Rajan, N. (2016). Within-season growth and spectral reflectance of cotton and their relation to lint yield. *Crop Science*, 56(5), 2688-2701. <https://doi.org/10.2135/cropsci2015.05.0296>
4. Bajwa, S.G.; Mishra, A.R.; Norman, R.J. (2009). Canopy reflectance response to plant nitrogen accumulation in rice. *Precision Agriculture*, 11(5), 488-506. <https://doi.org/10.1007/s11119-009-9142-0>
5. Baldrige, A.M.; Hook, S.J.; Grove, C.I.; Rivera, G. (2009). The ASTER spectral library version 2.0. *Remote Sensing of Environment*, 113(4), 711-715. <https://doi.org/10.1016/j.rse.2008.11.007>
6. Baranoski, G.V.G.; Rokne, J.G. (2005). A practical approach for estimating the red edge position of plant leaf reflectance. *International Journal of Remote Sensing*, 26(3), 503-521. <https://doi.org/10.1080/01431160512331314029>
7. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43(5), 772-777. <https://doi.org/10.1366/0003702894202201>
8. Boren, E.J.; Boschetti, L.; Johnson, D.M. (2019). Characterizing the variability of the structure parameter in the PROSPECT leaf optical properties model. *Remote Sensing*, 11(10), 1236. <https://doi.org/10.3390/rs11101236>
9. Bourriz, M.; Hajji, H.; Laamrani, A.; Elbouanani, N.; Ait Abdelali, H.; Bourzeix, F.; El-Battay, A.; Amazirh, A.; Chehbouni, A. (2025). Integration of hyperspectral imaging and AI techniques for crop type mapping: Present status, trends, and challenges. *Remote Sensing*, 17(9), 1574. <https://doi.org/10.3390/rs17091574>
10. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>

11. Burnett, A.C.; Anderson, J.; Davidson, K.J.; Ely, K.S.; Lamour, J.; Li, Q.; Morrison, B.D.; Yang, D.; Rogers, A.; Serbin, S.P. (2021). A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *Journal of Experimental Botany*, 72(18), 6175-6189. <https://doi.org/10.1093/jxb/erab295>
12. Cao, X.; Luo, Y.; Zhou, Y.; Fan, J.; Xu, X.; West, J.S.; Duan, X.; Cheng, D. (2015). Detection of powdery mildew in two winter wheat plant densities and prediction of grain yield using canopy hyperspectral reflectance. *PLOS ONE*, 10(4), e0121462. <https://doi.org/10.1371/journal.pone.0121462>
13. Chandra, H.; Nidamanuri, R.R. (2024). Object-based spectral library for knowledge-transfer-based crop detection in drone-based hyperspectral imagery. *Precision Agriculture*, 26(1). <https://doi.org/10.1007/s11119-024-10203-3>
14. Chang, K.-W.; Shen, Y.; Lo, J.-C. (2005). Predicting rice yield using canopy reflectance measured at booting stage. *Agronomy Journal*, 97(3), 872-878. <https://doi.org/10.2134/agronj2004.0162>
15. Chen, Y.; Monks, L.; Rubio, V.E.; Cox, A.J.; Swenson, N.G. (2025). Linking leaf hyperspectral reflectance to gene expression. *Communications Earth & Environment*, 6(1), 694. <https://doi.org/10.1038/s43247-025-02696-1>
16. Cogliati, S.; Sarti, F.; Chiarantini, L.; Cosi, M.; Lorusso, R.; Lopinto, E.; Miglietta, F.; Genesio, L.; Guanter, L.; Damm, A.; Pérez-López, S.; Scheffler, D.; et al. (18 authors) (2021). The PRISMA imaging spectroscopy mission: overview and first performance analysis. *Remote Sensing of Environment*, 262, 112499. <https://doi.org/10.1016/j.rse.2021.112499>
17. Cortes, C.; Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>
18. Coussement, J.; Henke, M.; Lootens, P.; Roldán-Ruiz, I.; Steppe, K.; De Swaef, T. (2018). Modelling leaf spectral properties in a soybean functional-structural plant model by integrating the prospect radiative transfer model. *Annals of Botany*, 122(4), 669-676. <https://doi.org/10.1093/aob/mcy105>
