Improving estimation of days to maturity in field pea using RGB aerial imagery and machine learning

Harry Navasca; Aliasghar Bazrafkan; Jeong-Hwa Kim; Hannah Worral; Shailesh Raj Acharya; Lisa Piche; Qi Zhang; Rebecca J. McGee; Paulo Flores; Nonoy Bandillo

Received: 4 March 2025 | Accepted: 15 July 2025

DOI: 10.1002/ppj2.70038

The Plant Phenome Journal OPEN ACCESS

ORIGINAL ARTICLE

Improving estimation of days to maturity in field pea using RGB aerial imagery and machine learning

Harry Navasca[1] | **Aliasghar Bazrafkan**[2] | **Françoise Dalprá Dariva**[1,3] | **Jeong-Hwa Kim**[1,4] | **Hannah Worral**[5] | **Josephine Princy Johnson**[1,6] | **Shailesh Raj Acharya**[1,7] | **Lisa Piche**[1] | **Andrew Ross**[1] | **Garrett Raymon**[1] | **Qi Zhang**[1] | **Rebecca J. McGee**[8,9] | **Kevin McPhee**[10] | **Clarice J. Coyne**[9,11] | **Paulo Flores**[2] | **Nonoy Bandillo**[1,7]

[1]Department of Plant Sciences, North Dakota State University, Fargo, North Dakota, USA

[2]Department of Agricultural & Biosystems Engineering, North Dakota State University, Fargo, North Dakota, USA

[3]Department of Agronomy, Horticulture & Plant Sciences, South Dakota State University, Brookings, South Dakota, USA

[4]Department of Plant & Soil Sciences, University of Delaware, Newark, Delaware, USA

[5]North Dakota State University North Central Research & Extension Center, Minot, North Dakota, USA

[6]Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[7]Department of Crop and Soil Sciences, North Carolina State University, Raleigh, North Carolina, USA

[8]USDA-ARS Grain Legume Genetics and Physiology Research, Pullman, Washington, USA

[9]Department of Horticulture, Washington State University, Pullman, Washington, USA

[10]Department of Plant Science & Plant Pathology, Montana State University, Bozeman, Montana, USA

[11]USDA-ARS Plant Germplasm Introduction & Testing, Washington State University, Pullman, Washington, USA

Correspondence
Nonoy Bandillo, Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC, USA. Email: nbbandil@ncsu.edu

Assigned to Associate Editor Hannah Schneider.

Funding information
Northern Pulse Growers Association; North Dakota Department of Agriculture specialty Crop Block Grant Program, Grant/Award Numbers: 20-489, 21-316; USDA-NIFA, Grant/Award Numbers: Bandillo Hatch Project ND01513, Flores Hatch Project ND01488

Abstract

Accurately estimating days to maturity (DTM) is essential for assessing local adaptation and yield potential in field pea (*Pisum sativum* L.) breeding programs. However, traditional manual scoring of DTM is labor-intensive and inefficient for large-scale, multi-environment trials. To address this challenge, we developed a high-throughput, low-cost phenotyping framework using uncrewed aerial systems (UASs) equipped with red-green-blue cameras, implemented within the North Dakota State University Pulse Crop Breeding Program. This study aimed to (1) compare aerial and manual phenotyping for DTM estimation, (2) identify the optimal assessment time point, and (3) detect significant loci associated with DTM in a panel of 300 genetically diverse pea accessions. Image-derived vegetation indices (VIs) collected 71 days after planting exhibited strong correlations with manually assessed DTM. Notably, vegetation indices demonstrated higher heritability ($H^2 = 0.91$) compared to traditional

DTM scores (H^2 = 0.84). eXtreme Gradient Boosting models identified the visible atmospherically resistant index (31%), modified green-red vegetation index (17%), and redness index (13%) as the most predictive VIs. Genome-wide association mapping using these indices revealed three significant single nucleotide polymorphisms on chromosomes 3 and 5—variants not detected using traditional maturity data—highlighting the potential enhanced detection power of image-derived traits. This work demonstrates the utility of low-cost UAS platforms for scalable, non-destructive maturity estimation and illustrates their potential to uncover genetic components of economically important traits, offering new avenues for addressing missing heritability in legume breeding.

Plain Language Summary

We developed a cost-effective aerial phenotyping platform to estimate and map associated genes with days to maturity (DTM) in field peas, which can be applied to breeding programs. Using drones and red-green-blue sensors similar to regular phone cameras, weekly imaging in the field plots captures the canopy greenness from the onset of flowering to the physiological maturity stages of field pea. The goal is to identify the optimal machine learning model and the best window for estimating DTM. Aerial estimates increased the detection of genetic variants more than visual assessments, making this platform effective for quantifying plant maturity in scalable and reliable data-driven decisions.

1 | INTRODUCTION

Accurate estimation of days to maturity (DTM) is critical in field pea (*Pisum sativum* L.) breeding, influencing cultivar development for local adaptation, stress resilience, and market alignment. As a key phenological trait, DTM enables effective comparison of breeding lines during varietal development trials (Parihar et al., 2022). The timing of maturity is driven by genetic responses to environmental cues such as temperature and water availability (Bueckert et al., 2015; Sadras et al., 2012). Regional variation in germination and flowering across agroecological zones further emphasizes the need for precise maturity assessments to support yield optimization and line selection. Given its strong positive correlation with yield (Hara et al., 2023), DTM remains a vital selection criterion in breeding programs.

Field pea is well adapted to temperate climates (Warkentin et al., 2015), yet climate change presents increasing challenges, particularly during the flowering stage when heat stress can reduce productivity. Breeding efforts have therefore prioritized traits that shorten flowering and ripening periods to minimize late-season stress exposure (Miller et al., 2005). Physiological maturity in pea is typically indicated by pod and foliage color transition to yellow or brown (Gritton, 1980), with notable pod growth occurring between 80 and 100 days after planting (DAP) (McKay et al., 2003). However, manual DTM phenotyping based on visual scoring is time-consuming, subjective, and prone to observer variability (Tefera et al., 2022), limiting scalability in multi-environment trials. These limitations highlight the need for more efficient and accurate phenotyping approaches (Chawade et al., 2019; G. Yang et al., 2017).

High-throughput field phenotyping (HTFP) has emerged as a transformative tool in plant breeding, enabling non-invasive, repeatable measurement of key traits through proximal and remote sensing technologies (Tao et al., 2022). When integrated with genome-wide association (GWA) mapping, HTFP enhances our ability to unravel genotype-by-phenotype-by-environment interactions (Xiao et al., 2022). While ground-based platforms can capture detailed images, they often lack the speed and spatial coverage needed for large-scale trials (Haghighattalab et al., 2016). In contrast, uncrewed aerial systems (UASs) provide a cost-effective and scalable means to capture high-resolution imagery over extensive field plots (Song et al., 2021), supporting frequent and biologically meaningful trait measurements.

The utility of aerial phenotyping depends on sensor type and the choice of vegetation indices (VIs) coupled with appropriate machine learning (ML) models (Govaichelvan et al., 2023). Multispectral (MS) sensors, which leverage red-edge

and near-infrared reflectance, are powerful tools for detecting senescence-related changes (Bazrafkan et al., 2023; S. Zhang et al., 2022), but often sacrifice spatial resolution due to larger pixel sizes and limited band specificity (Z. Khan et al., 2018; López-García et al., 2022). Red-green-blue (RGB) sensors, by contrast, offer high spatial resolution at a lower cost and are well-suited to capturing morphological traits such as color changes during maturation (Volpato et al., 2021). Their compact form factor and ease of deployment make RGB sensors particularly attractive for frequent time-point imaging in field trials. In this study, RGB imagery was used to monitor greenness decay, which is a key indicator of physiological maturity in pea at fine temporal detail.

Selecting robust statistical and ML models is essential for extracting trait information from aerial imagery. Tree-based ensemble methods such as random forest (RF) are widely used due to their ability to model non-linear relationships and resist overfitting by averaging multiple decision trees (Berk, 2016; Zou et al., 2015). Despite their strengths, RF models can become computationally intensive and less interpretable as complexity grows (Duroux & Scornet, 2018). Extremely randomized trees (Extra Trees) improve training efficiency by introducing additional randomness during tree construction (Galelli & Castelletti, 2013), and have demonstrated success in agricultural trait prediction (N. Khan et al., 2022).

The eXtreme Gradient Boosting (XGBoost) represents a state-of-the-art ML approach that builds an ensemble of optimized weak learners using regularized boosting techniques. XGBoost has shown high performance in pea phenotyping tasks such as lodging classification (Bazrafkan et al., 2024). Similarly, adaptive boosting (Adaboost) improves model accuracy by emphasizing misclassified samples, though it can be sensitive to noise and outliers (Sagi & Rokach, 2018). While classification approaches have been applied to maturity stage prediction, there is a gap in leveraging regression-based modeling to identify continuous VI predictors of DTM in field pea. This study introduces a framework that utilizes RGB-derived VIs in combination with ML regression for fine-scale maturity estimation.

To our knowledge, this is the first report to combine RGB-based aerial phenotyping with genome-wide association study (GWAS) to study DTM in field pea. Our objectives were to (1) compare DTM estimates from aerial and manual phenotyping across multiple time points, (2) determine the optimal time point for DTM estimation using aerial VIs, and (3) identify genetic loci associated with maturity using GWAS based on aerial image-derived indices. This integrated framework offers a low-cost, scalable solution for improving phenotyping efficiency and capturing biologically relevant variation in crop maturity, advancing the goals of field pea public breeding programs under increasing environmental constraints.

Core Ideas

- A low-cost unmanned aerial system framework was developed to estimate days to maturity (DTM) in field pea.
- Aerial estimates of plant maturity showed a strong correlation with ground-based observations.
- Aerial-derived vegetation indices showed higher heritability than ground-based DTM scores.
- Genome-wide association mapping using vegetation indices uncovered novel single nucleotide polymorphisms (SNPs) associated with maturity that were missed by conventional phenotyping.

2 | MATERIALS AND METHODS

2.1 | Location and experimental design

In this study, a set of 300 field pea accessions were selected to evaluate a range of agronomic traits, including seed yield, protein content, maturity, and plant architecture. To provide benchmarks for comparison, commercial checks including Hampton, AC Agassiz, DS Admiral (Andersen et al., 2002), and Arcadia were included. Utilizing FieldHub (Murillo et al., 2021), a randomization tool developed at North Dakota State University (NDSU), an augmented incomplete block design was created, with the repeating checks strategically arranged diagonally throughout the experiment. The study was conducted with two replications, resulting in a total of 700 evaluated plots. Each plot consisted of eight evenly spaced rows covering an area of 4 m^2, with a spacing of 1.52 m between adjacent plots to prevent overlapping canopies. This experiment was carried out at the NDSU Prosper Research Site, located 5 miles northwest of Prosper, ND (97.11707° W 46.99432° N) as shown in Figure 1.

