



## OPEN ACCESS

## EDITED BY

Jennifer Clarke,  
University of Nebraska-Lincoln,  
United States

## REVIEWED BY

Ce Yang,  
University of Minnesota Twin Cities,  
United States  
Balpreet Kaur Dhatt,  
Bayer Crop Science, United States

## \*CORRESPONDENCE

Diane R. Wang  
[drwang@purdue.edu](mailto:drwang@purdue.edu)

RECEIVED 26 May 2023

ACCEPTED 21 August 2023

PUBLISHED 20 September 2023

## CITATION

Ting T-C, Souza ACM, Imel RK, Guadagno CR, Hoagland C, Yang Y and Wang DR (2023) Quantifying physiological trait variation with automated hyperspectral imaging in rice. *Front. Plant Sci.* 14:1229161. doi: 10.3389/fpls.2023.1229161

## COPYRIGHT

© 2023 Ting, Souza, Imel, Guadagno, Hoagland, Yang and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Quantifying physiological trait variation with automated hyperspectral imaging in rice

To-Chia Ting<sup>1</sup>, Augusto C. M. Souza<sup>2</sup>, Rachel K. Imel<sup>1</sup>, Carmela R. Guadagno<sup>3</sup>, Chris Hoagland<sup>2</sup>, Yang Yang<sup>2</sup> and Diane R. Wang<sup>1\*</sup>

<sup>1</sup>Agronomy Department, Purdue University, West Lafayette, IN, United States, <sup>2</sup>Institute for Plant Sciences, Purdue University, West Lafayette, IN, United States, <sup>3</sup>Botany Department, University of Wyoming, Laramie, WY, United States

Advancements in hyperspectral imaging (HSI) together with the establishment of dedicated plant phenotyping facilities worldwide have enabled high-throughput collection of plant spectral images with the aim of inferring target phenotypes. Here, we test the utility of HSI-derived canopy data, which were collected as part of an automated plant phenotyping system, to predict physiological traits in cultivated Asian rice (*Oryza sativa*). We evaluated 23 genetically diverse rice accessions from two subpopulations under two contrasting nitrogen conditions and measured 14 leaf- and canopy-level parameters to serve as ground-reference observations. HSI-derived data were used to (1) classify treatment groups across multiple vegetative stages using support vector machines ( $\geq 83\%$  accuracy) and (2) predict leaf-level nitrogen content (N, %,  $n=88$ ) and carbon to nitrogen ratio (C:N,  $n=88$ ) with Partial Least Squares Regression (PLSR) following RRReliefF wavelength selection (validation:  $R^2 = 0.797$  and RMSEP = 0.264 for N;  $R^2 = 0.592$  and RMSEP = 1.688 for C:N). Results demonstrated that models developed using training data from one rice subpopulation were able to predict N and C:N in the other subpopulation, while models trained on a single treatment group were not able to predict samples from the other treatment. Finally, optimization of PLSR-RRReliefF hyperparameters showed that 300–400 wavelengths generally yielded the best model performance with a minimum calibration sample size of 62. Results support the use of canopy-level hyperspectral imaging data to estimate leaf-level N and C:N across diverse rice, and this work highlights the importance of considering calibration set design prior to data collection as well as hyperparameter optimization for model development in future studies.

## KEYWORDS

*Oryza sativa*, genetic diversity, growth traits, high-throughput phenotyping, nitrogen

## 1 Introduction

Variation in plant traits reflect differences in genetics, the environment, and their interactions integrated over time (Allard and Bradshaw, 1964). Understanding these relationships could provide mechanistically-based insights into breeding climate-resilient crops, however, conventional methods to measure plant traits can be destructive, may be time-consuming, and often require specific technical skills that vary across methods. These characteristics make such approaches challenging to scale across the large, and often diverse, germplasm panels that are relevant to geneticists and breeders. Deriving trait data from imaging as a systematic means of non-destructive, high-throughput phenotyping has therefore emerged as an important research area for the plant research community over the past decade (Yang et al., 2014; Mir et al., 2019; Yang et al., 2020). For example, morphometric traits (e.g., plant surface area and plant height) are now routinely extracted from Red-Green-Blue (RGB) images using established analysis pipelines (Yang et al., 2014; Gehan et al., 2017; Berry et al., 2018; Kim et al., 2020). In contrast, analogous pipelines for predicting physiological responses and/or biochemical traits from image-derived data have not been well-established (Pasala and Bb, 2020; Yang et al., 2020) and represents a significant gap for plant research communities.

Out of the various types of imaging technologies available to plant researchers, those with hyperspectral sensors have shown the greatest promise for estimating physiological and biochemical traits in plants. These imaging systems are made up of a light source, objective lenses, an imaging spectrograph, hyperspectral sensor(s), and a computer. Results are stored as quantitative electrical signals derived from a vast number of images, each corresponding to the reflectance value – the ratio of reflected radiant flux to the incident flux – of wavelengths ranging between 400 to 2500 nm (Sarić et al., 2022). As early as the 1970s, plant scientists have documented the relationship between leaf traits (e.g., thickness, water content, presence of wax and hairs, and age) and hyperspectral reflectance (Gausman and Allen, 1973; Grant, 1987). General signatures observed for reflectance of plant tissue have also been linked to function and composition. For example, low reflectance in the visible light region (VIS, 480 – 510 nm and 640 – 670 nm) is due to the absorption of light by photosynthetic pigments, reflectance in the near infrared region (NIR, 700 – 1100 nm) is influenced by the arrangement of mesophyll tissues of leaves (Rouse et al., 1974), and the two troughs observed in the short-wave infrared region (SWIR, 1000 – 2500 nm) are affected by plant cell water content (Cotrozzi et al., 2020).

Previous studies have related HSI-derived data to both quantitative and qualitative crop responses to abiotic factors. From controlled-environment phenotyping facilities, these include studies on maize response to different watering regimes (Ge et al., 2016; Asaari et al., 2019; Mertens et al., 2021), quantification of macronutrients in both maize and soybean Pandey et al. (2017), and generation of nitrogen distribution maps at the whole-plant level in wheat Bruning et al. (2019). Cultivated Asian rice (*Oryza sativa*), consumed directly by more than half of the world's population (Muthayya et al., 2014), has also been characterized for its spectral features under field and controlled environment conditions (Arias et al., 2021). For example, Din et al. (2017) reported that leaf area

index during the vegetative stage could be estimated from Vegetation Indices (VIs) derived from hyperspectral data using a single *japonica* variety. These experiments were carried out in the field using a spectroradiometer to collect canopy-level data. Spectral data have also been used to detect common leaf diseases across four varieties of rice grown in greenhouse conditions (Feng et al., 2021). For that study, leaves were first sampled destructively and subsequently placed in a hyperspectral imaging system. Yu et al. (2020) developed leaf-level nitrogen content models from spectral data for one *japonica* variety and one *indica* variety under field conditions. Their canopy-level HSI data were collected with a spectroradiometer, and specific VIs were used as model input. While results were promising in each of these rice studies, previous work focused on small numbers of accessions; the potential scalability and general utility of hyperspectral imaging data across more diverse germplasm remain to be tested.

To help address this gap, the current study evaluates 23 genetically diverse *O. sativa* accessions from two divergent subpopulations grown under two nitrogen levels in a phenotyping facility equipped with automated hyperspectral imaging. To establish potential relationships between physiological traits and HSI-derived data, the automated imaging is complemented by a suite of ground-reference observations. While numerous approaches exist that could be considered for analyzing these high dimensional and multicollinear hyperspectral data Mir et al. (2019); Mishra et al. (2020); Arias et al. (2021), we choose to leverage support vector machines (SVMs) for classification, as they have been recognized as an effective image classification algorithm Noble (2006); Gewali et al. (2018), and Partial Least Squares Regression (PLSR) as a computationally-tractable method for trait prediction Burnett et al. (2021). In this study, application of SVM and PLSR follow Principal Components Analysis (PCA) and the RReliefF algorithm, respectively, to retain only the most critical information from the original hyperspectral data. PCA is a classic example of an unsupervised method to reduce data dimensionality Gewali et al. (2018); Yu et al. (2020), while the RReliefF algorithm is a supervised, ranking-based method that calculates an importance score of each wavelength by considering the similarity and dissimilarity between wavelengths Ren et al. (2020).

The overall goal of our study is to test the utility of data derived from an automated hyperspectral imaging system as surrogates for physiological traits across genetically diverse rice. Specific objectives are to (1) assess whether HSI-derived data can classify subpopulation and treatment groupings across time, (2) understand which types of plant traits have the most potential to be predicted, (3) evaluate whether models developed using a single subpopulation or treatment grouping can be used to predict values in the other, and (4) quantify the effects of hyperparameter combinations on predictions by the RReliefF-PLSR framework.

## 2 Materials and methods

### 2.1 Plant materials

A set of 23 bio-geographically diverse accessions from two publicly-available, purified germplasm collections, the Rice Diversity Panel (RDP) 1 and RDP 2 (McCouch et al., 2016), were

evaluated for this study. These 23 lines encompassed 15 *indica* and eight *tropical japonica* accessions that originated from 17 countries (Figure S1). To limit potential confounding effects of development on target traits, these accessions were selected out of the tropically-adapted and phenologically-similar RDP subset screened by Wang et al. (2016b) (Table S1). Seeds were obtained from the USDA-ARS, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas, Genetic Stocks Oryza Collection (<https://www.ars.usda.gov/GSOR>).

## 2.2 Growth conditions

The selected diversity panel was raised at the Ag Alumni Seed Phenotyping Facility (AAPF), a controlled environment high-throughput plant phenotyping facility at Purdue University (West Lafayette, Indiana, U.S.A.) for 94 days during the Summer and Fall of 2020 (Figure S2). The facility is made up of a fully automated growth chamber (Conviron®, Winnipeg, Canada) and weight-based irrigation system (Bosman Van Zaal, Aalsmeer, The Netherlands). A virtual tour of the facility may be found at <https://ag.purdue.edu/aapf/virtual-tour.html>. Three conveyer belts with 32 positions per belt were allocated to this study. Of the 96 total positions, two were designated as “purge pots,” i.e., pots into which the system flushes solutions in between changing fertigation/irrigation regimes. The remaining 94 positions were occupied by 22 genotypes x 2 replicates x 2 nutrient levels and one final genotype (cv. Cybonnet) x 3 replicates x 2 nutrient levels (described below). Having two or more replicates per genotype allowed us to calculate accession-level means for exploring data structure, i.e., subpopulation or treatment. The temperature setpoint in the chamber was 26/22°C day/night, relative humidity at 60%, and photosynthetically active radiation (PAR) levels were recorded between 550–600  $\mu\text{mol photons m}^{-2}\text{s}^{-1}$ .