19. Cundill, S.; van der Werff, H.; van der Meijde, M. (2015). Adjusting spectral indices for spectral response function differences of very high spatial resolution sensors simulated from field spectra. *Sensors*, 15(3), 6221-6240. <https://doi.org/10.3390/s150306221>
20. Curran, P.J.; Dungan, J.L.; Gholz, H.L. (1990). Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree Physiology*, 7(1-4), 33-48. <https://doi.org/10.1093/treephys/7.1-2-3-4.33>
21. Ćwiek-Kupczyńska, H.; Altmann, T.; Arend, D.; Arnaud, E.; Chen, D.; Cornut, G.; Fiorani, F.; Frohberg, W.; Junker, A.; Klukas, C.; Lange, M.; Mazurek, C.; et al. (25 authors) (2016). Measures for interoperability of phenotypic data: minimum information requirements and formatting. *Plant Methods*, 12(1), 44. <https://doi.org/10.1186/s13007-016-0144-4>
22. Dobbels, A.A.; Lorenz, A.J. (2019). Soybean iron deficiency chlorosis high-throughput phenotyping using an unmanned aircraft system. *Plant Methods*, 15(1), 97. <https://doi.org/10.1186/s13007-019-0478-9>
23. Fahlgren, N.; Gehan, M.A.; Baxter, I. (2015). Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology*, 24, 93-99. <https://doi.org/10.1016/j.pbi.2015.02.006>
24. Feng, X.; Zhan, Y.; Wang, Q.; Yang, X.; Yu, C.; Wang, H.; Tang, Z.; Jiang, D.; Peng, C.; He, Y. (2019). Hyperspectral imaging combined with machine learning as a tool to obtain high-throughput plant salt-stress phenotyping. *The Plant Journal*, 101(6), 1448-1461. <https://doi.org/10.1111/tpj.14597>
25. Geladi, P.; MacDougall, D.; Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy*, 39(3), 491-500. <https://doi.org/10.1366/0003702854248656>
26. Gitelson, A.A.; Merzlyak, M.N.; Lichtenthaler, H.K. (1996). Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *Journal of Plant Physiology*, 148(3-4), 501-508. [https://doi.org/10.1016/S0176-1617\(96\)80285-9](https://doi.org/10.1016/S0176-1617(96)80285-9)
27. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C.; Straif, C.; Fischer, S.; et al. (39 authors) (2015). The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing*, 7(7), 8830-8857. <https://doi.org/10.3390/rs70708830>
28. Hassan, M.A.; Chang, C.Y.Y. (2025). PhenoGazer: A high-throughput phenotyping system to track plant stress responses using hyperspectral reflectance, nighttime chlorophyll fluorescence and RGB imaging in controlled environments. *Plant Phenomics*, 7(2), 100047. <https://doi.org/10.1016/j.plaphe.2025.100047>
29. Herr, A.W.; Adak, A.; Carroll, M.E.; Elango, D.; Kar, S.; Li, C.; Jones, S.E.; Carter, A.H.; Murray, S.C.; Paterson, A.; Sankaran, S.; Singh, A.; Singh, A.K. (2023). Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. *Crop Science*, 63(4), 1722-1749. <https://doi.org/10.1002/csc2.21028>

30. Herritt, M.; Dhanapal, A.P.; Fritschi, F.B. (2016). Identification of genomic loci associated with the photochemical reflectance index by genome-wide association study in soybean. *The Plant Genome*, 9(2). <https://doi.org/10.3835/plantgenome2015.08.0072>
31. Horler, D.N.H.; Dockray, M.; Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2), 273-288. <https://doi.org/10.1080/01431168308948546>
32. Jacquemoud, S.; Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34(1), 75-91. [https://doi.org/10.1016/0034-4257\(90\)90100-Z](https://doi.org/10.1016/0034-4257(90)90100-Z)
33. Jacquemoud, S.; Ustin, S.L.; Verdebout, J.; Schmuck, G.; Andreoli, G.; Hosgood, B. (1996). Estimating leaf biochemistry using the PROSPECT leaf optical properties model. *Remote Sensing of Environment*, 56(2), 194-202. [https://doi.org/10.1016/0034-4257\(95\)00238-3](https://doi.org/10.1016/0034-4257(95)00238-3)