2.2 | Field image and data acquisition

Field pea maturity phenotyping was assessed using two methods: ground phenotyping and aerial phenotyping, as illustrated in Figure 1. Ground phenotyping focused on determining the DTM by walking through the field and visually scoring each plot based on the extent of pod and leaf yellowing and browning. This assessment typically begins at the R7 growth stage and continues until all plots reach physiological maturity (Erskine et al., 1990). In this study, evaluations were conducted every 3 days, with a plot deemed physiologically mature when 90% of its leaves, stems, and pods had turned brown and leathery (Carlson-Nilsson et al., 2021). The DTM

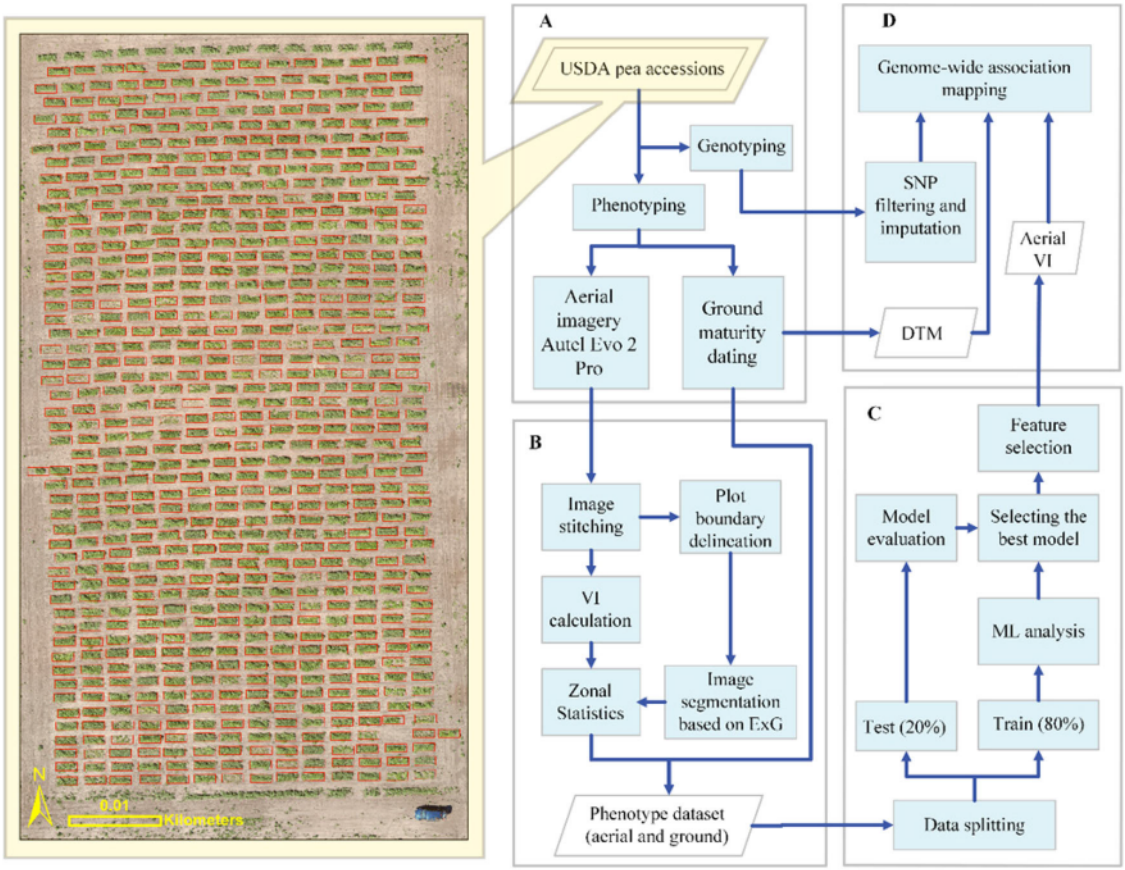

FIGURE 1 High-throughput field phenotyping (HTFP) framework for assessing field pea maturity; (A) image and data acquisition with field plot map as inset, (B) image pre-processing, (C) model validation, (D) genetic mapping. DTM, days to maturity; ExG, excess green index; ML, machine learning; SNP, single nucleotide polymorphism; VI, vegetation index; USDA, United States Department of Agriculture.

is recorded as DAP, calculated by subtracting the sowing date from the maturity date. This DTM metric served as the dependent variable for evaluating model performance in tree-based ML regression tasks.

Aerial maturity estimates were obtained using UASs through nadir imaging, designed to avoid canopy occlusions and image distortions that could impact VI calculations. Multiple flights were conducted at the R7 growth stage and beyond in the field pea plots. To enhance georeferencing accuracy, a handheld GPS-based 19984 SparkFun Real-time Kinematics (RTK) Facet (SparkFun Electronics) device was utilized to survey the experimental site and establish boundary coordinates. Orange plastic lids were placed at each corner of the experimental area to serve as ground control points, facilitating boundary traceability and precise referencing during image stitching and analysis. Checkpoints were also marked with black plastic lids to improve georeferencing of plot images. The high-throughput phenotyping workflow, as depicted in Figure 1A, commenced with the collection of a series of raw, geotagged images captured by a 6K, 20-megapixel, 1-in. complementary metal-oxide semiconductor sensor mounted on an FAA-category II Autel EVO II Pro quadcopter drone (Autel Robotics). Flights were conducted at an altitude of 50 m, with a cruising speed of 7.20 km/h, ensuring at least 80% side overlap and 85% front overlap in images. This setup yielded a camera sensor length of 10.59 m, a sensor width of 12.80 mm, and an image width of 5472 mm, resulting in a ground sample distance of 1.104 cm/pixel. Flights were scheduled between 10:00 a.m. and 2:00 p.m. to minimize shadow effects. Variations in sky conditions, such as cloud cover, were addressed using MAPIR calibration ground target version 2 (MAPIR Inc.) to maintain consistent reflectance during image collection. Additionally, a handheld Dr. Meter LX1330B digital light meter recorded illuminance values at each collection time point to aid in the reflectance correction of the images. A summary of UAS platform specifications, flight parameter settings, and experimental site information is shown in Table S1.

2.3 | Image pre-processing: Segmentation, VI calculation, and extraction

A series of plot-level images captured by the UAS were stitched and radiometrically corrected using Pix4Dmapper (Pix4D, version 4.9.0) software. This process produced a

raster known as the orthomosaic image, which was then imported into ArcGIS Pro version 3.1.2 (ESRI Inc., version 3.1.2) for further feature extraction. The current study selected 20 VIs derived from the visible light (RGB) of the electromagnetic radiation spectrum to effectively predict traits related to greenness and senescence in the test genotypes. The green-dominated indices are used for monitoring greenness, canopy cover, and crop vigor (Meyer & Neto, 2008; Woebbecke et al., 1995). The red-based indices are defined by the red band usually absorbed by plants for photosynthesis and are capable of profiling senescence, stress, soil, and crop delineation (Kawashima & Nikatani, 1998). The blue-based index is suitable for delineating vegetation pixels from wet or shadowed backgrounds (Meyer & Neto, 2008). Composite indices derived from modified expression of RGB bands excel in lighting correction under variable conditions (Woebbecke et al., 1995), soil-vegetation discrimination (Guijarro et al., 2010), estimating chlorophyll content in the absence of near-infrared band (Hunt et al., 2005, 2011), and capturing color transition during senescence phase (Meyer & Neto, 2008). Overall, this set of VI measures the extent of greenness in field pea maturity. A summary of VI notation and mathematical expressions is shown in Table S2.

Prior to generating the dataset, the three bands—RGB—were normalized by dividing each digital number (DN) by the total bands, ensuring standardized computed values. In total, 20 image-derived VI datasets were generated as aerial VI estimates. Among these, the excess green (ExG) index (Woebbecke et al., 1995) proved effective in distinguishing vegetation from soil by removing background interference, a technique previously applied in crops like cabbage (*Brassica rapa* L. subsp. *pekinensis*), radish (*Raphanus sativus* L.) (D. Kim et al., 2018), and common bean (*Phaseolus vulgaris* L.) (Parker et al., 2020). The ExG derived from the raw orthomosaic image served as the input raster for the reclassify geoprocessing tool within ArcGIS, enhancing the accuracy of vegetation pixel extraction. Analysis of the ExG histogram identified an optimal threshold of 0.65 for image classification between soil/background and vegetation pixels. This was confirmed by overlaying the results with a true-color orthomosaic image using the swipe function in ArcGIS Pro. Using the reclassify tool, DNs exceeding 0.65 were distinguished as vegetation, while DNs below were 0.65 distinguished as soil/background, resulting in a binarized raster that separated soil and vegetation canopy into two classes, assigned binary values of zero and one, respectively.

Using the ArcGIS raster calculator tool, the two input datasets—the binarized raster and the RGB index raster (orthomosaic image)—were processed to mask soil pixels corresponding to zero values in the binarized raster. This masking left only non-zero vegetation pixels in the output raster, now referred to as the vegetation RGB index raster, effectively removing soil and shadow effects. The zonal statistics analysis tool in the ArcGIS Pro software was employed to analyze the vegetation RGB index raster by overlaying it with the plot map polygon shape annotated with identification codes from the geodatabase. This extraction process generated a dataset table containing localized summary statistics—mean, median, count, sum, range, and 95th percentile—for the twenty VI formulas based on the visible light spectrum, as detailed in Table S2. Ultimately, these 20 extracted aerial VIs, along with DTM, were combined to form the comprehensive phenotype data.

2.4 | Model cross-validation: ML and statistical relationships

The statistical analysis framework, illustrated in Figure 2, utilized two distinct datasets to identify the most effective image-derived VI based on its accuracy in predicting maturity response in field pea. Ground estimates (visual maturity scores) were treated as dependent variables, while aerial estimates (VIs) served as predictor variables for the analysis. This resulted in a comprehensive evaluation of how well each VI correlated with the actual maturity observed in the field, ultimately guiding the selection of the most reliable indices for assessing field pea maturity.