The environment was additionally tracked by affixing Lascar EL-USB-2-LCD Data Loggers to seven randomly selected pots at the sowing time, which recorded temperature and relative humidity every 10 minutes (Figure S3A). Average temperature and humidity from the loggers across the experimental period were  $29.48 \pm 0.22/23.34 \pm 0.09^\circ\text{C}$  day/night, relative humidity at  $64.62 \pm 0.69\%$ . Lighting conditions followed long day (14 h day/10 h night) scheduling due to the need to accommodate experiments on other species in the same facility: lights turned on daily at 0600 h and turned off at 2000 h. Two seeds were sown per pot (6L in volume) in horticultural substrate, which was made up of a mixture of Profile Porous Ceramic Greens Grade and Berger BM6 All Purpose at a one-to-one ratio by volume. Plants were hand-watered until 10 days after sowing (DAS), at which point the seedlings were thinned and a weight-based automated irrigation was initiated. The experiment was designed with two nutrient treatment levels: high (300 ppm nitrogen, N1) and low (50 ppm nitrogen, N2). The fertigation solution was created by mixing Peters Excel 15-5-15 Cal Mag Special in reverse osmosis (RO) water. Each morning prior to chamber lights turning on, plants were irrigated to a preset weight with RO water. This target weight was increased by about 1.15 times during the experiment to account for the increase in transpirational demand of the growing plants (Figure S3B). RO water irrigation occurred every day except on scheduled days when a fixed volume of fertigation solution (either

high or low concentration, depending on the treatment) was applied instead of RO water. This RO water irrigation and fertigation regime was designed so that each plant should theoretically receive enough water to meet individual transpirational demands while also receiving the fertilizer amount prescribed by their treatment.

## 2.3 Imaging and image processing

Plants were imaged approximately three times per week beginning on 20 DAS and continuing until the end of the experiment using the automated imaging booth in AAPF. This temporal frequency enabled us to average HSI data within each week to reduce the influence of noisy spectra. During each imaging event, one side-view and one top-view images were acquired, though the current study analyzed side-view data only as top-view images after 46 DAS were unavailable due to a camera malfunction. The HSI cameras used a scanning range that encompassed the VIS to NIR region (VNIR; 400 – 1000 nm, MSV 500 VNIR Spectral Camera, Middleton, Spectral Vision, WI, U.S.A.) and the SWIR region (Specim, Oulu, Finland); thus, generation two hypercubes of data image. White and dark reference tests were conducted for each hyperspectral cameras (SVNIR and SWIR) for post-processing images, where the relative light reflectance was estimated for each wavelength. This step was based on the work by Zhang et al. (2019). For the white reference, two panels of known material and spectral signature was scanned with all lights inside the imaging chamber on. The dark reference test was conducted with the lens shutter closed and no lights. The relative light reflectance (%) for each plant was calculated based on the normalized difference between these tests, as seen in Equation 1.

$$\text{reflectance } (\%) = 100 \times \frac{\text{Plant} - \text{Dark}}{\text{White} - \text{Dark}} \quad (1)$$

where,

Plant – the raw digital number measured for the rice plants

Dark - the raw digital number measured during the dark reference tests

White - the raw digital number measured during the white reference tests

For the VNIR hypercube, the rice plants were segmented out of the background using the typical attenuation between Red Edge wavelengths reflectance. The SWIR hypercube segmentation was done using the SURF image registration (Bay et al., 2008) to map the rice plants from the segmented VNIR hypercube (fixed image) using the VNIR-segmented rice plant as the fixed image. After the segmentation, the average light reflectance was calculated and stored in a spreadsheet. Hyperspectral images were processed by using a proprietary processing script in MATLAB (MATLAB, 2018).

## 2.4 Growth and physiological measurements

Pre-dawn gas exchange measurements were taken in the facility growth chamber during early vegetative growth between 0430–

0530 h prior to chamber lights turning on using LI-6800 Portable Photosynthesis System (LI-COR, Lincoln, NE, USA) (Table 1). Leaf-level net assimilation ( $A$ ,  $\mu\text{mol m}^{-2}\text{s}^{-1}$ ), stomatal conductance to water vapor ( $g_{\text{sw}}$ ,  $\text{mol m}^{-2}\text{s}^{-1}$ ), and nighttime transpiration ( $E$ ,  $\text{mol m}^{-2}\text{s}^{-1}$ ) were extracted from the measurements. Environmental conditions in the cuvette matched ambient conditions in the growth chamber: reference  $\text{CO}_2$ , 415  $\mu\text{mol mol}^{-1}$ ; vapor pressure deficit, < 1.5 kPa (average  $\text{VPD} = 1.09 \text{kPa}$ ); PAR, 0  $\mu\text{mol photons m}^{-2}\text{s}^{-1}$ ; and leaf temperature was  $26.87 \pm 0.016^\circ\text{C}$  (mean  $\pm$  SE). As rice leaves are generally too narrow to cover the entire cuvette, leaf width was first measured with a digital caliper prior to each measurement in order to adjust gas exchange measurements to the observed leaf area. A/Ci curves were collected in the facility growth chamber using both LI-6800 and LI-6400XT Portable Photosynthesis Systems (LI-COR, Lincoln, NE, USA) during 59–63 DAS and 87–91 DAS (mid-tillering and late vegetative stages, respectively) with constant PAR of 1200  $\mu\text{mol photons m}^{-2}\text{s}^{-1}$  and reference  $\text{CO}_2$  concentrations were set in the following order: 415, 300, 200, 100, 50, 10, 415, 415, 600, 800, 100, 1200, 415  $\mu\text{mol mol}^{-1}$ . Light response curves were initially conducted on randomly selected plants to determine the PAR level for running A/Ci curves. A/Ci curves were collected between 1000 and 1400 h on the youngest fully expanded leaf as indicated by the emergence of the leaf collar. Leaf temperatures were between 25 and 27  $^\circ\text{C}$  and relative humidity was maintained between 50–70%. After A/Ci curves were run during 59–63 DAS, the leaf used for each curve was excised and

its area measured using a leaf area meter (LAI-2200C; LI-COR, Lincoln, NE, USA). Leaves were oven dried for at least five days at 65  $^\circ\text{C}$  and their dry weights recorded. They were then finely ground using a mortar and pestle and subsequently analyzed for carbon (C, %) and nitrogen (N, %) content using the FlashEA<sup>®</sup> 1112 Nitrogen and Carbon Analyzer for Soils, Sediments and Filters (Thermo Scientific, CE Elantech, Lakewood, NJ) with two replicates per leaf (30 mg per replicate). The equipment was operated according to the flash dynamic combustion method, and resultant signals were translated into the percentage of carbon and nitrogen by the Eager 300 software. During 67–91 DAS, chlorophyll a fluorescence was measured at predawn (0430–0530 h) and midday (1000–1400 h) conditions with a hand-held fluorometer (Fluorpen FP110, Photon System Instruments, Drasov, Czech Republic) on the youngest, fully expanded leaf per plant on three different areas of the leaf blade to minimize possible variation in the efficiency of PSII due to the spatial response to sudden environmental changes in monocotyledons (Oberhuber et al., 1993): the basal one-third, the middle one-third, and the distal one-third. Measurements of Fv/Fm, the maximum efficiency of PSII in dark acclimated leaves, were taken according to Murchie and Lawson (2013). The measuring light of the FluorPen was set at approximately 900  $\mu\text{mol photons m}^{-2}\text{s}^{-1}$  throughout the experiment. Then, we applied a saturation pulse at approximately 1500  $\mu\text{mol photons m}^{-2}\text{s}^{-1}$  to obtain Fv/Fm (pQY), whereas those taken at midday were the maximum efficiency of PSII in light conditions, Fv'/Fm'(mQY)

TABLE 1 Overview of growth and physiology measurements.

Abbreviation	Units	Description
$A$	$\mu\text{mol m}^{-2}\text{s}^{-1}$	pre-dawn leaf-level net assimilation
$C$	%	carbon
$C:N$ (or $\text{CN}$ in Figures)	–	leaf-level carbon to nitrogen ratio
$E$	$\text{mol m}^{-2}\text{s}^{-1}$	pre-dawn leaf-level transpiration
$FB$	g	final biomass
$g_{\text{sw}}$	$\text{mol m}^{-2}\text{s}^{-1}$	pre-dawn leaf-level stomatal conductance to water
$J_{\text{max}}$	$\mu\text{mol m}^{-2}\text{s}^{-1}$	light-saturated potential rate of electron transport
$\log(E)$	–	log transformed pre-dawn leaf-level transpiration
$\log(g_{\text{sw}})$	–	log transformed pre-dawn leaf-level stomatal conductance to water
$\log(\text{SLN})$	–	log(specific leaf nitrogen)
$m\text{QY}$	–	midday quantum yield
$N$	%	nitrogen
$\text{normmQY}$	–	normalized midday quantum yield
$\text{normpQY}$	–	normalized pre-dawn quantum yield
$p\text{QY}$	–	pre-dawn quantum yield
$\text{SLA}$	$\text{cm}^2\text{g}^{-1}\text{dw}$	specific leaf area with respect to dry weight
$\text{SLA}_C$	$\text{cm}^2\text{mg}^{-1}\text{C}$	specific leaf area with respect to carbon
$\text{SLN}$	$\text{mg N cm}^{-2}$	specific leaf nitrogen
$V_{\text{cmax}}$	$\mu\text{mol m}^{-2}\text{s}^{-1}$	the maximum carboxylation rate limited by Rubisco

(Henriques, 2009). Total above-ground biomass from all plants were harvested at the end of the experiment on 94 DAS, oven-dried at 65 °C for at least five days, and their dry weights recorded. Table S2 summarizes the relationship between the timing for physiological trait measurements and the imaging dates and Figure 1 provides an overview of the experiment.

## 2.5 Data analysis

Data were formatted and analyzed in R 4.1.1 (R Core Team, 2021) with packages dplyr (Wickham et al., 2023), tidyverse (Wickham et al., 2019) and reshape2 (Wickham, 2007). Plots were generated using the package ggplot2 (Wickham, 2016) or in base R environment. The code for each physiological trait prediction model can be accessed through GitHub ([https://github.com/To-Chia/rice\\_imaging\\_ms](https://github.com/To-Chia/rice_imaging_ms)).