34. Jiang, Y.; Carrow, R.N. (2007). Broadband spectral reflectance models of turfgrass species and cultivars to drought stress. *Crop Science*, 47(4), 1611-1618. <https://doi.org/10.2135/cropsci2006.09.0617>
35. Jiang, Y.; Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: A review. *Plant Phenomics*, 2020, 4152816. <https://doi.org/10.34133/2020/4152816>
36. Jin, B.; Zhang, C.; Jia, L.; Tang, Q.; Gao, L.; Zhao, G.; Qi, H. (2022). Identification of rice seed varieties based on near-infrared hyperspectral imaging technology combined with deep learning. *ACS Omega*, 7(6), 4735-4749. <https://doi.org/10.1021/acsomega.1c04102>
37. Ju, C.-H.; Tian, Y.-C.; Yao, X.; Cao, W.-X.; Zhu, Y.; Hannaway, D. (2010). Estimating leaf chlorophyll content using red edge parameters. *Pedosphere*, 20(5), 633-644. [https://doi.org/10.1016/S1002-0160\(10\)60053-7](https://doi.org/10.1016/S1002-0160(10)60053-7)
38. Kaur, S.; Kakani, V.G.; Carver, B.; Jarquin, D.; Singh, A. (2024). Hyperspectral imaging combined with machine learning for high-throughput phenotyping in winter wheat. *The Plant Phenome Journal*, 7(1), e20111. <https://doi.org/10.1002/ppj2.20111>
39. Király, P.; Tóth, G. (2025). Being aware of data leakage and cross-validation scaling in chemometric model validation. *Journal of Chemometrics*, 39(4), e70026. <https://doi.org/10.1002/cem.70026>
40. Koh, J.C.O.; Banerjee, B.P.; Spangenberg, G.; Kant, S. (2022). Automated hyperspectral vegetation index derivation using a hyperparameter optimisation framework for high-throughput plant phenotyping. *New Phytologist*, 233(6), 2659-2670. <https://doi.org/10.1111/nph.17947>
41. Kong, X.; Zhao, J.; Liu, Y.; Bai, B.; Yang, J.; Fan, Y.; Li, G.; Li, Z.; Wan, S. (2025). The PROSPECT model in high-throughput phenotyping for peanut leaf parameter estimation: Comparative performance of hyperspectral inversion models. *Current Plant Biology*, 43, 100498. <https://doi.org/10.1016/j.cpb.2025.100498>
42. Kothari, S.; Beauchamp-Rioux, R.; Blanchard, F.; Crofts, A.L.; Girard, A.; Guilbeault-Mayers, X.; Hacker, P.W.; Pardo, J.; Schweiger, A.K.; Demers-Thibeault, S.; Bruneau, A.; Coops, N.C.; Kalacska, M.; Vellend, M.; Laliberté, E. (2023). Predicting leaf traits across functional groups using reflectance spectroscopy. *New Phytologist*, 238(2), 549-566. <https://doi.org/10.1111/nph.18713>
43. Krause, M.R.; González-Pérez, L.; Crossa, J.; Pérez-Rodríguez, P.; Montesinos-López, O.; Singh, R.P.; Dreisigacker, S.; Poland, J.; Rutkoski, J.; Sorrells, M.; Gore, M.A.; Mondal, S. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3 Genes|Genomes|Genetics*, 9(4), 1231-1247. <https://doi.org/10.1534/g3.118.200856>
44. Kursu, M.B.; Rudnicki, W.R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, 36(11), 1-13. <https://doi.org/10.18637/jss.v036.i11>
45. Li, H.; Liang, Y.; Xu, Q.; Cao, D. (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta*, 648(1), 77-84. <https://doi.org/10.1016/j.aca.2009.06.046>
46. Li, D.; Cheng, T.; Zhou, K.; Zheng, H.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W. (2017). WREP: A wavelet-based technique for extracting the red edge position from reflectance spectra for estimating leaf and canopy chlorophyll contents of cereal crops. *ISPRS Journal of Photogrammetry and Remote Sensing*, 129, 103-117. <https://doi.org/10.1016/j.isprsjprs.2017.04.024>
47. Liang, L.; Liu, Z.X.; Yang, M.H.; Zhang, Y.X.; Wang, C.H. (2009). Discrimination of variety and authenticity for rice based on visual/near infrared reflection spectra. *Journal of Infrared and Millimeter Waves*, 28(5), 353-356.