2.4.1 | Optimal time point and ML model

Four ensemble ML models were evaluated for regression tasks: RF, Extra Trees, Adaboost, and XGBoost. Each model produced mean predictions in continuous values. RF operates by constructing multiple decision trees, where each tree is built from a subset of features and a specified node size. This randomness in feature selection and node size helps to prevent overfitting and enhances the model's ability to generalize to unseen variable relationships (Breiman, 2001). Additionally, RF allows for fine-tuning the number of decision trees and node sizes, enabling effective predictions with relative simplicity.

Adaboost, on the other hand, employs a boosting framework that iteratively combines multiple single-split decision trees, known as stumps or weak learners. It assigns larger weights to less accurately predicted samples, progressively improving the model with each iteration until the final prediction is optimized (Freund & Schapire, 1997). XGBoost is another gradient-boosting algorithm that enhances predictive performance through gradient descent optimization, combining the strengths of sample learners to generate optimal model predictions (Panigrahi et al., 2023). Similar to RF, Extra Trees utilizes decision trees but incorporates randomness at each

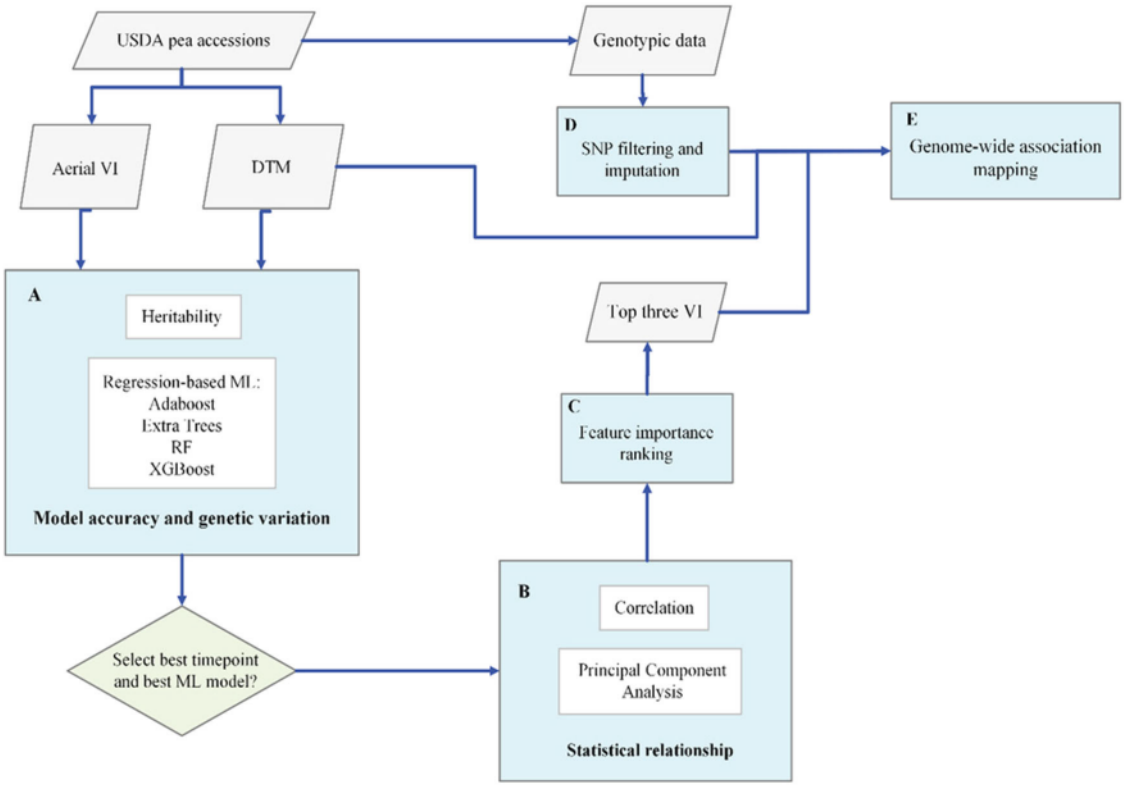

FIGURE 2 Model analysis and validation framework for selecting optimal timepoint-machine learning (ML) model, and identifying the top three vegetation indices (VIs) indicative of pea ground maturity. Ground days to maturity [DTM] and image-derived estimates (aerial VIs) were (A) compared based on broad-sense heritability and fed into ML regression models to select the best model at a specified time point. Twenty indices from the selected time point were (B) evaluated and (C) ranked using the feature importance function of a selected model to determine the top three most informative VIs. United States Department of Agriculture (USDA) pea accessions were filtered and imputed (D) to generate high-quality single nucleotide polymorphisms (SNPs) to be used as genotype data. (E) Genome-wide association mapping was employed using the selected VIs, DTM, and filtered genotype data. Adaboost, adaptive boosting; Extra Trees, extremely randomized tree; RF, random forest; XGBoost, eXtreme Gradient Boosting.

node without replacement, adding an extra layer of variability (Geurts et al., 2006).

In this analysis, aerial image-derived VIs and DTM were designated as predictor and dependent variables, respectively. Any plot IDs with missing phenotype data were to be filtered out. An 80:20 split ratio was applied to create training and testing datasets prior to model tuning.

Effective model fitting relies on fine-tuned hyperparameters, often optimizing the accuracy in predicting DTM. The hyperparameter tuning strategy in this study employs the grid search technique with cross-validation, where the combination of model configurations is identified based on the assigned range values of hyperparameters. Using the ML models, 700 test observations, 20 aerial estimates (predictor variables), and fivefold cross-validation, the RF and Extra Trees typically recommend the search range n_estimators of 100–300, with max_depth of up to 10, achieving a performance and error rate (Breiman, 2001; Geurts et al., 2006; Probst et al., 2019). Conversely, XGBoost and Adaboost are sensitive to overfitting and suggest a range of 50–200 estimators with max_depth of 3–7 (or stumps in Adaboost) (Freund & Schapire, 1997) at the maximum learning rate of 0.2–0.1. Hyperparameter tuning and feature importance ranking for each model were conducted using GridSearchCV and the variable importance function from the Scikit-learn ML ensemble module (Pedregosa et al., 2011) within the Jupyter Notebook interface (version 6.5.4) in a Python 3.11 environment. The evaluation of the tree-based regression models was conducted using the following equations:

$$R^2 = 1 - \left\{ \left[\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \right] \cdot \left[\sum_{i=1}^{n} \left(Y_i - \hat{Y}\right)^2 \right]^{-1} \right\} \quad (1)$$

$$\text{RMSE} = \left\{ \left[\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right) \right] \cdot [n]^{-1} \right\}^{0.5} \quad (2)$$

where R^2 is the coefficient of determination, which measures the goodness of fit of an ML model, RMSE is the root mean square error, Y_i is the actual measured dependent values, $\hat{Y}_i$ is the predicted dependent values, $\hat{Y}$ is the expected value of the

measured value, and n is the number of samples. The model with the lowest value RMSE and the highest R^2 was specified to generate a plot ranking the feature relative importance with DTM among different time points. The optimal number of model inputs was obtained by running all 20 indices across all models and iteratively removing one index at a time using the backward feature elimination with leave-one-out cross-validation method to search for the lowest RMSE value. Using the optimal model, variable importance ranking from lowest to highest was executed to identify the top three indices to be used as aerial-based phenotype data for DTM.

2.4.2 | Ground- and aerial-based heritability estimates

Variance components were estimated based on a linear mixed model summarized in the equations below:

$$Yij = \mu + r_i + g_j + \varepsilon_{ij} \quad (3)$$

$$H^2 = \{\sigma^2{}_{\text{genotype}}\} \bullet \{\sigma^2{}_{\text{gentoype}} + [\sigma^2{}_{\text{error}} \bullet (r)^{-1}]\}^{-1} \quad (4)$$

where Y_{ij} is the maturity value of observation plots of ith replicate and jth genotype, μ is the overall population mean, r_i is the fixed term effect of ith replicate, g_j is the random term effect of jth genotype, and ε_{ij} is the random residual error term associated with ith replicate and jth genotype. Heritability (H^2) measures the ratio of observed phenotypic variation in a particular trait within the population that is attributed to genetic factors. Broad sense heritability was calculated by dividing the genotypic variance with the total phenotypic variance. The total phenotypic variance consisted of the sum of two components: genotypic and residual error variance, which is divided by the total number of replications (r). The ASReml-R package (Butler et al., 2017) fitted the linear mixed model using Equation (3) and calculated the heritability using Equation (4) for ground and aerial trait estimates.

2.4.3 | Feature similarity and correlation analysis

Rescaled phenotypic data were used to perform principal component analysis (PCA), with the first two principal components (PC1 and PC2) selected based on the highest proportion of explained variance across all features. DTM was assessed through direct field observation, conducted every 3 days by visually scoring plots for signs of maturity, such as browning of leaves and pods, following the method described by Erskine et al. (1990) and Carlson-Nilsson et al. (2021). PCA outputs included variance loadings (eigenvalues) and directions (eigenvectors), computed from standardized data. Features with high variance loadings and eigenvectors closely aligned with the DTM vector were interpreted as having strong similarity and were grouped accordingly with ground-truth maturity scores, following an approach similar to soybean (*Glycine max* L.) maturity evaluation (C. Yang et al., 2025). A comprehensive overview of trait relationships and sample clustering is illustrated in the two-dimensional PCA biplot.

The Pearson correlation method was then used to analyze the rescaled datasets, visualizing the results with a heat map that displayed correlation values ranging from −1 (highly negatively correlated) to 1 (highly positively correlated). VIs that exhibited a strong correlation with DTM were identified as the most informative and influential. To assess the statistical significance of all variables, a t-test was conducted, using a threshold p-value of 0.05.

2.5 | GWA mapping

Using the ground collected-DTM and the top three highly heritable aerial VIs, GWA mapping was performed. The single nucleotide polymorphism (SNP) marker dataset was generated through genotyping-by-sequencing, as described by Al Bari et al. (2023). The SNP data were filtered using Tassel version 5.0 software (Bradbury et al., 2007) with specific criteria: a minor allele frequency threshold of 5% (<0.05), heterozygosity set at 80%, and a missingness threshold capped at 20%. This filtering process refined the genotype data to 19,826 SNP markers, which were then utilized for GWA mapping implemented in Genome Association and Prediction Integrated Tool (Wang & Zhang, 2021) in R environment.