### 2.5.1 Physiological trait analysis

From the ground-reference observations, we derived specific leaf area ( $SLA, cm^2 g^{-1}$ ), CN ratio ( $C: N$ ), specific leaf area with respect to carbon ( $SLA_C (cm^2 mg^{-1}(C))$ ) and specific leaf nitrogen ( $SLN, mg(N)cm^{-2}$ ). The summary statistics may be found in Table S3.  $E$  and  $gsw$  had right-skewed distributions and thus were log-transformed (denoted as  $\log(E)$  and  $\log(gsw)$ , respectively), prior to conducting the correlation tests and a PCA. Pearson's correlation coefficient was calculated for all pairwise combinations of the physiological traits and plotted with package corrrplot (Wei and Simko, 2021) (Figure S4A). Data for  $SLA$ ,  $C$  and  $N$  on six rice plants were unavailable and were mean-imputed before the pairwise correlation was calculated. The same methods were applied to  $C: N$ ,  $SLA_C$ , and  $SLN$ . A/Ci curves were first assessed for quality control (QC) using the PEcAn.photosynthesis package (<https://pecanproject.github.io/modules/photosynthesis/docs/articles/ResponseCurves.html>). For the 89 curves that passed QC,  $V_{cmax}$  and  $J_{max}$  were estimated using the R plantecophys package (Duursma,

2015). PCA based on correlation matrix was conducted on accession-mean data using the prcomp function in base R (R Core Team, 2021). For unavailable data, their replicates were used to represent the accessionmean value. Agglomerative hierarchical clustering was performed on the Euclidean distance matrix from accession-mean data. Clustering was based on the average linkage method and plotted with package dendextend (Galili, 2015).

### 2.5.2 Hyperspectral imaging data

To obtain stable signals for modeling, imaging events that occurred approximately within the same weeks were averaged (Table S2; Figure 1). Averaged datasets are hereafter referred to as Week 6, 9, 10 and 13 HSI data. PCA with variance-covariance matrix was conducted on weekly accession-mean HSI data. From the results of PCA, wavelengths that had the top ten loadings of the PCs that cumulatively accounted for > 90% of total variance were selected and termed as  $W_{SVM}$ .

### 2.5.3 Signal variation in HSI data through time

SVMs were trained to quantify the prediction accuracy of treatment groupings from  $W_{SVM}$  using the package e1071 (Meyer et al., 2021). First, the classifier of the training week was built with  $W_{SVM}$ . Then, predictions made on the evaluation week were achieved by selecting  $W_{SVM}$  in the training week dataset from the evaluation week. We tested two kernels, the radial basis function (RBF) kernel and the linear kernel. Grid-search of the parameters for both kernels were conducted with 23-fold cross validation. Both rough and fine grid-searches were conducted. In the rough grid-search of RBF kernel, parameters  $cost$  and  $gamma$  were evaluated within the range of  $(2^{-5}, 2^{15})$ ,  $(2^{-15}, 2^{23})$ , respectively, both with an interval of  $2^2$ .

The best parameters resulting from the rough grid-search,  $+/- 2^{0.5}$ , set the range of the subsequent fine-grid search with an interval of  $2^{0.1}$ . When the best parameters from the rough grid-search fell on the boundary, ranges for the fine grid-search used the best value plus or minus 2 within the specified ranges. Parameters found in the

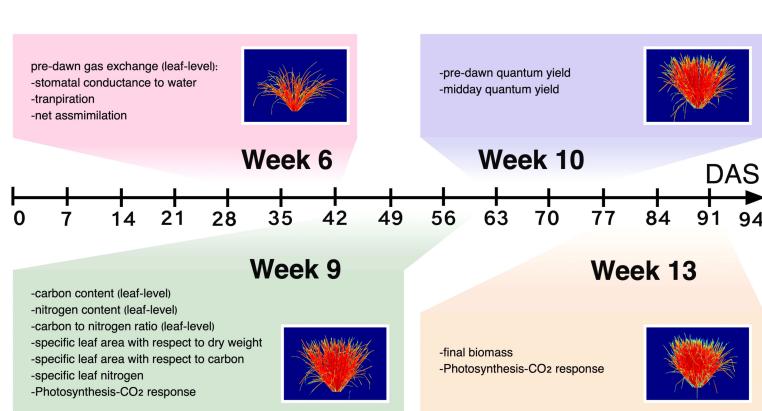


FIGURE 1

Overview of experimental data collection. Fourteen physiological traits were collected at different timepoints over the course of plant development. Hyperspectral imaging was carried out several times per week and imaging events that occurred approximately within the same weeks as phenotyping campaigns were averaged for analyses. The false-colored images in boxes show the same individual rice plant over the four measurement weeks.

fine grid-search were adopted only if model accuracy was higher in the fine grid-search than the rough grid-search; in some cases, accuracies were the same. The parameter, *cost*, which was the only parameter in the linear kernel, was determined with the same method as *cost* in the RBF kernel.

### 2.5.4 Prediction of physiological traits with HSI data

We used the RReliefF algorithm implemented by the CORElearn package (Robnik-Sikonja and Kononenko, 2003) to select relevant wavelengths before utilizing partial least squares regression (PLSR) for predictions (Figure S5). RReliefF algorithms are derived from Relief, developed by Kira and Rendell (1992). We chose RReliefF to filter HSI data as it takes conditional dependencies between variables into account (Robnik-Sikonja and Kononenko, 2003). Wavelengths selected from RReliefF were termed as  $W_{PLSR}$ . The estimator applied in the current study was RReliefF expRank, and the number of iterations was determined by the number of calibration samples multiplied by 100. Note that calibration in this study refers to model training in the context of machine learning while model validation is equivalent to model testing. For PLSR modeling, the R package, pls (Liland et al., 2021), was used, and the procedure for building the models was adapted from Burnett et al. (2021). Specifically, models were fit using the classical orthogonal scores algorithm. Eighty percent of the full dataset was used for calibration and the remaining was for model validation. Sampling was carried out with the criterion that each treatment level contributed equally to the datasets (*i.e.*, the full-data was grouped by treatments prior to sampling). Calibration datasets were used to determine the number of components in the final models by selecting the lowest root mean square error of prediction (RMSEP, Equation 2) in leave-one-out (LOO) cross-validation. The coefficient of determination ( $R^2$ , Equation 3), RMSEP and normalized root mean square of error in predictions (%RMSEP, Equation 4) served as model evaluation metrics.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

$$\% RMSEP = \frac{RMSEP}{range(y)} \quad (4)$$

, where  $y_i$  is the  $i^{th}$  measured value,  $\hat{y}_i$  is the  $i^{th}$  predicted value from the LOO cross-validation for model calibration or the  $i^{th}$  predicted value for model validation,  $\bar{y}_i$  is the mean of the measured values and  $n$  is the sample size in calibration or validation datasets. To account for model variation, jackknife permutation during model calibration was carried out. The derived coefficients were used to compute predicted values in validation datasets. From there, 95% confidence intervals of the predicted values were derived.

To test whether models trained on HSI data from one rice subpopulation or one nitrogen treatment level could be used to predict traits in the other subpopulation or treatment, follow-up

PLSR models were developed for leaf-level  $N$  and C:  $N$ . Procedures for using the RReliefF algorithm and for building PLSR models were the same as described above (Figure S5). Additional analysis on leaf-level  $N$  was carried out to determine the effect of number of  $W_{PLSR}$  and sample size on prediction metrics. This analysis aimed to sort out whether the number of wavelengths used as input to PLSR models described previously was appropriate.  $W_{PLSR}$  at ten values (10, 50, 100, 150, 200, 300, 400, 500, 600 and 700) and sample sizes at five values (48, 60, 70, 78 and 88) were examined using one hundred iterations of each combination of these hyperparameters. We used a grid search method to evaluate this relationship (*i.e.*, effects of sample size and wavelength number on prediction results); grid search was able to be leveraged as PLSR is more computationally tractable than advanced machine learning methods.  $R^2$ , RMSEP and %RMSEP were calculated and their means and standard errors were derived.

## 3 Results

### 3.1 Treatment and subpopulation effects on multivariate physiological traits

Examining pairwise relationships among the traits, we found that all traits had significant correlations with at least one other trait ( $\alpha = 0.01$ ), except for pre-dawn leaf-level net assimilation ( $A$ ,  $\mu\text{mol m}^{-2}\text{s}^{-1}$ ) (Figure S4). As expected, traits that reflected similar aspects of physiology were more correlated. For example, nitrogen content ( $N$ , %) and leaf-level carbon to nitrogen ratio (C:  $N$ ) had a strong negative correlation (-0.96), log transformed pre-dawn leaf-level transpiration ( $\log(E)$ ) and log transformed pre-dawn leaf-level stomatal conductance to water ( $\log(gsw)$ ) were perfectly positively correlated, and  $pQY$  (Fv/Fm) and  $mQY$  (Fv'/Fm') had a weak positive correlation (0.48). We found that overall health status of the rice plants improved with nitrogen enrichment, as reflected by higher  $pQY$  and  $mQY$  in N1 (high nitrogen) than in N2 (low nitrogen) treatments.  $N$  was also found to have moderate and weak positive correlations with  $pQY$  and  $mQY$ , respectively (0.53 and 0.41). It is worth noting that specific leaf area with respect to dry weight (SLA,  $\text{cm}^2\text{g}^{-1}\text{dw}$ ) had correlations with most of these traits; it had a weak negative correlation with C:  $N$ , SLN, and  $\log(E)$  (-0.46, -0.34, and -0.27) and a weak positive correlation with  $N$ , SLA<sub>C</sub>,  $mQY$  and  $pQY$  (0.49, 0.42, 0.41, and 0.29). On the other hand, final biomass (FB, g) only had a weak negative correlation with  $\log(E)$  and  $\log(gsw)$  (-0.35 for both), suggesting that dry matter accumulation of rice in this experiment may have been more associated with water status than with nutrient status.

Due to the correlated nature of the ground-reference dataset, we employed two multivariate approaches to evaluate potential effects of subpopulation and treatment on high-level data structure. PCA on accessionmean trait data suggested both treatment and subpopulation effects (Figure 2A), with PC1 and PC2 reflecting treatment and subpopulation effects, respectively. The traits,  $N$  and C:  $N$ , contributed to PC1 (treatment) the most whereas SLA<sub>C</sub> and SLN had the most influence on PC2 (subpopulation). In contrast, hierarchical clustering showed a clear separation by treatment only

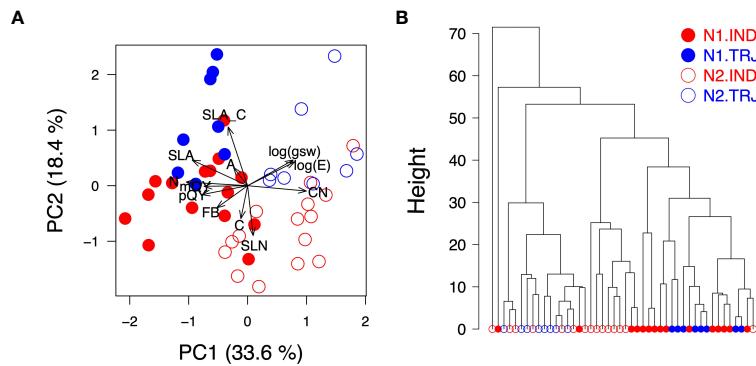


FIGURE 2

Treatment and subpopulation effects on physiological traits. (A) Principal components analysis biplot of PC1 (x-axis) and PC2 (y-axis). (B) Agglomerative hierarchical clustering using accession-mean physiological data. Open circles are observations from the low nitrogen treatment (N2); filled circles indicate the high nitrogen treatment (N1). Blue circles indicate accessions from the *tropical japonica* subpopulation (TRJ); red circles show accessions from the *indica* subpopulation (IND).

and not subpopulation (Figure 2B). Interestingly, the dendrogram revealed that N2 *indica* observations were more closely clustered with the N1 group (both *indica* and *tropical japonica*) than with the cluster that primarily contained N2 *tropical japonica*, suggesting that *indica* may be less responsive to nitrogen enrichment than *tropical japonica*.