48. Liu, Z.Y.; Shi, J.J.; Zhang, L.W.; Huang, J.F. (2010). Discrimination of rice panicles by hyperspectral reflectance data based on principal component analysis and support vector classification. *Journal of Zhejiang University SCIENCE B*, 11(1), 71-78. <https://doi.org/10.1631/jzus.B0900193>

49. Lozada, D.N.; Godoy, J.V.; Ward, B.P.; Carter, A.H. (2019). Genomic prediction and indirect selection for grain yield in US Pacific Northwest winter wheat using spectral reflectance indices from high-throughput phenotyping. *International Journal of Molecular Sciences*, 21(1), 165. <https://doi.org/10.3390/ijms21010165>
50. Meerdink, S.K.; Hook, S.J.; Roberts, D.A.; Abbott, E.A. (2019). The ECOSTRESS spectral library version 1.0. *Remote Sensing of Environment*, 230, 111196. <https://doi.org/10.1016/j.rse.2019.05.015>
51. Momen, M.; Campbell, M.T.; Walia, H.; Morota, G. (2019). Utilizing trait networks and structural equation models as tools to interpret multi-trait genome-wide association studies. *Plant Methods*, 15(1), 107. <https://doi.org/10.1186/s13007-019-0493-x>
52. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; de los Campos, G.; Alvarado, G.; Suchismita, M.; Rutkoski, J.; González-Pérez, L.; Burgueño, J. (2017). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods*, 13, 4. <https://doi.org/10.1186/s13007-016-0154-2>
53. Mróz, T.; Shafiee, S.; Crossa, J.; Montesinos-López, O.A.; Lillemo, M. (2024). Multispectral-derived genotypic similarities from budget cameras allow grain yield prediction and genomic selection augmentation in single and multi-environment scenarios in spring wheat. *Molecular Breeding*, 44. <https://doi.org/10.1007/s11032-024-01449-w>
54. Nguyen, H.T.; Lee, B.-W. (2006). Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression. *European Journal of Agronomy*, 24(4), 349-356. <https://doi.org/10.1016/j.eja.2006.01.001>
55. Nidamanuri, R.R.; Ramiya, A.M. (2013). Spectral identification of materials by reflectance spectral library search. *Geocarto International*, 29(6), 609-624. <https://doi.org/10.1080/10106049.2013.821175>
56. Pacheco-Labrador, J.; Martín, M. (2015). Characterization of a field spectroradiometer for unattended vegetation monitoring. Key sensor models and impacts on reflectance. *Sensors*, 15(2), 4154-4175. <https://doi.org/10.3390/s150204154>
57. Rao, N.R.; Garg, P.K.; Ghosh, S.K. (2007). Development of an agricultural crops spectral library and classification of crops at cultivar level using hyperspectral data. *Precision Agriculture*, 8(4-5), 173-185. <https://doi.org/10.1007/s11119-007-9037-x>
58. Rinnan, Å.; van den Berg, F.; Engelsen, S.B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10), 1201-1222. <https://doi.org/10.1016/j.trac.2009.07.007>
59. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Aroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; Warton, D.I.; Wintle, B.A.; Hartig, F.; Dormann, C.F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913-929. <https://doi.org/10.1111/ecog.02881>
60. Ryu, C.; Suguri, M.; Umeda, M. (2011). Multivariate analysis of nitrogen content for rice at the heading stage using reflectance of airborne hyperspectral remote sensing. *Field Crops Research*, 122(3), 214-224. <https://doi.org/10.1016/j.fcr.2011.03.013>
61. Savitzky, A.; Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627-1639. <https://doi.org/10.1021/ac60214a047>