Marker-trait associations were tested using a linear mixed model as follows:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{C\gamma} + \mathbf{Zu} + \boldsymbol{\varepsilon} \quad (5)$$

where $\mathbf{y}$ is a vector of the best linear unbiased predictors (BLUPs), $\mathbf{X}$, $\mathbf{C}$, and $\mathbf{Z}$ are design matrix for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\mathbf{u}$, respectively. $\boldsymbol{\beta}$ is a vector of fixed marker effects; $\boldsymbol{\gamma}$ is a vector of subpopulation effects; $\mathbf{u}$ is a vector of the random additive genetic background effects associated with the lines; $\boldsymbol{\varepsilon}$ is the unexplained variation. PCA was conducted to summarize the genetic structure and variation present in the collection. To control spurious association due to population structure, the first three PCs were used in the standard mixed linear model (MLM), as shown in (5).

The method of Li and Ji (2005), which accounts for linkage disequilibrium among markers, was used to calculate a comparison-wise error rate and control the experiment-wise error rate. The effective number of independent tests (Meff)

was calculated from the correlation matrix and eigenvalue decomposition of 19,826 SNP markers. The test criteria were adjusted with the following equation (Šidák, 1967):

$$\alpha_P = 1 - (1 - \alpha_e)^{1/\text{Meff}} \quad (6)$$

where, α_P is the computed comparison-wise error rate, and α_e is the trial-wise error rate. In this study, trial-wise error rate was $\alpha_e = 0.05$.

2.6 | Candidate gene annotation

A range within the ±100 kb for upstream and downstream of the SNP physical location was established as a search window (200 kb) in selecting candidate genes (Pavan et al., 2022). Annotated genes within the search region were obtained from the French reference genome annotation file Pisum_sativum_v1a (Kreplak et al., 2019).

3 | RESULTS

3.1 | Heritability and ML identify the peak timing for estimating maturity

To determine the optimal time point for estimating DTM, we assessed trait heritability and ML model accuracy across multiple imaging dates. The average broad-sense heritability (H^2) of RGB-derived VIs increased by 6% at 71 DAP, declined by 3% at 85 DAP, and dropped to near nonsignificant levels by 94 DAP (Figure 3A). Heritability peaked at 65 DAP ($H^2 = 0.90$) and plateaued at 71 DAP ($H^2 = 0.91$), indicating maximal genetic differentiation among genotypes at this stage.

Model performance followed a similar trend. At 71 DAP, the average R^2 across all ML regression models increased by 14%, with XGBoost achieving the highest performance ($R^2 = 0.60$; Figure 3B). Standard deviation of R^2 values from fivefold cross-validation decreased substantially (26%–47%) from 46 to 71 DAP, followed by a sharp increase (219%–226%) at 85 DAP (Table S3), indicating higher model stability at 71 DAP. After 79 DAP, accuracy declined due to reduced variation in canopy greenness (Figure 3C), which limited model signal. Based on heritability, ML performance, and prediction stability, 71 DAP was selected as the optimal time point for phenotyping maturity. Additionally, the top three VIs at this time point exhibited the lowest RMSE in predicting DTM (Figure S1).

3.2 | Aerial-derived vegetation indices are highly heritable

At the optimal imaging date (71 DAP), 14 of the 20 RGB-derived VIs were positively correlated with ground-based DTM (Figure 4A). Modified green-red vegetation index (MGRVI), GRVI, and visible atmospherically resistant index (VARI) had the strongest positive correlations ($r = 0.76$), while redness index (RI) and red-green difference ratio index (IRG) had the most negative correlations ($r = -0.76$), indicating their value for distinguishing early- and late-maturing genotypes. PCA supported these relationships, with MGRVI, GRVI, VARI, MExG, and ExGR clustering near DTM in the biplot (Figure 4B). The heritability of ground-based DTM was 0.84, consistent with previous findings in field pea (Ofga & Petros, 2017). In contrast, RGB-derived VIs had higher average heritability ($H^2 = 0.91$; Figure 4C and Figure S2), indicating stronger genetic control. Notably, ExR and WI had relatively low heritability ($H^2 = 0.83$) and weak correlations with DTM ($r = 0.17$ and -0.38, respectively), reducing their utility for maturity assessment. These results underscore the advantage of high-heritability, image-derived VIs, particularly those related to canopy greenness and senescence, for high-throughput maturity phenotyping.

3.3 | Feature importance highlights key vegetation indices for maturity prediction

Feature importance rankings from ML models revealed temporal shifts in the relevance of individual VIs for predicting DTM. At early time points (46–64 DAP), XGBoost identified MGRVI, GRVI, and MExG as the most informative features (Figures S3–S5), reflecting early canopy pigment dynamics.

At 71 DAP, when phenotypic differentiation was greatest, VARI, MGRVI, and RI were the top predictors across models (Figure 5A and Figure S6). These indices are sensitive to visible reflectance changes in greenness and red channels, which correspond with early senescence. In later stages (79–94 DAP), Extra Trees outperformed other models and prioritized MExG, red-green ratio indices, and senescence-related metrics (Figures S7–S9).

Across all models, VARI had the highest overall feature importance—ranking first in RF (34%) and XGBoost (31%). MGRVI and IRG were top features in Adaboost (13%) and Extra Trees (15%), respectively (Figure 5A). VARI was consistently prioritized across three ML models, while MGRVI and IRG were important in two, reinforcing their role in maturity estimation (Figure S6). At 71 DAP, the top three

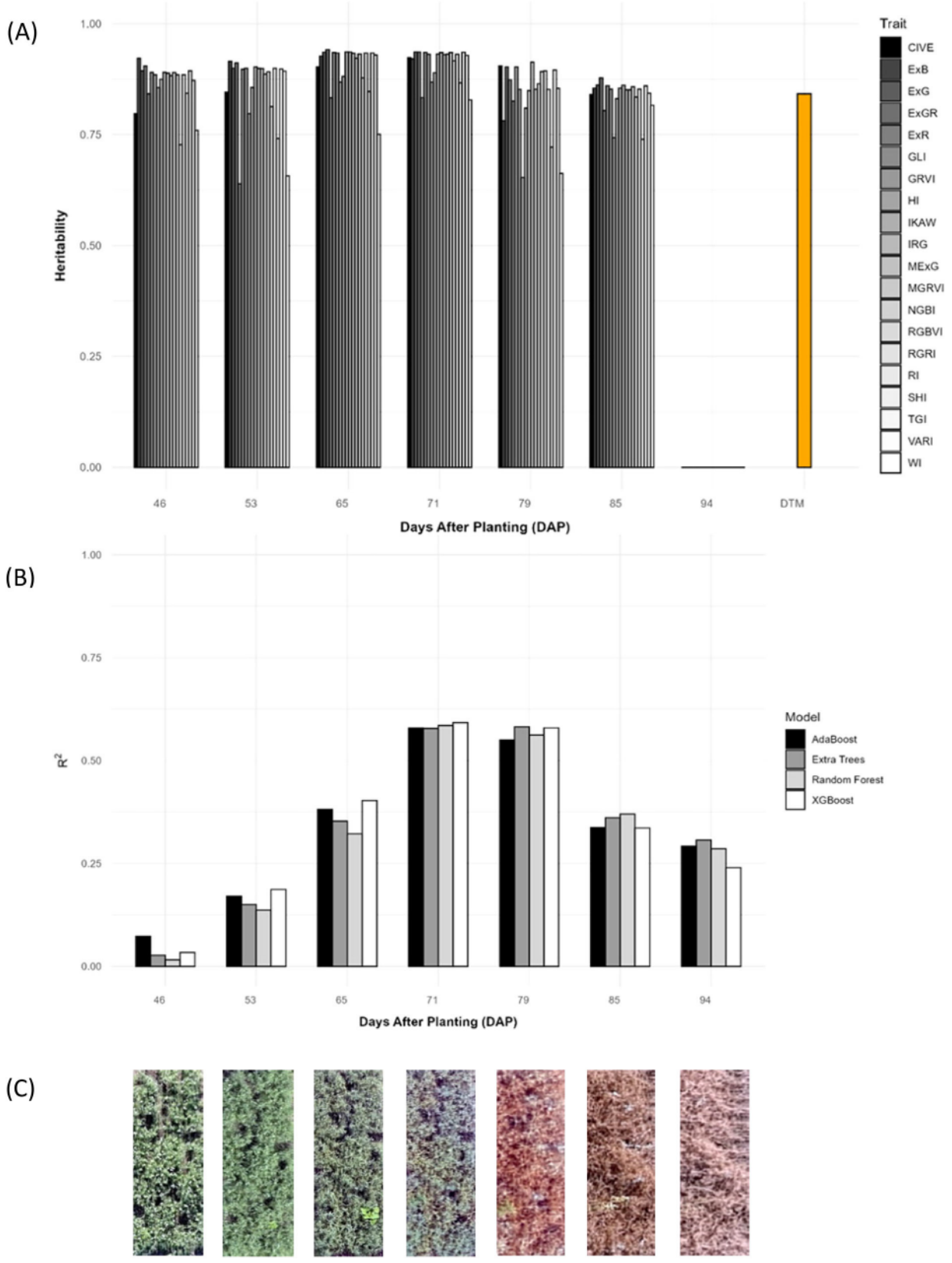

FIGURE 3 (A) Broad-sense heritability of 20 red-green-blue (RGB)-based aerial vegetation index (VI) at different time points. (B) Accuracy performance (R^2) of different machine learning (ML) regression models: adaptive boosting (Adaboost), extremely randomized trees (Extra Trees), random forest, and eXtreme Gradient Boosting (XGBoost) of all aerial-based VI at every time point. (C) Series of orthomosaic RGB images represented by a single field plot (PID#1325) showing field pea growth stages from flowering to full maturity. DTM, days to maturity.