### 3.2 Effects of nitrogen on hyperspectral reflectance in rice

Weekly accession mean HSI data gave rise to spectra typical of plants (Figure 3A). Due to the high dimensionality and correlated nature of these data, PCA was applied on accession-mean HSI datasets to examine the potential effects of nitrogen treatment and subpopulation identity. Similar to results of hierarchical clustering of the ground-reference data, only the treatment effect was clear throughout the time period analyzed (Figures 3B–E); separation of treatment levels was primarily determined by PC2 from Weeks 6 to 10 and by PC3 on Week 13. However, nitrogen treatment signals appeared to be lower on Week 13 than on earlier weeks.

Using these PCA results, wavelengths were selected for use in support vector machines (SVM); these are referred to hereafter as  $W_{SVM}$  (see Materials and Methods for details).  $W_{SVM}$  detected in PC2 from Weeks 6 to 10 were similar to those detected in PC3 on Week 13 and were all centered around 715 nm (Figure 3A) in the Red Edge region, indicating that the treatment signals could be attributed to similar spectral regions across the full experimental period. A detailed examination showed that wavelengths around 715 and 910 nm accounted for the highest and the second highest proportions of  $W_{SVM}$ , respectively. Wavelengths around 2200 – 2400 nm, a highly variable region, was found to contribute to the set of  $W_{SVM}$  as well. This was observed for all weeks except Week 9. Lastly,  $W_{SVM}$  of Weeks 10 and 13 included wavelengths around 1400 and 1800 nm, a region known to be informative of water absorption.

We next quantified treatment classification accuracy of these selected wavelengths by applying SVM. Results showed that

$W_{SVM}$  were able to classify nitrogen treatments for any one week using information from any other weeks (Figure 4). Overall, accuracies ranged from 0.63 to 0.91, with accuracy greatest for Weeks 6 through 10 and poorest for Week 13. Accuracy of prediction made by Week 13 dropped as evaluation weeks became more temporally distant from the training week; the lowest accuracy using the Week 13 classifier was observed for Week 6 (0.7 for both kernels) while for Weeks 9 and 10, the accuracy was about 0.78.

### 3.3 Use of hyperspectral imaging as surrogates for physiological traits in rice

Based on promising results of the SVM classification, we next wanted to test the utility of HSI-derived data to estimate trait values. The RReliefF algorithm was first used to derive  $W_{PLSR}$ , i.e. wavelengths selected for partial least squares regression (PLSR) (Materials and Methods). Reflectance data of  $W_{PLSR}$  were then used directly as the predictors of single-response PLSR models. Model results are summarized below based on several trait categories: biomass constituent, growth, photosynthetic capacity, and water transport.

#### 3.3.1 Biomass constituent traits

Ground-reference observations for biomass constituent traits (leaf-level carbon (C, %), nitrogen (N, %) and C:N ratio (C: N)) were collected during Week 9. Distributions of N and C: N were aligned with nitrogen treatment (Figures 5A, E), and  $W_{PLSR}$  for both N and C: N were composed of 400 wavelengths; these sets were similar to each other and included wavelengths from the VIS and Red Edge regions and covered wavelengths around 1500 nm and 2000 nm, regions sensitive to water inside plant cells (Figures 5B, F). In calibration models,  $R^2$  for C: N and N were 0.808 and 0.786, respectively (Figures 5C, G, S6). RMSEP were 1.687 and 0.246 (%), respectively, and %RMSEP were 10.01% and 12.50%, respectively. For model validation, validation of N models ( $R^2 = 0.797$ , RMSEP = 0.264 and %RMSEP = 11.24) were better

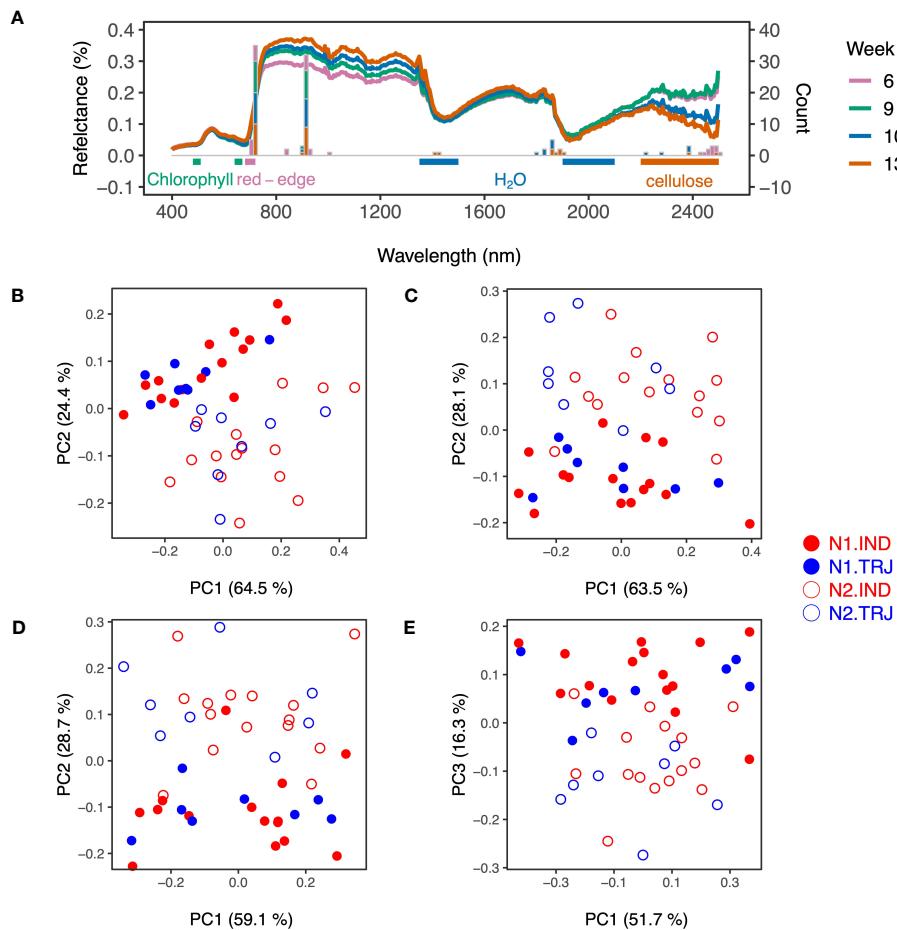


FIGURE 3

Variation in weekly hyperspectral imaging (HSI) reflectance data over plant developmental time. (A) Spectra show accession-mean HSI reflectance data (medians) across weeks. Stacked bar plots indicate key wavelengths identified based on the top 20 - 30 loadings in principal component analysis (PCA) from each week. Horizontal bars represent regions that are sensitive to chlorophyll concentration (green; 480 – 510 nm and 640 – 670 nm), water content (blue; 1350 – 1500 nm and 1900 – 2100 nm) and cellulose (orange; 2200 – 2500 nm). The pink horizontal bar marks the Red Edge region (680 – 720 nm). Plots shown in (B–E) are results from PCA conducted on accession-mean HSI reflectance data on Weeks 6, 9, 10 and 13, respectively.

than validation of C: N models ( $R^2 = 0.592$ ,  $RMSEP = 1.688$  and %  $RMSEP = 18.49$ ) (Figures 5D, H). In contrast, PLSR models could not be built with C as the lowest RMSEP during LOO-calibration suggested that a 0-component model was the most appropriate.

### 3.3.2 Growth traits

Traits of this group included specific leaf area with respect to leaf dry weight ( $SLA$ ,  $cm^2 g^{-1}$ ), specific leaf area with respect to carbon ( $SLA_C$ ,  $cm^2 mg^{-1}(C)$ ), specific leaf nitrogen ( $SLN$ ,  $mg(N)$ )

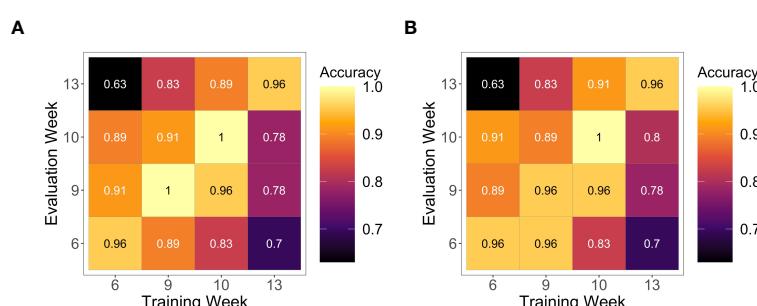


FIGURE 4

Nitrogen treatment levels predicted by support vector machines. Classification results from support vector machines using (A) a radial basis function kernel and (B) a linear kernel. These were trained using  $W_{SV, M}$  from the training week.

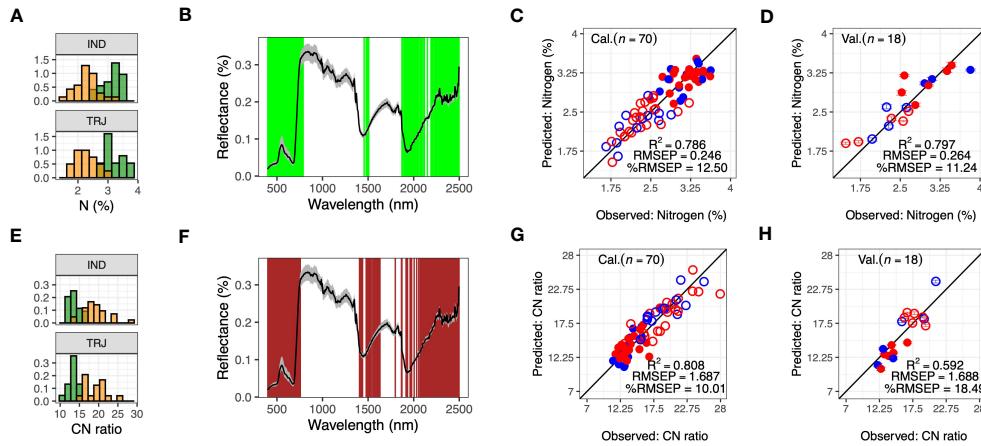


FIGURE 5

Predicting leaf-level nitrogen (N) and CN ratio (C: N) using hyperspectral reflectance imaging data. (A) Distribution of leaf-level nitrogen (N) in *indica* (IND) and *tropical japonica* (TRJ) accessions under high nitrogen (N1, green) and low nitrogen (N2, yellow) treatments. (B) Selected wavelengths (green lines) for predicting N from Week 9 HSI data. (C, D) are model calibration and validation for predicting N, respectively. (E) Distribution of C: N in IND and TRJ in N1 (green) and N2 (orange) treatments. (F) the selected wavelengths (dark red lines) for predicting C: N from the weekly averaged HSI data, and (G, H) are model calibration and validation for predicting C: N, respectively. Error bars are 95% confidence intervals. In (B) and (F), black lines show reflectance values and gray shaded areas indicate reflectances between the 5<sup>th</sup> and the 95<sup>th</sup> percentiles. In panels (C, D, G, H): Open circles = N2 treatment; filled circles are N1 treatment; red = IND and blue = TRJ.