62. Sims, D.A.; Gamon, J.A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81(2-3), 337-354. [https://doi.org/10.1016/S0034-4257\(02\)00010-X](https://doi.org/10.1016/S0034-4257(02)00010-X)
63. Song, S.; Gong, W.; Zhu, B.; Huang, X. (2011). Wavelength selection and spectral discrimination for paddy rice, with laboratory measurements of hyperspectral leaf reflectance. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(5), 672-682. <https://doi.org/10.1016/j.isprsjprs.2011.05.002>
64. Soubry, I.; Guo, X. (2021). Seasonal spectral separation of western snowberry and wolfwillow in grasslands with field spectroradiometer and simulated multispectral bands. *Environments*, 8(7), 60. <https://doi.org/10.3390/environments8070060>
65. Sozzani, R.; Benfey, P.N. (2011). High-throughput phenotyping of multicellular organisms: finding the link between genotype and phenotype. *Genome Biology*, 12(3), 219. <https://doi.org/10.1186/gb-2011-12-3-219>
66. Tan, K.; Wang, S.; Song, Y.; Liu, Y.; Gong, Z. (2018). Estimating nitrogen status of rice canopy using hyperspectral reflectance combined with BPSO-SVR in cold region. *Chemometrics and Intelligent Laboratory Systems*, 172, 68-79. <https://doi.org/10.1016/j.chemolab.2017.11.014>

67. Thenkabail, P.S.; Smith, R.B.; De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, 71(2), 158-182. [https://doi.org/10.1016/S0034-4257\(99\)00067-X](https://doi.org/10.1016/S0034-4257(99)00067-X)
68. Thorp, K.R.; Thompson, A.L.; Herritt, M.T. (2024). Phenotyping cotton leaf chlorophyll via in situ hyperspectral reflectance sensing, spectral vegetation indices, and machine learning. *Frontiers in Plant Science*, 15, 1495593. <https://doi.org/10.3389/fpls.2024.1495593>
69. Ting, T.C.; Souza, A.C.M.; Imel, R.K.; Guadagno, C.R.; Hoagland, C.; Yang, Y.; Wang, D.R. (2023). Quantifying physiological trait variation with automated hyperspectral imaging in rice. *Frontiers in Plant Science*, 14, 1229161. <https://doi.org/10.3389/fpls.2023.1229161>
70. Ustin, S.L.; Gitelson, A.A.; Jacquemoud, S.; Schaepman, M.; Asner, G.P.; Gamon, J.A.; Zarco-Tejada, P. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment*, 113, S67-S77. <https://doi.org/10.1016/j.rse.2008.10.019>
71. Wang, L.; Liu, D.; Pu, H.; Sun, D.-W.; Gao, W.; Xiong, Z. (2014). Use of hyperspectral imaging to discriminate the variety and quality of rice. *Food Analytical Methods*, 8(2), 515-523. <https://doi.org/10.1007/s12161-014-9916-5>
72. Wang, R.; Gamon, J.A.; Cavender-Bares, J.; Townsend, P.A.; Zygielbaum, A.I. (2018). The spatial sensitivity of the spectral diversity-biodiversity relationship: an experimental test in a prairie grassland. *Ecological Applications*, 28(2), 541-556. <https://doi.org/10.1002/eap.1669>
73. Wang, J.; Chu, Y.; Chen, G.; Zhao, M.; Wu, J.; Qu, R.; Wang, Z. (2024). Characterization and identification of NPK stress in rice using terrestrial hyperspectral images. *Plant Phenomics*, 6, 0197. <https://doi.org/10.34133/plantphenomics.0197>
74. Wang, Z.; Wang, Z.; Zhan, Y.; Yu, S. (2025). Deep learning-based rice seed variety classification model with hyperspectral imaging. *Journal of Electronic Imaging*, 34(5), 053040. <https://doi.org/10.1117/1.JEI.34.5.053040>
75. Wei, Y.; Li, X.; Pan, X.; Li, L. (2020). Nondestructive classification of soybean seed varieties by hyperspectral imaging and ensemble machine learning algorithms. *Sensors*, 20(23), 6980. <https://doi.org/10.3390/s20236980>