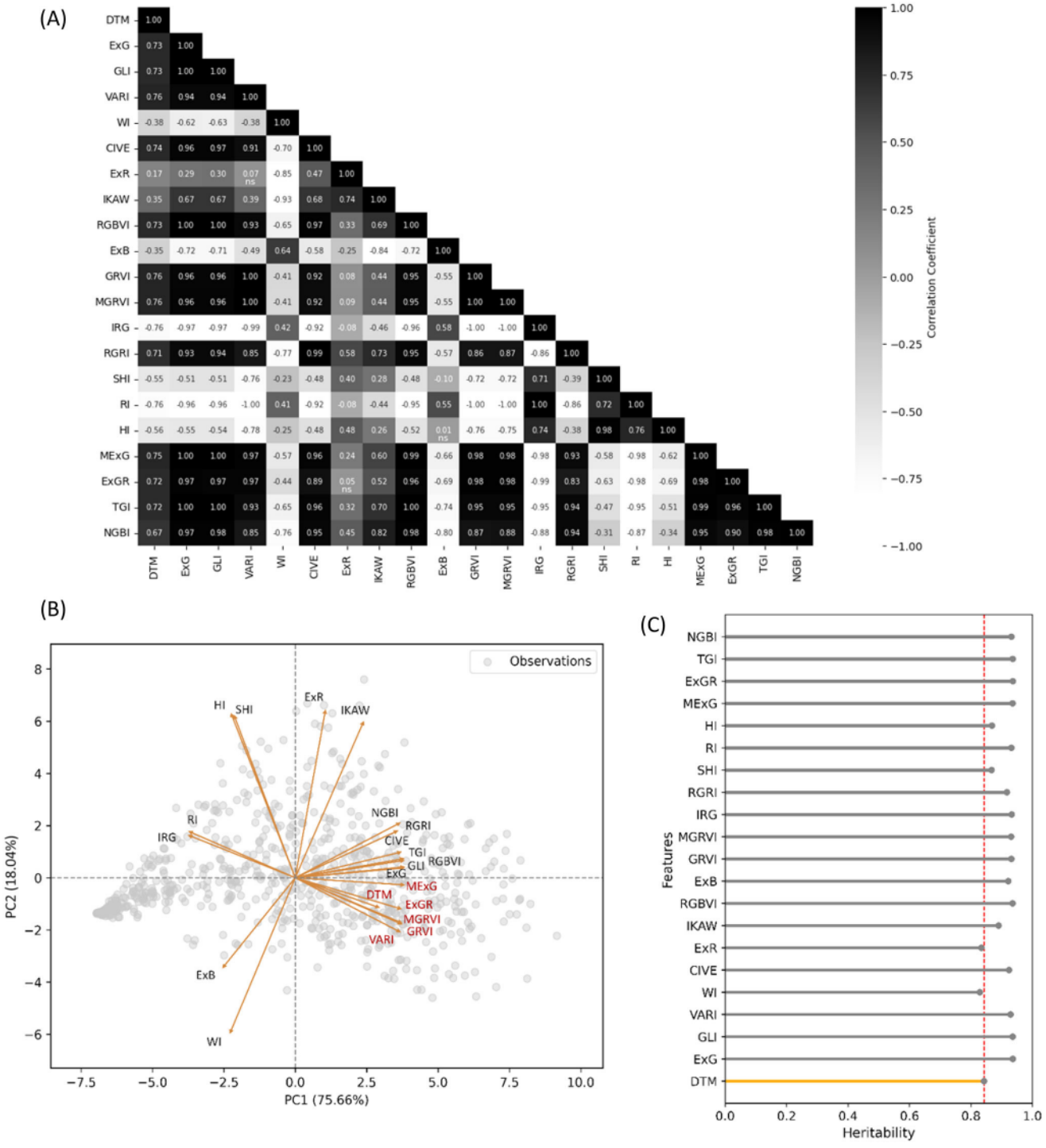

FIGURE 4 (A) Pearson correlation heat map, (B) principal component analysis (PCA)-based biplot quadrant grouping, and (C) heritability values of 20 different mean red-green-blue (RGB)-based vegetation indices (features) with the days to maturity (DTM) at 71 days to planting (DAP). Black color was designated by highly negative values ($r = -1.0$) while white color of the spectrum designated the highly positive correlation ($r = 1.0$). The red dashed line denoted the threshold and the broad sense heritability value of ground maturity (H^2, 0.84). ns, not significant at p-value = 0.05.

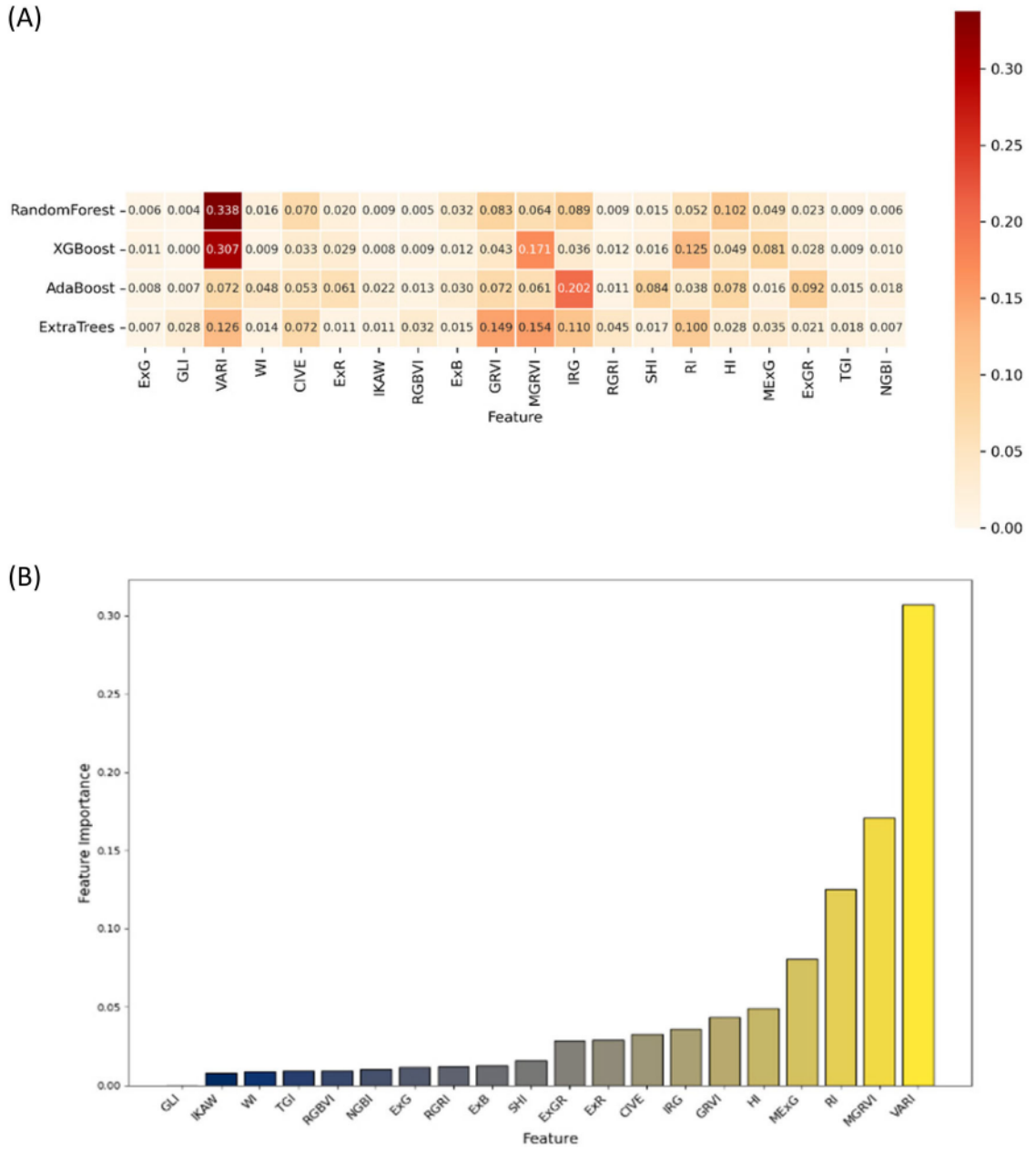

FIGURE 5 Feature importance heat map, in percentage of (A) of machine learning (ML) regression models, and (B) feature importance ranking (in ascending order) of 20 image-derived vegetation indices (VIs) using the eXtreme Gradient Boosting (XGBoost) regression model at 71 days after planting (DAP). A shade of red represented the color bar for minimum and maximum value for dark red and pale red, respectively.

indices—VARI, MGRVI, and RI—had feature importance scores of 31%, 17%, and 13%, respectively, under XGBoost (Figure 5B), which also had the highest predictive R^2 (0.60; Figure 3B).

3.4 | Aerial vegetation indices-based GWAS enhances detection power and identifies novel maturity genes

GWA mapping using aerial image-derived traits identified four quantitative trait nucleotides (QTNs) associated with pea maturity that were not detected through conventional visual maturity ratings via a MLM incorporating three PCAs and a kinship matrix. Two QTNs, S5LG3_188788921 and S5LG3_188788926, both located on chromosome 5, and S3LG5_423138033 on chromosome 3, were consistently associated with the top three VIs: VARI, MGRVI, and RI (Table 1; Figure 6). An additional QTN, S5LG3_141169470 on chromosome 5, was identified for MGRVI and RI only. Notably, none of these SNPs exceeded the GWA significance threshold ($-\log10[p] = 3.92$) for ground-based DTM scores, highlighting the increased mapping power afforded by aerial VI-based phenotyping. However, genotypic differences

TABLE 1 Significant single nucleotide polymorphism (SNP) and previously-mapped genes detected by image-derived aerial vegetation indices (VIs) relating to days to maturity (DTM) in field pea.

Traits	SNP	Chr No.	Position (Mb)	MAF	Effect	$-\log_{10}[p]$	Gene model	Annotation
VARI	S5LG3_188788921	5	188788921	0.36	−0.03	5.899363798	*Psat5g105760*	Heavy metal-associated isoprenylated plant protein 16-like
							Psat5g105720	RRM domain-containing protein
	S5LG3_188788926	5	188788926	0.36	−0.03	5.899363798	*Psat5g105760*	Heavy metal-associated isoprenylated plant protein 16-like
							Psat5g105720	RRM domain-containing protein
	S3LG5_423138033	3	423138033	0.29	0.03	4.136488107	*Psat3g199400*	Late exocytosis + associated with Golgi transport
							Psat3g199440	Vesicle transport v-SNARE protein N-terminus
							Psat3g199480	Pectinesterase
							Psat3g199520	PPR repeat family
							Psat3g199560	PPR repeat family
							Psat3g199600	Emp24/gp25L/p24 family/GOLD
							Psat3g199640	EXS family
							Psat3g199680	SPX domain
							Psat3g199720	SPX domain
MGRVI	S5LG3_188788921	5	188788921	0.36	−0.05	6.655774067	*Psat5g105760*	Heavy metal-associated isoprenylated plant protein 16-like
							Psat5g105720	RRM domain-containing protein
	S5LG3_188788926	5	188788926	0.36	−0.05	6.655774067	*Psat5g105760*	Heavy metal-associated isoprenylated plant protein 16-like
							Psat5g105720	RRM domain-containing protein
	S3LG5_423138033	3	423138033	0.29	0.03	4.245303241	*Psat3g199400*	Late exocytosis + associated with Golgi transport
							Psat3g199440	Vesicle transport v-SNARE protein N-terminus
							Psat3g199480	Pectinesterase
							Psat3g199520	PPR repeat family
							Psat3g199560	PPR repeat family
							Psat3g199600	Emp24/gp25L/p24 family/GOLD