$\text{cm}^{-2}$ ), final biomass ( $FB$ ), and pre-dawn leaf-level net assimilation ( $A, \mu\text{mol m}^{-2} \text{s}^{-1}$ ), which reflects dark respiration.  $FB$  was normalized and  $SLN$  was log transformed, denoted as  $norm_{FB}$ , and  $\log(SLN)$ , respectively, prior to building prediction models. PLSR modeling results are summarized in Figure S7.  $W_{PLSR}$  across the three traits were similar in that they did not include wavelengths in the green light region (520 – 600 nm) but did include the Red Edge region. Compared to  $W_{PLSR}$  for N and C: N, SLA,  $SLA_C$  and  $\log(SLN)$  included  $W_{PLSR}$  in NIR plateau region. We found that calibration models for these three traits were not ideal ( $R^2 \leq 0.295$ ,  $\%RMSEP \geq 14.00$ ) and that their validation models were unstable as model metrics varied greatly when datasets were permuted to generate different combinations of calibration and validation datasets. For  $FB$  and  $A$ , both their calibration and validation models were very poor (Figure S7).

### 3.3.3 Photosynthetic capacity and water transport traits

Photosynthetic capacity traits included pre-dawn quantum yield  $Fv/Fm$  ( $pQY$ ), midday quantum yield  $Fv'/Fm'$  ( $mQY$ ) and  $V_{cmax}$  and  $J_{max}$ .  $pQY$  and  $mQY$  were normalized prior to building models, denoted as  $norm_{mQY}$  and  $norm_{pQY}$ , respectively. Calibration models were not very predictive with  $R^2 \leq 0.344$ ,  $RMSEP \geq 0.832$ , and  $\%RMSEP \geq 14.91$ , and validation models for both traits revealed that the models were not applicable ( $R^2 \leq 0.19$  and  $\%RMSEP \geq 20.90$ ) (Figure S8).  $W_{PLSR}$  for  $V_{cmax}$  and  $J_{max}$  spanned nearly the entire spectral region examined, and similar to the performance of growth trait models, these models were not stable as when the datasets were permuted or could not be established. Likewise, PLSR models could not be established for water transport traits,  $\log(E)$  and  $\log(gsw)$  (Figure S9).

### 3.3.4 Trait predictions in one treatment level using models developed from another

As models to predict leaf-level N and C: N from HSI data performed well, we further examined whether we could develop models using only one treatment level to predict traits in the other treatment level, termed treatment-based models. Due to the smaller sample size for model calibration, sizes of  $W_{PLSR}$  shrunk to 300.  $W_{PLSR}$  for the treatment-based models generally covered the same wavelengths as  $W_{PLSR}$  for model calibration utilizing the full dataset. However, treatment-based models tended to have  $W_{PLSR}$  more dispersed in NIR and SWIR regions, especially for models built from the N2 treatment level (Figure 6). In addition, while  $W_{PLSR}$  from full datasets of N and C: N included wavelengths around 1350–1500 nm, treatment-based models omitted wavelengths in this region, possibly due to the smaller number of wavelengths considered (300 vs 400). Calibration for treatment-based models performed moderately well with  $R^2 \geq 0.35$ , and  $\%RMSEP \leq 18.37$ . However, it was apparent from validation results that models developed using data from one treatment should not be applied to predict samples in the other treatment ( $R^2 \leq 0.127$  and  $\%RMSEP \geq 19.48$ ).

### 3.3.5 Trait predictions in one subpopulation using models developed from another

Rice has two deeply diverged subpopulations, both of which were represented in this study, so we next examined whether models developed from one subpopulation could be used to predict leaf-level N and C: N in the other subpopulation. These are termed subpopulation-based models. In the case of N,  $W_{PLSR}$  for subpopulation-based models appeared similar and included wavelengths in the VIS and Red

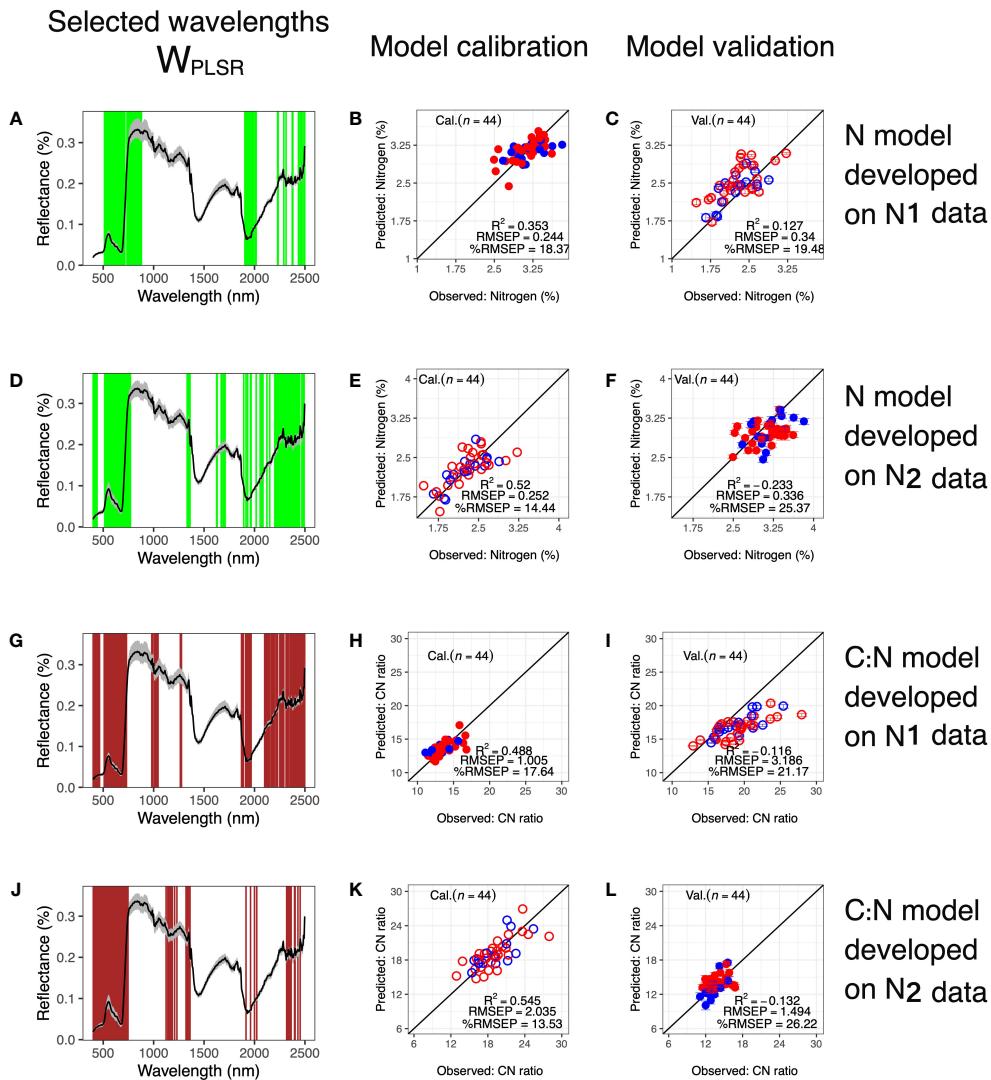


FIGURE 6

Predicting traits in one treatment level using models developed from another. Wavelengths selected based on calibration datasets of (A) *N* data in N1 treatment, (D) *N* data in N2 treatment, (G) C: *N* data in N1 treatment and (J) C: *N* data in N2 treatment, respectively. Corresponding calibration results are in (B, E, H, K), respectively. Calibrated models were used to predict (C) *N* in N2 treatment, (F) *N* in N1 treatment, (I) C: *N* in N1 treatment and (L) C: *N* in N2 treatment, respectively. In (A, D, G, J), black lines show mean reflectance values and gray shaded areas indicate reflectances between the 5<sup>th</sup> and the 95<sup>th</sup> percentiles. In calibration and validation plots: Open circles = N2 treatment; filled circles = N1 treatment; red = IND and blue = TRJ.

Edge regions, wavelengths around 1900 – 2100 nm, and wavelengths at the end of the SWIR region (Figures 7, S10). Both model calibration and validation models performed well, with performance similar to that of models developed from the full dataset (calibration:  $R^2 \geq 0.775$ , RMSEP  $\leq 0.671$  and %RMSEP  $\leq 12.36$ ; validation:  $R^2 \geq 0.689$ , RMSEP  $\leq 0.298$  and %RMSEP  $\leq 13.91$ ).

For the subpopulation-based C: *N* models,  $W_{PLSR}$  was similar to those of the subpopulation-based *N* models (Figure 7), with the exception of when TRJ was used for model calibration,  $W_{PLSR}$  did not include any wavelengths greater than 2300 nm. This differentiated it from the other three subpopulation-based models. All subpopulation-based models excluded wavelengths around 1350–1500 nm, which were used in full dataset models (Figure 5). Model calibration and validation performed well in both

subpopulation-based models for C: *N* (calibration:  $R^2 \geq 0.772$ , RMSEP  $\leq 1.743$  and %RMSEP  $\leq 10.58$ ; validation:  $R^2 \geq 0.682$ , RMSEP  $\leq 2.057$  and %RMSEP  $\leq 12.49$ ). Comparing models developed from the full dataset with subpopulation-based models, the three model metrics (RMSEP,  $R^2$ , and %RMSEP) did not point to the same conclusion, *i.e.* high  $R^2$  in one model did not correspond with low RMSEP or low %RMSEP. From the perspective of RMSEP, models developed with data from IND had better calibration results compared with models developed from the full dataset or with model developed from data in TRJ subpopulation. However, validation results from the full dataset remained the best, as may be expected given its greater variability around the mean. RMSEP from validation for full dataset was 0.133 and 0.369 lower than RMSEP from models developed from IND and TRJ data, respectively.