76. Weng, H.; Wu, M.; Li, X.; Wu, L.; Li, J.; Atoba, T.O.; Zhao, J.; Wu, R.; Ye, D. (2023). High-throughput phenotyping salt tolerance in JUNCAOs by combining prompt chlorophyll a fluorescence with hyperspectral spectroscopy. *Plant Science*, 330, 111660. <https://doi.org/10.1016/j.plantsci.2023.111660>
77. Xue, L.; Cao, W.; Luo, W.; Dai, T.; Zhu, Y. (2004). Monitoring leaf nitrogen status in rice with canopy spectral reflectance. *Agronomy Journal*, 96(1), 135-142. <https://doi.org/10.2134/agronj2004.1350>
78. Xue, L.; Yang, L. (2008). Recommendations for nitrogen fertiliser topdressing rates in rice using canopy reflectance spectra. *Biosystems Engineering*, 100(4), 524-534. <https://doi.org/10.1016/j.biosystemseng.2008.05.005>
79. Yang, C.-M.; Chen, R.-K. (2004). Modeling rice growth with hyperspectral reflectance data. *Crop Science*, 44(4), 1283-1290. <https://doi.org/10.2135/cropsci2004.1283>
80. Yassue, R.M.; Galli, G.; Chen, C.P.; Fritsche-Neto, R.; Morota, G. (2023). Genome-wide association analysis of hyperspectral reflectance data to dissect the genetic architecture of growth-related traits in maize under plant growth-promoting bacteria inoculation. *Plant Direct*, 7(4), e492. <https://doi.org/10.1002/pld3.492>
81. Yendrek, C.R.; Tomaz, T.; Montes, C.M.; Cao, Y.; Morse, A.M.; Brown, P.J.; McIntyre, L.M.; Leahey, A.D.B.; Ainsworth, E.A. (2016). High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. *Plant Physiology*, 173(1), 614-626. <https://doi.org/10.1104/pp.16.01447>
82. Yi, Q.; Huang, J.; Wang, F.; Wang, X. (2010). Evaluating the performance of PC-ANN for the estimation of rice nitrogen concentration from canopy hyperspectral reflectance. *International Journal of Remote Sensing*, 31(4), 931-940. <https://doi.org/10.1080/01431160902912061>
83. Yoosefzadeh-Najafabadi, M.; Torabi, S.; Tulpan, D.; Rajcan, I.; Eskandari, M. (2021). Genome-wide association studies of soybean yield-related hyperspectral reflectance bands using machine learning-mediated data integration methods. *Frontiers in Plant Science*, 12, 777028. <https://doi.org/10.3389/fpls.2021.777028>
84. Yu, F.; Xiang, S.; Jin, Z.; Bai, J.; Xu, T. (2026). Advancements in leaf-scale radiative transfer modeling: A review with focus on the PROSPECT model. *Agriculture Communications*, 4, 100122. <https://doi.org/10.1016/j.agrcom.2026.100122>

85. Zarco-Tejada, P.J.; Pushnik, J.C.; Dobrowski, S.; Ustin, S.L. (2003). Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sensing of Environment*, 84(3), 283-294. [https://doi.org/10.1016/S0034-4257\(02\)00113-X](https://doi.org/10.1016/S0034-4257(02)00113-X)
86. Zhai, Y.; Zhou, L.; Qi, H.; Gao, P.; Zhang, C. (2023). Application of visible/near-infrared spectroscopy and hyperspectral imaging with machine learning for high-throughput plant heavy metal stress phenotyping: A review. *Plant Phenomics*, 5, 0124. <https://doi.org/10.34133/plantphenomics.0124>
87. Zhang, X.; Liu, F.; He, Y.; Li, X. (2012). Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds. *Sensors*, 12(12), 17234-17246. <https://doi.org/10.3390/s121217234>
88. Zhang, S.; Zhu, Y.; Wang, M.; Fei, T. (2019). Selection of the optimal spectral resolution for the cadmium-lead cross contamination diagnosing based on the hyperspectral reflectance of rice canopy. *Sensors*, 19(18), 3889. <https://doi.org/10.3390/s19183889>
89. Zhu, Q.; Liu, Q.; Ma, D.; Zhu, Y.; Zhang, L.; Wang, A.; Fan, S. (2025). Maize seed variety classification based on hyperspectral imaging and a CNN-LSTM learning framework. *Agronomy*, 15(7), 1585. <https://doi.org/10.3390/agronomy15071585>