(Continues)

TABLE 1 (Continued)

Traits	SNP	Chr No.	Position (Mb)	MAF	Effect	$-\log_{10}[p]$	Gene model	Annotation
							Psat3g199640	EXS family
							Psat3g199680	SPX domain
							Psat3g199720	SPX domain
	S5LG3_141169470	5	141169470	0.44	−0.05	4.200511334	*Psat5g078640*	Dipeptidyl peptidase IV (DPP IV) N-terminal region
							Psat5g078680	Phosphatidylinositol 3- and 4-kinase
RI	S5LG3_188788921	5	188788921	0.36	0.02	6.686968275	*Psat5g105760*	Heavy metal-associated isoprenylated plant protein 16-like
							Psat5g105720	RRM domain-containing protein
	S5LG3_188788926	5	188788926	0.36	0.36	6.686968275	*Psat5g105760*	Heavy metal-associated isoprenylated plant protein 16-like
							Psat5g105720	RRM domain-containing protein
	S3LG5_423138033	3	423138033	0.29	−0.02	4.239950042	*Psat3g199400*	Late exocytosis + associated with Golgi transport
							Psat3g199440	Vesicle transport v-SNARE protein N-terminus
							Psat3g199480	Pectinesterase
							Psat3g199520	PPR repeat family
							Psat3g199560	PPR repeat family
							Psat3g199600	Emp24/gp25L/p24 family/GOLD
							Psat3g199640	EXS family
							Psat3g199680	SPX domain
							Psat3g199720	SPX domain
	S5LG3_141169470	5	141169470	0.44	0.03	4.239257815	*Psat5g078640*	Dipeptidyl peptidase IV (DPP IV) N-terminal region
							Psat5g078680	Phosphatidylinositol 3- and 4-kinase

Abbreviations: MGRVI, modified green-red vegetation index; RI, redness index; RRM, RNA recognition motif; VARI, visible atmospherically resistant index.

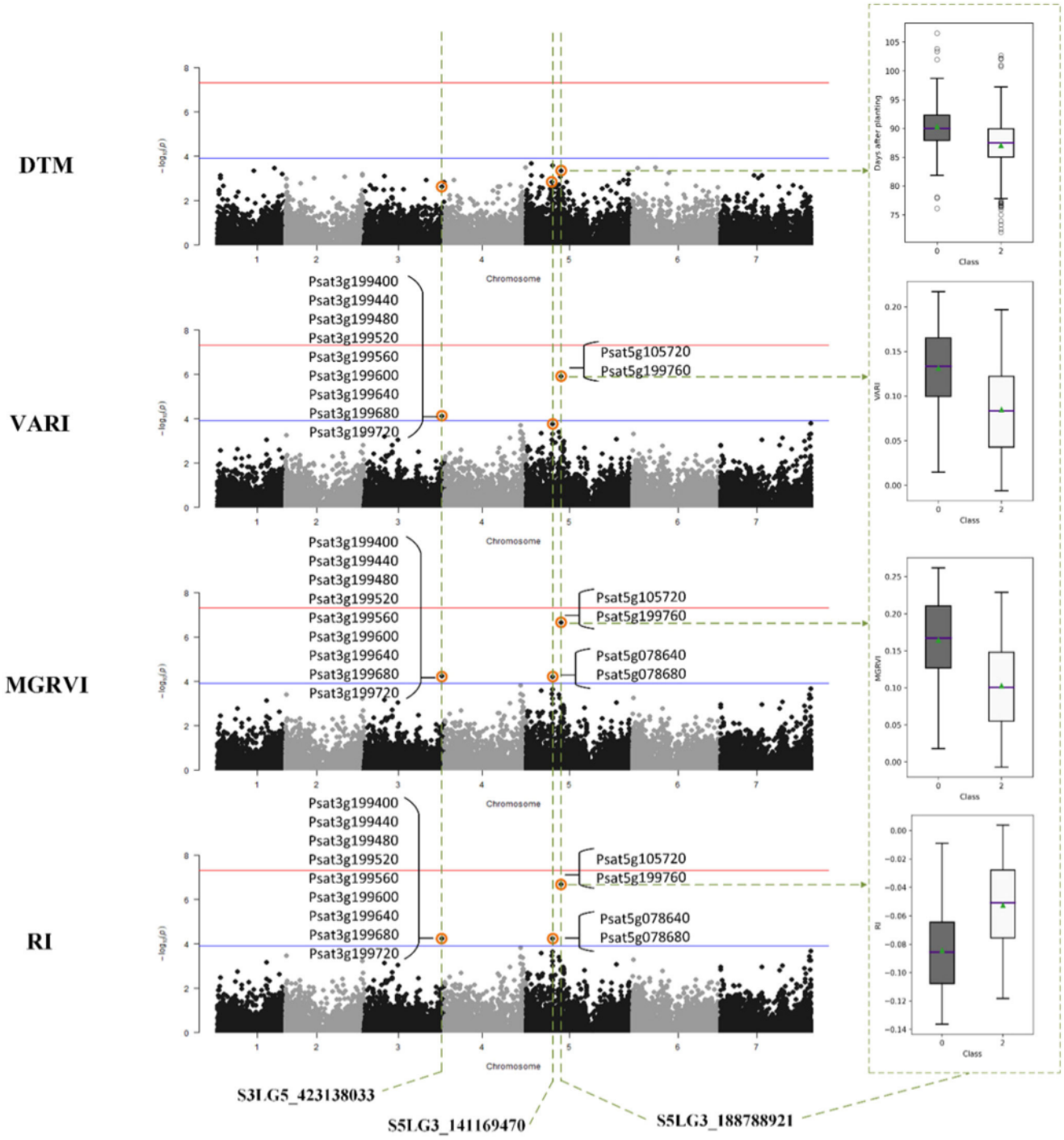

FIGURE 6 Manhattan plots of the ground maturity and the top three indices selected from best regression tree-based model (eXtreme Gradient Boosting [XGBoost]) at 71 days after planting (DAP). Traits: DTM (days to maturity), VARI (visible atmospherically resistant index), MGRVI (modified green-red vegetation index), and RI (redness index). Threshold value represented by a solid blue horizontal line with transformed *p*-value ($-\log_{10}$[*p*-value]) obtained by Li and Ji (2005) and Šidák (1967) correction ≈3.92. Broken green vertical line projected the marker with single nucleotide polymorphism (SNP) position (423, 141, 188 Mb) in chromosomes 3 and 5, respectively, extended across image derived indices and ground maturity. Encircled in orange are the significant signals surpassing the *p*-value threshold (blue solid horizontal line). Mean comparison using *t*-test at *p*-value of 0.05 of two homozygous classes (0 and 2) at the chromosome 5 (188 Mb) across all traits.

in average DTM were significant for the VI-associated SNP S5LG3_188788921 (Figure 6), indicating potential biological relevance of this locus in regulating maturity.

In total, 13 stress-related genes were located near the identified QTNs. Nine genes were found in proximity to S3LG5_423138033 on chromosome 3, while S5LG3_188788921 and S5LG3_141169470 on chromosome 5 were each associated with two candidate genes. The QTN S5LG3_188788921 is linked to *Psat5g10720*, an RNA recognition motif domain-containing protein—implicated in abiotic stress responses and developmental regulation (Shi et al., 2017). QTN S5LG3_141169470 is associated with *Psat5g078680*, a gene encoding phosphatidylinositol 3- and 4-kinase (PI3K), which plays roles in cell division, vacuolar reorganization during pollen development (Lee et al., 2008), and reactive oxygen species response via endocytosis in *Arabidopsis thaliana* (Serrazina et al., 2014). Meanwhile, S3LG5_423138033 is linked to *Psat3g199480*, annotated as a pectinesterase (PE), known for its involvement in cell wall remodeling (Derbyshire et al., 2007; Pelletier et al., 2010; Wu et al., 2018) and pathogen defense through pectin methyl esterification (Bethke et al., 2013; Lionetti et al., 2012). A summary of the gene models and their functional annotations is provided in Table 1.

4 | DISCUSSION

This study underscores the significant potential of HTFP platforms utilizing off-the-shelf RGB sensors to advance public breeding programs. Compared to manual phenotyping, HTFP offers more accurate and consistent trait estimates, reducing the inherent subjectivity of visual ratings. Aerial VIs derived from the visible light spectrum enable precise temporal monitoring of plant growth and development, particularly for assessing maturity (Vélez et al., 2023). Our results align with prior HTFP studies in field pea (Vargas et al., 2019; C. Zhang et al., 2021), which demonstrated increased correlations between aerial VIs and DTM.

Heritability estimates for 20 image-derived VIs surpassed those based on ground-based DTM measurements by 8.3%, suggesting aerial phenotyping captures genetic variation more effectively. ML model performance varied across developmental stages: predictive accuracy was lower during early (46–53 DAP) and late (85–94 DAP) phases, likely due to heterogeneous onset of leaf yellowing and uniform senescence across plots. Peak accuracy across models (Adaboost, Extra Trees, XGBoost, and RF) occurred at 71 DAP, effectively differentiating matured from immature plots. This pattern is consistent with Volpato et al. (2021), who reported maximal predictive power at similar maturity stages, followed by a decline concurrent with loss of leaf greenness.

Green-dominated VIs ranked highest in feature importance across ML models, highlighting their robustness in characterizing canopy color under varying illumination. RF and XGBoost models prioritized the VARI, emphasizing its utility in capturing corrected canopy greenness and red-green balance while minimizing atmospheric interference (Larrinaga & Brotons, 2019). Adaboost favored the IRG, sensitive to pigment transitions linked to senescence and stress. Red-dominated indices were more prominent in XGBoost and Extra Trees models, whereas blue-based indices contributed less consistently. XGBoost was selected as the optimal model due to its superior regularization capacity, reducing overfitting relative to RF. These ML approaches have also demonstrated efficacy in predicting related traits such as flowering time (Deva et al., 2023) and leaf age (Bai et al., 2023). PCA further identified VARI, MGRVI, and RI as the most informative VIs, with MGRVI and RI providing complementary yet contrasting signals relevant to maturity.

GWA mapping revealed that drone-derived traits outperformed visual ratings in detecting significant SNPs associated with maturity. Key QTNs identified at ~188 and ~141 Mbp on chromosome 5, and ~423 Mbp on chromosome 3, showed additive effects on VI slopes (VARI, MGRVI) correlating with maturity progression. These loci were linked to genes implicated in abiotic stress adaptation, critical for plant defense (Krinke et al., 2007; Takahashi et al., 2017). Similar associations on these chromosomes have been reported by Gali et al. (2019), supporting our findings. Additionally, our results corroborate Tar'an et al. (2004), who reported that reduced DTM combined with aerial VI improvements positively affect grain yield in field pea.