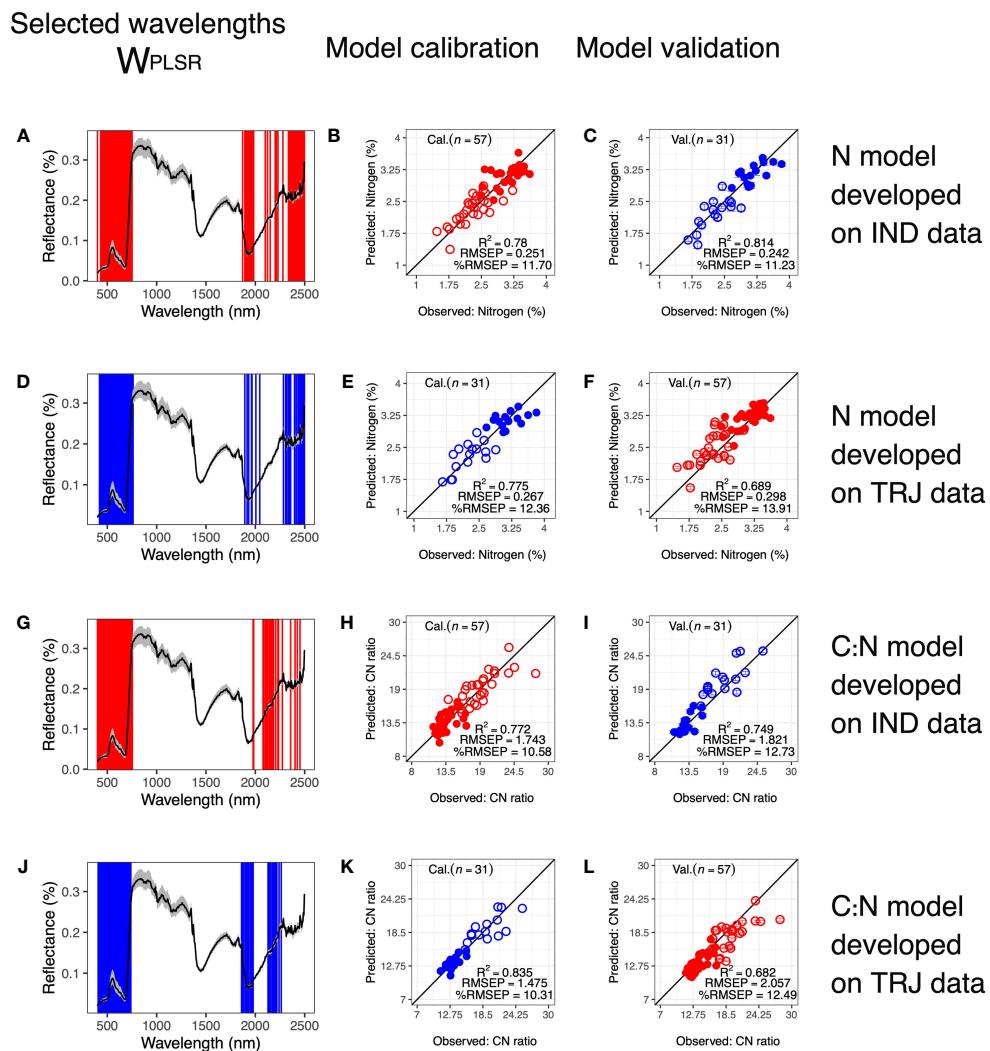


FIGURE 7

Predicting nitrogen (N) and CN ratio (C: N) in one subpopulation using models developed from another. Wavelengths selected based on calibration datasets of (A) N data in IND subpopulation, (B) N data in TRJ subpopulation, (G) C: N data in IND subpopulation and (J) C: N data in TRJ subpopulation, respectively. Corresponding calibration results are in (B, E, H, K), respectively. Calibrated models were used to predict (C) N in TRJ subpopulation, (F) N in IND subpopulation, (I) C: N in TRJ subpopulation and (L) C: N in IND subpopulation, respectively. In (A, D, G, J), black lines show mean reflectance values and gray shaded areas indicate reflectances between the 5<sup>th</sup> and the 95<sup>th</sup> percentiles. In calibration and validation plots: Open circles = N2 treatment; filled circles = N1 treatment; red = IND and blue = TRJ.

### 3.3.6 Effect of sample size and number of wavelengths on model performance

We next investigated the effects of sample size and number of wavelengths on model performance using N as the target trait (Figure 8). Results showed that when total sample size was greater than 70, models with 300 or 400 wavelengths had the best performance. When the number of wavelengths were greater than 400, model performance dropped as indicated by the decrease in  $R^2$  (increase in RMSEP and %RMSEP) in model validation. This suggested that: (1) We needed at least total sample size of 78, which was equivalent to 62 samples in model calibration, to afford model complexity, and (2) models suffered from over-fitting when number of wavelengths were greater than 400. In addition, we noticed that when the total sample size was less than 60, standard errors of the model metrics in validation tended to be large; this

meant models were too complex. Lastly, from the perspectives of  $R^2$  and RMSEP, sample sizes of 48 and 60 caused under-fitting, observed as large differences in model metrics between validation and calibration. Model performance started to converge when total sample sizes were greater than 78. On the other hand, %RMSEP did not converge as much as for  $R^2$  or RMSEP. This may be due to the fact that %RMSEP was a normalized metric and is more robust than  $R^2$  or RMSEP.

## 4 Discussion

In this study, we leveraged the genetic divergence of two rice subpopulations along with two levels of nitrogen application to drive the range of phenotypic variation observed for a suite of

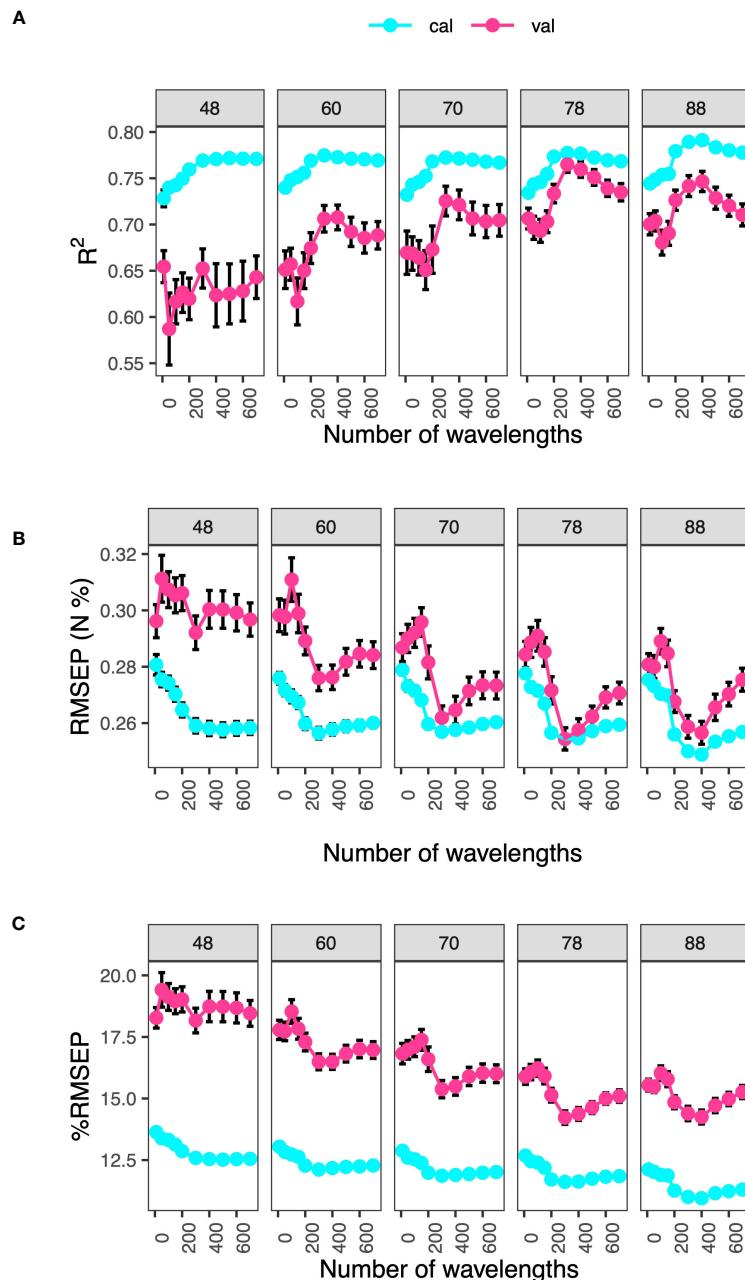


FIGURE 8

Effect of sample size and number of wavelengths on model performance. Model metrics, (A)  $R^2$ , (B) RMSEP (N, %) and (C) %RMSEP, are derived from nitrogen prediction models. Selected wavelengths are those with the top rankings based on the RReliefF algorithm in calibration datasets. Labels denote total sample sizes, with 80% of it used for model calibration. Error bars are standard errors from 100 iterations.

physiological traits and hyperspectral images. As previous work relating spectral features to rice traits focused on limited sets of varieties, our overarching goal was to assess the potential application of hyperspectral reflectance as trait surrogates across diverse rice accessions. The relatively long experimental period (13 weeks) additionally afforded an opportunity to ask about the persistence of HSI utility over the course of plant development.

Overall, we found that the traits,  $N$  (leaf-level nitrogen content) and  $C:N$  (leaf-level carbon to nitrogen ratio) could be predicted using HSI data, whereas for other traits, models either could not be developed or resulted in poor validation. Of note,  $N$  and  $C:N$  were

traits that contributed most to the first principal component of the PCA using ground-reference data, which separated nitrogen treatment classes, and we speculate that this relationship contributed to the high performance of the PLSR models for these traits. Previous work has also demonstrated successful  $N$  prediction in crop species using HSI data (Vigneau et al., 2011; Homolová et al., 2013; Din et al., 2017; Tan et al., 2018; Bruning et al., 2019; He et al., 2020; Meacham-Hensold et al., 2020; Vergara-Díaz et al., 2020; Yu et al., 2020; Baath et al., 2021; Lin et al., 2022). Under various growth conditions, reflectance variation in the VIS and Red-Edge regions along with the end of the SWIR region have

been identified as important for  $N$  prediction; our findings were consistent with these previous studies (Figure 5). The close relationship between  $N$  and chlorophyll may explain why model prediction is often successful, as plant spectral features in the VIS region are largely determined by pigments. In contrast with  $N$ , C:N has not been frequently reported as a predictable trait from hyperspectral reflectance. Gao et al. (2020) found that predicting C:N was related to wavelengths around the red and Red-Edge regions (560 nm, 600 nm, 660 nm, 680 nm, 720 nm, 760 nm and 1470 nm) of forage-grass in the Tibetan Plateau. The  $W_{PLSR}$  for the full C:N model as well as subpopulation-based models largely overlapped with the wavelengths that previous study had identified, except for 760 nm, which was not in  $W_{PLSR}$  for TRJ model, and 1470 nm, which was not found in either TRJ- or IND-based models. Our results showed that  $W_{PLSR}$  between  $N$  and C:N were similar (Figure 5, consistent with the high correlation between the two traits Figure S4). This observation suggests that interpretation of selected wavelengths should be made with caution, as they may not represent cause-effect relationships with the target trait.