The integration of RGB-based VIs with ML techniques proved effective in enhancing maturity estimation and identifying relevant genetic markers. Although this platform facilitates large-scale screening for maturity traits, limitations persist. These include within-line heterogeneity, weed interference, and uneven germination, echoing challenges described by Volpato et al. (2021) and Al Bari et al. (2023). Furthermore, variation in RGB image quality due to fluctuating field illumination can impact greenness estimation (C. Kim et al., 2006; Chopin et al., 2018). Advanced image correction methods (C. Kim et al., 2006) and the integration of neural-network-based object detection with VI calculations (Hashim et al., 2021) could substantially improve vegetation-soil discrimination accuracy.

A potential limitation of this study is its single-year, single-location design, which may limit the broader applicability of the identified QTNs across diverse genetic backgrounds and environmental conditions. Although a 2-year experiment was initially planned, extreme drought during the first year severely constrained data collection. Therefore, validation in independent populations and across multiple environments is essential to confirm the robustness of these genomic asso-

ciations. Nevertheless, the dataset used in this study is of high quality, as evidenced by its high heritability, making it well-suited for genetic mapping.

In conclusion, continued improvements in image analysis algorithms and the use of MS or hyperspectral sensors promise to further enhance HTFP model performance. Overall, this study demonstrates that HTFP is a cost-effective and powerful approach for dissecting the genetic architecture of maturity in field pea, with significant implications for accelerating public breeding programs.

5 | CONCLUSION

This study demonstrates that HTFP using accessible RGB sensors, combined with ML, offers a robust and scalable approach to improve estimation of maturity in field pea. The use of aerial vegetation indices facilitated the identification of key genetic loci associated with maturity in field pea. While some limitations remain, such as environmental variability and image quality, ongoing technological advances and multi-environment validation will further strengthen the utility of HTFP platforms. Overall, this approach holds strong potential to accelerate genetic gain in public breeding programs like the NDSU Pulse Crop Breeding Program.

AUTHOR CONTRIBUTIONS

Harry Navasca: Data curation; formal analysis; investigation; methodology; validation; visualization; writing—original draft; writing—review and editing. **Aliasghar Bazrafkan**: Data curation; investigation; methodology; validation; writing—review and editing. **Françoise Dalprá Dariva**: Methodology; validation; writing—review and editing. **Jeong-Hwa Kim**: Data curation; investigation; writing—review and editing. **Hannah Worral**: Data curation; methodology; writing—review and editing. **Josephine Princy Johnson**: Data curation; investigation; writing—review and editing. **Shailesh Raj Acharya**: Investigation; writing—review and editing. **Lisa Piche**: Investigation; methodology; writing—review and editing. **Andrew Ross**: Data curation; investigation. **Garrett Raymon**: Data curation; writing—review and editing. **Qi Zhang**: Writing—review and editing. **Rebecca J. McGee**: Resources; writing—review and editing. **Kevin McPhee**: Writing—review and editing. **Clarice J. Coyne**: Resources; writing—review and editing. **Paulo Flores**: Conceptualization; investigation; methodology; writing—review and editing. **Nonoy Bandillo**: Conceptualization; funding acquisition; investigation; methodology; project administration; supervision; writing—review and editing.

ACKNOWLEDGMENTS

The authors would like to acknowledge the funding provided by the North Dakota Department of Agriculture through the specialty Crop Block Grant Program (20-489 and 21–316), USDA-NIFA (Bandillo Hatch Project ND01513; Flores Hatch Project ND01488), and the Northern Pulse Growers Association for continued funding support on development of the NDSU advanced pea breeding lines. Technical assistance from Andrew Ross, and hourly students is gratefully acknowledged. This investigation used resources from the Center for Computationally Assisted Science and Technology (CCAST) at North Dakota State University, Fargo, ND, USA, which were made possible in part by NSF MRI Award No. 2019077.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The SNP dataset used in this study is publicly available through the NCBI Sequence Read Archive (SRA) under the accession number PRJNA730349 and can be accessed here: https://www.ncbi.nlm.nih.gov/sra/PRJNA730349. The other datasets generated and/or analyzed during the current study are available at the following GitHub repository: https://www.github.com/harrYskA/PeaDaysToMaturity.

ORCID

Harry Navasca https://orcid.org/0009-0007-2923-8580
Aliasghar Bazrafkan https://orcid.org/0000-0002-7862-0520
Françoise Dalprá Dariva https://orcid.org/0000-0002-4259-5212
Jeong-Hwa Kim https://orcid.org/0000-0003-4501-3940
Hannah Worral https://orcid.org/0000-0003-3832-5815
Josephine Princy Johnson https://orcid.org/0009-0003-8962-1582
Shailesh Raj Acharya https://orcid.org/0009-0007-5692-9248
Lisa Piche https://orcid.org/0009-0006-2970-9648
Qi Zhang https://orcid.org/0000-0002-8277-6798
Rebecca J. McGee https://orcid.org/0000-0002-5613-1283
Clarice J. Coyne https://orcid.org/0000-0003-1374-4618
Paulo Flores https://orcid.org/0000-0002-3964-6904
Nonoy Bandillo https://orcid.org/0000-0002-5941-9047

REFERENCES

Andersen, L., Warkentin, T. D., Philipp, O., Xue, A., & Sloan, A. (2002). DS Admiral field pea. *Canadian Journal of Plant Science/Canadian Journal of Plant Science*, *82*(4), 751–752. https://doi.org/10.4141/p02-023

Bai, Y., Shi, L., Zha, Y., Liu, S., Nie, C., Xu, H., Yang, H., Shao, M., Yu, X., Cheng, M., Liu, Y., Lin, T., Cui, N., Wu, W., & Jin, X. (2023). Estimating leaf age of maize seedlings using UAV-based RGB and multispectral images. *Computers and Electronics in Agriculture*, *215*, 108349. https://doi.org/10.1016/j.compag.2023.108349

Al Bari, M. A., Fonseka, D., Stenger, J., Zitnick-Anderson, K., Atanda, S. A., Morales, M., Worral, H., Piche, L., Kim, J., Johnson, J., Saludares, R. A., Flores, P., Pasche, J., & Bandillo, N. (2023). A greenhouse-based high-throughput phenotyping platform for identification and genetic dissection of resistance to Aphanomyces root rot in field pea. *The Plant Phenome Journal*, *6*(1), e20063. https://doi.org/10.1002/ppj2.20063

Bazrafkan, A., Navasca, H., Kim, J. H., Morales, M., Johnson, J., Delavarpour, N., Fareed, N., Bandillo, N., & Flores, J. P. C. (2023). Predicting dry pea maturity using machine learning and advanced sensor fusion with unmanned aerial systems (UASs). *Remote Sensing*, *15*(11), 2758. https://doi.org/10.3390/rs15112758

Bazrafkan, A., Navasca, H., Worral, H., Oduor, P., Delavarpour, N., Morales, M., Bandillo, N., & Flores, P. (2024). Predicting lodging severity in dry peas using UAS-mounted RGB, LIDAR, and multispectral sensors. *Remote Sensing Applications Society and Environment*, *34*, 101157. https://doi.org/10.1016/j.rsase.2024.101157

Berk, R. A. (2016). Random forests. In *Statistical learning from a regression perspective* (pp. 205–258). Springer texts in statistics. Springer. https://doi.org/10.1007/978-3-319-44048-4_5

Bethke, G., Grundman, R. E., Sreekanta, S., Truman, W., Katagiri, F., & Glazebrook, J. (2013). Arabidopsis pectin methylesterases contribute to immunity against *Pseudomonas syringae*. *Plant Physiology*, *164*(2), 1093–1107. https://doi.org/10.1104/pp.113.227637

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

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

Bueckert, R. A., Wagenhoffer, S., Hnatowich, G., & Warkentin, T. D. (2015). Effect of heat and precipitation on pea yield and reproductive performance in the field. *Canadian Journal of Plant Science*, *95*(4), 629–639. https://doi.org/10.4141/cjps-2014-342

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2017). ASReml-R reference visual version 4. VSN International Ltd, hemel hempstead. HP1 1ES, UK. https://asreml.kb.vsni.co.uk/wpcontent/uploads/sites/3/ASReml-R-Reference-Visual-4.pdf

Carlson-Nilsson, U., Aloisi, K., Vågen, I. M., Rajala, A., Mølmann, J. B., Rasmussen, S. K., Niemi, M., Wojciechowska, E., Pärssinen, P., Poulsen, G., & Leino, M. W. (2021). Trait expression and environmental responses of pea (*Pisum sativum* L.) genetic resources targeting cultivation in the arctic. *Frontiers in Plant Science*, *12*, 688067. https://doi.org/10.3389/fpls.2021.688067

Chawade, A., van Ham, J., Blomquist, H., Bagge, O., Alexandersson, E., & Ortiz, R. (2019). High-throughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy*, *9*(5), 258. https://doi.org/10.3390/agronomy9050258

Chopin, J., Kumar, P., & Miklavcic, S. J. (2018). Land-based crop phenotyping by image analysis: Consistent canopy characterization from inconsistent field illumination. *Plant Methods*, *14*(1), Article 39. https://doi.org/10.1186/s13007-018-0308-5

Derbyshire, P., McCann, M. C., & Roberts, K. (2007). Restricted cell elongation in Arabidopsis hypocotyls is associated with a reduced average pectin esterification level. *BMC Plant Biology*, *7*(1), Article 31. https://doi.org/10.1186/1471-2229-7-31

Deva, C., Dixon, L. E., Urban, M. O., Ramírez-Villegas, J., Droutsas, I., & Challinor, A. J. (2023). A new framework for predicting and understanding flowering time for crop breeding. *Plants, People, Planet*, *6*(1), 197–209. https://doi.org/10.1002/ppp3.10427

Duroux, R., & Scornet, E. (2018). Impact of subsampling and tree depth on random forests. *ESAIM: Probability and Statistics*, *22*, 96–128. https://doi.org/10.1051/ps/2018008

Erskine, W., Muehlbauer, F. J., & Short, R. (1990). Stages of development in lentil. *Experimental Agriculture*, *26*(3), 297–302. https://doi.org/10.1017/s0014479700018457

Freund, Y., & Schapire, R. E. (1997). A Decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Galelli, S., & Castelletti, A. (2013). Assessing the predictive capability of randomized tree-based ensembles in streamflow modeling. *Hydrology and Earth System Sciences*, *17*(7), 2669–2684. https://doi.org/10.5194/hess-17-2669-2013