HSI data were unable to predict the other traits evaluated in this study. In the case of leaf-level carbon, C, past literature also concluded that C was not predictable Gao et al. (2020), even though specific chemical compounds like lignin, cellulose and starch could be inferred from spectral data (Curran, 1989). We speculate that while those chemical compounds are responsive to specific wavelengths, signals disappear when all of the compounds are considered at once, as in the case of C. For FB, the inability to predict this trait using HSI data sits in contrast to previous research that was able to leverage HSI to predict biomass (Cho et al., 2007). Since FB was only collected at the end of the experimental period in our study, there may have been too much architectural complexity by that late stage in development. In this case, image transformation may be helpful to ameliorate the influence of plant architecture on spectral characteristics, and previous studies have reported various methods to extract relevant information from complex HSI data (Al Makdassi et al., 2019; Asaari et al., 2019; Mishra et al., 2020; Mertens et al., 2021). Additionally, the observed variation of FB between the two treatments in our study was small (Figure S7M). Collecting biomass data at multiple time points would increase the overall size and variation of the ground-reference dataset, which would likely improve model performance. However, this requires more destructive sampling, which was not amenable to the experimental design considered here.

For other traits, we surmise that one general underlying reason for the inability to build satisfactory models is due to the differences in spatial scale between imaging data (taken at canopy-level) and ground-reference observations (taken at leaf-level of the youngest fully expanded leaf of a single tiller). For these particular traits, heterogeneity among individual leaves is likely present, which would not have been captured by our ground-reference observations. Previous research successfully predicted SLA from HSI data by utilizing a hand-held spectroradiometer on the same individual leaves used for ground-reference measurements (Cotrozzi et al., 2020). To improve model performance, future work using canopy-level HSI systems should strive to capture some intra-canopy variation in ground-reference trait data. Similarly, temporal lags between ground-reference measurements

and imaging events should be minimized, especially for traits that represent fast processes in plants that are very sensitive to environmental differences Ge et al. (2016). In our particular setup at the AAPF, plants travel on conveyor belts out of the growth chamber environment to the imaging booth (Figure S2). This takes approximately four minutes, during which temperature drops to around 21 °C. A sensors-to-plants approach would alleviate the effects of temporal decoupling Yang et al. (2020), however, this would depend on available infrastructure. Lastly, Meacham-Hensold et al. (2019) and Fu et al. (2020) were able to predict  $V_{cmax}$  and  $J_{max}$  in tobacco with PLSR. Compared to the current work, these past studies had many more observations; we speculate that our dataset did not have enough variation in A/Ci curves for successful PLSR predictions.

One question we were keen to address here was whether HSI signatures differed in accordance with subpopulation identity. We determined that, despite clear subpopulation differentiation in ground-reference trait data, there were no background genetic signatures in HSI information for rice, based on PCA and model predictions of  $N$  and C:N (Figures 3, 7). This has several implications for moving forward in conducting larger-scale experiments leveraging HSI-predicted physiological traits in diverse rice: (1) subpopulation genetic background does not need to be controlled for in prediction models; and (2) calibration models can be developed from a combination of genetic materials, similar to NIR models for stem non-structural carbohydrate traits in rice (Wang et al., 2016b), conditional that there is adequate phenotypic variation. Contrary to our results, He et al. (2020) suggested that *indica* and *japonica* subpopulations required different models to estimate  $N$ . The main design difference is that their study utilized top-view HSI data, and they analyzed only derived Vegetation Indices. In contrast, we utilized side-view data and analyzed spectral data directly. One important point to make is that our experiment was carried out only during the vegetative stage of rice. Therefore, we speculate that these results may not extend to rice during the reproductive stage; there are several reasons: (1) occlusion of rice leaves or panicles would likely be more apparent in reproductive stage than in early vegetative stage (Din et al., 2017). In addition, some genotypes exert more panicles than the others (Wang et al., 2016a). And (2), during the reproductive stage, resources in leaves are transferred to grains, which would likely lead to changes in hyperspectral signature of leaves.

Automated phenotyping systems, such as Purdue's AAPF, readily support high frequency of hyperspectral imaging, as imaging events can be programmed prior to the start of the experiment and require limited human intervention (Figure S2). In our study, these events occurred two to three times per week throughout the experimental period. The frequency of imaging allowed us to average across several days to obtain pot-level spectral information to be used for downstream analysis and modeling; this attenuates potentially noisy individual spectra and helps detect the overall HSI signature of each plant during timepoints of interest. Having automated imaging occur from Week 6 through Week 13 also enabled us to ask about the consistency of HSI information over developmental time. To test this, we employed SVM to see if HSI could classify nitrogen treatments forwards and backwards in time (*i.e.*, using one

week to predict another week). Classification using HSI data performed very well overall (Figure 4), and prediction accuracies were greater than 0.83, excluding Week 13. Accuracies where Training and Evaluation Weeks were the same (diagonals in Figure 4), however, may be over-estimated due to overlap in datasets. Additional improvements in classification accuracy of nitrogen treatment could consider (1) testing different feature selection methods prior to SVM; (2) analyzing hyperspectral images directly, such as using Fourier Transformation to extract spectral features; or (3) using different classification methods such as CNN (Singh et al., 2016).

It is interesting to note that the latest timepoint, Week 13, showed the poorest performance out of all weeks, no matter whether it served as the Training or the Evaluation Week. This timepoint occurred at the very end of the experimental period, and we speculate that water may have played an interactive role with the nitrogen treatment to modify HSI signatures. We found that higher levels of nitrogen led to greater SLA in our experiment (Table S3), meaning greater transpiring leaf area per unit dry biomass; this translates to an increased demand. Watering regime, which was managed on a target weight basis using the facility's automated system, did not differ between the two nitrogen treatments; therefore, it is plausible that the high nitrogen treatment would have experienced some level of water stress by the end of the experiment. This explanation is supported by the observation that on Week 13, key wavelengths derived from PCA showed a shift from 2400 nm to 1800 nm, adjacent to the region known to be responsive to water (Figure 3A), and we noticed leaf-rolling (O'Toole and Cruz, 1980) in some plants during midday in the latter part of the experimental period.

With the establishment of automated plant phenotyping systems, the promise of hyperspectral images to serve as surrogates for plant physiological traits has motivated a wide range of studies across different species and treatments over the last decade. However, it is important to recognize that these image-derived reflectance data are influenced by multiple factors such as plant architecture, scaling, equipment (e.g., cameras), and the particular setup of each facility, which impacts the timing of imaging as plants move from their growing environment to imaging booths. For these reasons, prediction models are generally considered non-transferrable across species, experiments, or facilities (Cho et al., 2007; Verrelst et al., 2015). Flexible approaches to designing within-experiment calibration sets with robust ground-reference data are therefore needed; general frameworks such as RReliefF-PLSR prediction coupled with the optimization of hyperparameters for model development, as presented here, can be extended beyond any single study. As image processing methods continue to advance, researchers may be able to more readily extract reflectances of individual leaves from canopy-level images; this would enable plant scientists to explore within-plant trait distributions and how that varies across genotypes. These kinds of insights can help deepen understanding of plant physiological response mechanisms across intra-specific genetic variation.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://purr.purdue.edu/publications/4079/>.

## Author contributions

AS, CG, CH, YY, and DW conceptualized the study. RI, CH, and DW collected data. AS and T-CT analyzed data. T-CT and DW wrote the manuscript and all authors contributed feedback and/or edits. All authors contributed to the article and approved the submitted version.

## Funding

This work was partly funded by a USDA AFRI grant to DW (Grant number 2022-67013-36205).

## Acknowledgments

The authors thank Makala Hammons and Natalie Roth for help with plant measurements, Yujie Chen for statistical consulting, and John Couture and Yiwei Jiang for feedback on analyses.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1229161/full#supplementary-material>

## References

- Allard, R. W., and Bradshaw, A. D. (1964). Implications of genotype-environmental interactions in applied plant breeding1. *Crop Sci.* 4, cropsci1964.0011183X000400050021x. doi: 10.2135/cropsci1964.0011183X000400050021x
- Al Makdesi, N., Ecarnot, M., Roumet, P., and Rabaté, G. (2019). A spectral correction method for multi-scattering effects in close range hyperspectral imagery of vegetation scenes: application to nitrogen content assessment in wheat. *Precis. Agric.* 20, 237–259. doi: 10.1007/s11119-018-9613-2
- Arias, F., Zambrano, M., Broce, K., Medina, C., Pacheco, H., Nunez, Y., et al. (2021). Hyperspectral imaging for rice cultivation: Applications, methods and challenges. *AIMS Agric. Food* 6, 273–307. doi: 10.3934/agrfood.2021018
- Asaari, M. S. M., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., and Scheunders, P. (2019). Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Comput. Electron. Agric.* 162, 749–758. doi: 10.1016/j.compag.2019.05.018
- Baath, G. S., Flynn, K. C., Gowda, P. H., Kakani, V. G., and Northup, B. K. (2021). Detecting biophysical characteristics and nitrogen status of finger millet at hyperspectral and multispectral resolutions. *Front. Agron.* 2, 604598. doi: 10.3389/fagro.2020.604598
- Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (SURF). *Comput. Vision Image Understand.* 110, 346–359. doi: 10.1016/j.cviu.2007.09.014
- Berry, J. C., Fahlgren, N., Pokorny, A. A., Bart, R. S., and Velez, K. M. (2018). An accurate, highthroughput method for standardizing image color profiles to improve image-based plant phenotyping. *PeerJ* 6, e5727. doi: 10.7717/peerj.5727
- Bruning, B., Liu, H., Brien, C., Berger, B., Lewis, M., and Garnett, T. (2019). The development of hyperspectral distribution maps to predict the content and distribution of nitrogen and water in wheat (*Triticum aestivum*). *Front. Plant Sci.* 10, 1380. doi: 10.3389/fpls.2019.01380
- Burnett, A. C., Anderson, J., Davidson, K. J., Ely, K. S., Lamour, J., Li, Q., et al. (2021). A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *J. Exp. Bot.* 72, 6175–6189. doi: 10.1093/jxb/erab295
- Cho, M. A., Skidmore, A., Corsi, F., van Wieren, S. E., and Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *Int. J. Appl. Earth Observ. Geoinform.* 9, 414–424. doi: 10.1016/j.jag.2007.02.001
- Cotrozzi, L., Peron, R., Tuinstra, M. R., Mickelbart, M. V., and Couture, J. J. (2020). Spectral phenotyping of physiological and anatomical leaf traits related with maize water status. *Plant Physiol.* 184, 1363–1377. doi: 10.1104/pp.20.00577
- Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sens. Environ.* 30, 271–278. doi: 10.1016/0034-4257(89)90069-2
- Din, M., Zheng, W., Rashid, M., Wang, S., and Shi, Z. (2017). Evaluating hyperspectral vegetation indices for leaf area index estimation of *Oryza sativa* L. at diverse phenological stages. *Front. Plant Sci.* 8, 820. doi: 10.3389/fpls.2017.00820
- Duursma, R. A. (2015). Plantecophys - an R package for analysing and modelling leaf gas exchange data. *PLoS One* 10, e0143346. doi: 10.1371/journal.pone.0143346
- Feng, L., Wu, B., He, Y., and Zhang, C. (2021). Hyperspectral imaging combined with deep transfer learning for rice disease detection. *Front. Plant Sci.* 12, 693521. doi: 10.3389/fpls.2021.693521
- Fu, P., Meacham-Hensold, K., Guan, K., Wu, J., and Bernacchi, C. (2020). Estimating photosynthetic traits from reflectance spectra: A synthesis of spectral indices, numerical inversion, and partial least square regression. *Plant Cell Environ.* 43, 1241–1258. doi: 10.1111/pce.13718
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720. doi: 10.1093/bioinformatics/btv428
- Gao, J., Liang, T., Liu, J., Yin, J., Ge, J., Hou, M., et al. (2020). Potential of hyperspectral data and machine learning algorithms to estimate the forage carbon-nitrogen ratio in an alpine grassland ecosystem of the tibetan plateau. *ISPRS J. Photogram. Remote Sens.* 163, 362–374. doi: 10.1016/j.isprsjprs.2020.03.017
- Gausman, H. W., and Allen, W. A. (1973). Optical parameters of leaves of 30 plant species 1. *Plant Physiol.* 52, 57–62. doi: 10.1104/pp.52.1.57
- Ge, Y., Bai, G., Stoerger, V., and Schnable, J. C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Comput. Electron. Agric.* 127, 625–632. doi: 10.1016/j.compag.2016.07.028
- Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., et al. (2017). PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ* 5, e4088. doi: 10.7717/peerj.4088
- Gewali, U. B., Monteiro, S. T., and Saber, E. (2018). Machine learning based hyperspectral image analysis: a survey. *arXiv*. doi: 10.48550/arXiv.1802.08701
- Grant, L. (1987). Diffuse and specular characteristics of leaf reflectance. *Remote Sens. Environ.* 22, 309–322. doi: 10.1016/0034-4257(87)90064-2
- He, J., Zhang, X., Guo, W., Pan, Y., Yao, X., Cheng, T., et al. (2020). Estimation of vertical leaf nitrogen distribution within a rice canopy based on hyperspectral data. *Front. Plant Sci.* 10, 1802. doi: 10.3389/fpls.2019.01802
- Henriques, F. S. (2009). Leaf chlorophyll fluorescence: Background and fundamentals for plant biologists. *Bot. Rev.* 75, 249–270. doi: 10.1007/s12229-009-9035-y
- Homolová, L., Malenovský, Z., Clevers, J. G. P. W., Garcia-Santos, G., and Schaepman, M. E. (2013). Review of optical-based remote sensing for plant trait mapping. *Ecol. Complex.* 15, 1–16. doi: 10.1016/j.ecocom.2013.06.003
- Kim, S. L., Kim, N., Lee, H., Lee, E., Cheon, K.-S., Kim, M., et al. (2020). High-throughput phenotyping platform for analyzing drought tolerance in rice. *Planta* 252, 38. doi: 10.1007/s00425-020-03436-9
- Kira, K., and Rendell, L. A. (1992). “A practical approach to feature selection,” in *Machine Learning Proceedings 1992*. Eds. D. Sleeman and P. Edwards (Morgan Kaufmann), 249–256. doi: 10.1016/B978-1-55860-247-2.50037-1
- Liland, K. H., Mevik, B.-H., and Wehrens, R. (2021). *pls: Partial Least Squares and Principal Component Regression*.
- Lin, M.-Y., Lynch, V., Ma, D., Maki, H., Jin, J., and Tuinstra, M. (2022). Multi-species prediction of physiological traits with hyperspectral modeling. *Plants* 11, 676. doi: 10.3390/plants11050676
- MATLAB (2018). *version 9.7.0.1190202 (R2019b)* (The MathWorks Inc.).
- McCouch, S. R., Wright, M. H., Tung, C.-W., Maron, L. G., McNally, K. L., Fitzgerald, M., et al. (2016). Open access resources for genome-wide association mapping in rice. *Nat. Commun.* 7, 10532. doi: 10.1038/ncomms10532
- Meacham-Hensold, K., Fu, P., Wu, J., Serbin, S., Montes, C. M., Ainsworth, E., et al. (2020). Plotlevel rapid screening for photosynthetic parameters using proximal hyperspectral imaging. *J. Exp. Bot.* 71, 2312–2328. doi: 10.1093/jxb/eraa068
- Meacham-Hensold, K., Montes, C. M., Wu, J., Guan, K., Fu, P., Ainsworth, E. A., et al. (2019). Highthroughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. *Remote Sens. Environ.* 231, 111176. doi: 10.1016/j.rse.2019.04.029
- Mertens, S., Verbraeken, L., Sprenger, H., Demuynck, K., Maleux, K., Cannoot, B., et al. (2021). Proximal hyperspectral imaging detects diurnal and drought-induced changes in maize physiology. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.640914
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2021). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, TU Wien. 7–9.
- Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 282, 60–72. doi: 10.1016/j.plantsci.2019.01.007
- Mishra, P., Lohumi, S., Ahmad Khan, H., and Nordon, A. (2020). Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches. *Comput. Electron. Agric.* 178, 105780. doi: 10.1016/j.compag.2020.105780
- Murchie, E., and Lawson, T. (2013). Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *J. Exp. Bot.* 64, 3983–3998. doi: 10.1093/jxb/ert208
- Muthayya, S., Sugimoto, J. D., Montgomery, S., and Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption. *Ann. N. Y. Acad. Sci.* 1324, 7–14. doi: 10.1111/nyas.12540
- Noble, W. S. (2006). What is a support vector machine? *Nat. Biotechnol.* 24, 1565–1567. doi: 10.1038/nbt1206-1565
- Oberhuber, W., Dai, Z.-Y., and Edwards, G. E. (1993). Light dependence of quantum yields of photosystem II and CO<sub>2</sub> fixation in c3 and c4 plants. *Photosynth. Res.* 35, 265–274. doi: 10.1007/BF00016557
- O'Toole, J. C., and Cruz, R. T. (1980). Response of leaf water potential, stomatal resistance, and leaf rolling to water stress. *Plant Physiol.* 65, 428–432. doi: 10.1104/pp.65.3.428
- Pandey, P., Ge, Y., Stoerger, V., and Schnable, J. C. (2017). High throughput *in vivo* analysis of plant leaf chemical properties using hyperspectral imaging. *Front. Plant Sci.* 8, 1348. doi: 10.3389/fpls.2017.01348
- Pasala, R., and Bb, P. (2020). Plant phenomics: High-throughput technology for accelerating genomics. *J. Biosci.* 45. doi: 10.1007/s12038-020-00083-w
- R Core Team (2021). *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing).
- Ren, J., Wang, R., Liu, G., Feng, R., Wang, Y., and Wu, W. (2020). Partitioned ReliefF method for dimensionality reduction of hyperspectral images. *Remote Sens.* 12, 1104. doi: 10.1023/A:1025667309714
- Robnik-Šikonja, M., and Kononenko, I. (2003). Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning* 53, 23–69. doi: 10.1023/A:1025667309714
- Rouse, J., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). “Monitoring Vegetation Systems in the Great Plains with Erts,” in *NASA Special Publication*, vol. 351, , 309.
- Sarić, R., Nguyen, V. D., Burge, T., Berkowitz, O., Trt'ilek, M., Whelan, J., et al. (2022). Applications of hyperspectral imaging in plant phenotyping. *Trends Plant Sci.* 301–315. doi: 10.1016/j.tpls.2021.12.003

- Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21, 110–124. doi: 10.1016/j.tplants.2015.10.015
- Tan, C., Du, Y., Zhou, J., Wang, D., Luo, M., Zhang, Y., et al. (2018). Analysis of different hyperspectral variables for diagnosing leaf nitrogen accumulation in wheat. *Front. Plant Sci.* 9, 674. doi: 10.3389/fpls.2018.00674
- Vergara-Diaz, O., Vatter, T., Kefauver, S. C., Obata, T., Fernie, A. R., and Araus, J. L. (2020). Assessing durum wheat ear and leaf metabolomes in the field through hyperspectral data. *Plant J.* 102, 615–630. doi: 10.1111/tpj.14636
- Verrelst, J., Camps-Valls, G., Muñoz-Mari, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W., et al. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review. *ISPRS J. Photogram. Remote Sens.* 108, 273–290. doi: 10.1016/j.isprsjprs.2015.05.005
- Vigneau, N., Ecarnot, M., Rabaté, G., and Roumet, P. (2011). Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in wheat. *Field Crops Res.* 122, 25–31. doi: 10.1016/j.fcr.2011.02.003
- Wang, D. R., Bunce, J. A., Tomecek, M. B., Gealy, D., McClung, A., McCouch, S. R., et al. (2016a). Evidence for divergence of response in indica, japonica, and wild rice to high CO<sub>2</sub> × temperature interaction. *Global Change Biol.* 22, 2620–2632. doi: 10.1111/gcb.13279
- Wang, D. R., Wolfrum, E. J., Virk, P., Ismail, A., Greenberg, A. J., and McCouch, S. R. (2016b). Robust phenotyping strategies for evaluation of stem non-structural carbohydrates (NSC) in rice. *J. Exp. Bot.* 67, 6125–6138. doi: 10.1093/jxb/erw375
- Wei, T., and Simko, V. (2021). *R package ‘corrplot’: Visualization of a Correlation Matrix*.
- Wickham, H. (2007). Reshaping data with the reshape package. *J. Stat. Softw.* 21, 1–20. doi: 10.18637/jss.v021.i12
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (New York: Springer-Verlag).
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4, 1686. doi: 10.21105/joss.01686
- Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*.
- Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214. doi: 10.1016/j.molp.2020.01.008
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5, 5087. doi: 10.1038/ncomms6087
- Yu, F., Feng, S., Du, W., Wang, D., Guo, Z., Xing, S., et al. (2020). A study of nitrogen deficiency inversion in rice leaves based on the hyperspectral reflectance differential. *Front. Plant Sci.* 11, 573272. doi: 10.3389/fpls.2020.573272
- Zhang, L., Maki, H., Ma, D., Sanchez-Gallegó, J. A., Mickelbart, M. V., Wang, L., et al. (2019). Optimized angles of the swing hyperspectral imaging system for single corn plant. *Comput. Electron. Agric.* 156, 349–359. doi: 10.1016/j.compag.2018.11.030