Gali, K. K., Sackville, A., Tafesse, E. G., Lachagari, V. R., McPhee, K., Hybl, M., Mikić, A., Smýkal, P., McGee, R., Burstin, J., Domoney, C., Ellis, T. N., Tar'an, B., & Warkentin, T. D. (2019). Genome-wide association mapping for agronomic and seed quality traits of field pea (*Pisum sativum* L.). *Frontiers in Plant Science*, *10*, 1538. https://doi.org/10.3389/fpls.2019.01538

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Govaichelvan, K. N., Pathmanathan, D., Zainal-Abidin, R., & Abu, A. (2023). Machine learning for major food crops breeding: Applications, challenges, and ways forward. *Agronomy Journal*, *116*(3), 1112–1125. https://doi.org/10.1002/agj2.21393

Gritton, E. T. (1980). Field pea. In W. R. Fehr & H. H. Hadley (Eds.), *Hybridization of crop plants* (pp. 347–356). ASA and CSSA. https://doi.org/10.2135/1980.hybridizationofcrops.c23

Guijarro, M., Pajares, G., Riomoros, I., Herrera, P., Burgos-Artizzu, X., & Ribeiro, A. (2010). Automatic segmentation of relevant textures in agricultural images. *Computers and Electronics in Agriculture*, *75*(1), 75–83. https://doi.org/10.1016/j.compag.2010.09.013

Haghighattalab, A., González Pérez, L., Mondal, S., Singh, D., Schinstock, D., Rutkoski, J., Ortiz-Monasterio, I., Singh, R. P., Goodin, D., & Poland, J. (2016). Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods*, *12*(1), Article 35. https://doi.org/10.1186/s13007-016-0134-6

Hara, P., Piekutowska, M., & Niedbała, G. (2023). Prediction of pea (*Pisum sativum* L.) seeds yield using artificial neural networks. *Agriculture*, *13*(3), 661. https://doi.org/10.3390/agriculture13030661

Hashim, W., Eng, L. S., Alkawsi, G., Ismail, R., Alkahtani, A. A., Dzulkifly, S., Baashar, Y., & Hussain, A. (2021). A hybrid vegetation detection framework: Integrating vegetation indices and convolutional neural network. *Symmetry*, *13*(11), 2190. https://doi.org/10.3390/sym13112190

Hunt, E. R., Cavigelli, M. A., Daughtry, C. S. T., McMurtrey, J. E., & Walthall, C. L. (2005). Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precision Agriculture*, *6*(4), 359–378. https://doi.org/10.1007/s11119-005-2324-5

Hunt, E. R., Daughtry, C. S. T., Eitel, J. U. H., & Long, D. S. (2011). Remote sensing leaf chlorophyll content using a visible band index. *Agronomy Journal*, *103*(4), 1090–1099. https://doi.org/10.2134/agronj2010.0395

Kawashima, S., & Nakatani, M. (1998). An algorithm for estimating chlorophyll content in leaves using a video camera. *Annals of Botany*, *81*(1), 49–54. https://doi.org/10.1006/anbo.1997.0544

Khan, N., Kamaruddin, M. A., Sheikh, U. U., Zawawi, M. H., Yusup, Y., Bakht, M. P., & Noor, N. M. (2022). Prediction of oil palm yield using machine learning in the perspective of fluctuating weather and soil moisture conditions: Evaluation of a generic workflow. *Plants*, *11*(13), 1697. https://doi.org/10.3390/plants11131697

Khan, Z., Rahimi-Eichi, V., Haefele, S., Garnett, T., & Miklavcic, S. J. (2018). Estimation of vegetation indices for high-throughput phenotyping of wheat using aerial imaging. *Plant Methods*, *14*(1), Article 20. https://doi.org/10.1186/s13007-018-0287-6

Kim, C., You, B., & Kim, H. (2006). Color segmentation robust to brightness variations by using B-Spline curve modeling. In *2011 IEEE/RSJ International conference on intelligent robots and systems* (pp. 4874–4879). IEEE. https://doi.org/10.1109/iros.2006.282443

Kim, D., Yun, H., Jeong, S., Kwon, Y., Kim, S., Lee, W., & Kim, H. (2018). Modeling and testing of growth status for chinese cabbage and white radish with UAV-based RGB imagery. *Remote Sensing*, *10*(4), 563. https://doi.org/10.3390/rs10040563

Kreplak, J., Madoui, M., Cápal, P., Novák, P., Labadie, K., Aubert, G., Bayer, P. E., Gali, K. K., Syme, R. A., Main, D., Klein, A., Bérard, A., Vrbová, I., Fournier, C., D'Agata, L., Belser, C., Berrabah, W., Toegelová, H., Milec, Z., … Burstin, J. (2019). A reference genome for pea provides insight into legume genome evolution. *Nature Genetics*, *51*(9), 1411–1422. https://doi.org/10.1038/s41588-019-0480-1

Krinke, O., Ruelland, E., Valentova, O., Vergnolle, C., Renou, J., Taconnat, L., Flemr, M., Burketova, L., & Zachowski, A. (2007). Phosphatidylinositol 4-kinase activation is an early response to salicylic acid in arabidopsis suspension cells. *Plant Physiology*, *144*(3), 1347–1359. https://doi.org/10.1104/pp.107.100842

Larrinaga, A., & Brotons, L. (2019). Greenness indices from a low-cost UAV imagery as tools for monitoring post-fire forest recovery. *Drones*, *3*(1), 6. https://doi.org/10.3390/drones3010006

Lee, Y., Kim, E., Choi, Y., Hwang, I., Staiger, C. J., Chung, Y., & Lee, Y. (2008). The Arabidopsis phosphatidylinositol 3-kinase is important for pollen development. *Plant Physiology*, *147*(4), 1886–1897. https://doi.org/10.1104/pp.108.121590

Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*, *95*(3), 221–227. https://doi.org/10.1038/sj.hdy.6800717

Lionetti, V., Cervone, F., & Bellincampi, D. (2012). Methyl esterification of pectin plays a role during plant–pathogen interactions and affects plant resistance to diseases. *Journal of Plant Physiology*, *169*(16), 1623–1630. https://doi.org/10.1016/j.jplph.2012.05.006

López-García, P., Intrigliolo, D., Moreno, M. A., Martínez-Moreno, A., Ortega, J. F., Pérez-Álvarez, E. P., & Ballesteros, R. (2022). Machine learning-based processing of multispectral and RGB UAV imagery for the multitemporal monitoring of vineyard water status. *Agronomy*, *12*(9), 2122. https://doi.org/10.3390/agronomy12092122

McKay, K., Schatz, B. G., & Endres, G. (2003). *Field pea production*. NDSU Extension Service.

Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, *63*(2), 282–293. https://doi.org/10.1016/j.compag.2008.03.009

Miller, P., McKay, K., Jones, C., Blodgett, S., Menalled, F., Riesselman, J., & Wichman, D. (2005). *Growing dry pea in Montana* (MT-2500502-AG). Montana State University Extension Service.

Murillo, D., Gezan, S. A., Heilman, A. M., Walk, T. C., Aparicio, J., & Horsley, R. D. (2021). FielDHub: A shiny app for design of experiments in life sciences. *Journal of Open-Source Software*, *6*(61), 3122. https://doi.org/10.21105/joss.03122

Ofga, B., & Petros, Y. (2017). Genetic variability and association among agronomic characters in selected field pea (*Pisum sativum* L.) genotypes at Bale Zone, Sinana Research Center, Oromia Region, Ethiopia. *International Journal of Genetics and Genomics*, *5*(6), 63–75. https://doi.org/10.11648/j.ijgg.20170506.11

Panigrahi, B., Kathala, K. C. R., & Sujatha, M. (2023). A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models. *Procedia Computer Science*, *218*, 2684–2693. https://doi.org/10.1016/j.procs.2023.01.241

Parihar, A. K., Yadav, R., Lamichaney, A., Mishra, R. K., Chandra, A., Gupta, D. S., Tripathi, K., Hazra, K. K., & Dixit, G. P. (2022). Field pea breeding. In D. K. Yadava, H. K. Dikshit, G. P. Mishra, & S. Tripathi (Eds.), *Fundamentals of field crop breeding* (pp. 1237–1321). Springer. https://doi.org/10.1007/978-981-16-9257-4_25

Parker, T., Palkovic, A., & Gepts, P. (2020). Determining the genetic control of common bean early-growth rate using unmanned aerial vehicles. *Remote Sensing*, *12*(11), 1748. https://doi.org/10.3390/rs12111748

Pavan, S., Delvento, C., Nazzicari, N., Ferrari, B., D'Agostino, N., Taranto, F., Lotti, C., Ricciardi, L., & Annicchiarico, P. (2022). Merging genotyping-by-sequencing data from two ex situ collections provides insights on the pea evolutionary history. *Horticulture Research*, *9*, uhab062. https://doi.org/10.1093/hr/uhab062

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830. https://doi.org/10.5555/1953048.2078195

Pelletier, S., Van Orden, J., Wolf, S., Vissenberg, K., Delacourt, J., Ndong, Y. A., Pelloux, J., Bischoff, V., Urbain, A., Mouille, G., Lemonnier, G., Renou, J., & Höfte, H. (2010). A role for pectin de-methylesterification in a developmentally regulated growth acceleration in dark-grown Arabidopsis hypocotyls. *New Phytologist*, *188*(3), 726–739. https://doi.org/10.1111/j.1469-8137.2010.03409.x

Probst, P., Wright, M. N., & Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, *9*(3), e1301. https://doi.org/10.1002/widm.1301

Sadras, V. O., Lake, L., Chenu, K., McMurray, L. S., & Leonforte, A. (2012). Water and thermal regimes for field pea in Australia and their implications for breeding. *Crop and Pasture Science*, *63*(1), 33–44. https://doi.org/10.1071/cp11321

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, *8*(4), e1249. https://doi.org/10.1002/widm.1249

Serrazina, S., Dias, F. V., & Malhó, R. (2014). Characterization of FAB1 phosphatidylinositol kinases in Arabidopsis pollen tube growth and

fertilization. *New Phytologist*, *203*(3), 784–793. https://doi.org/10.1111/nph.12836

Shi, X., Hanson, M. R., & Bentolila, S. (2017). Functional diversity of Arabidopsis organelle-localized RNA-recognition motif-containing proteins. *Wiley Interdisciplinary Reviews—RNA*, *8*(5), e1420. https://doi.org/10.1002/wrna.1420

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*(318), 626–633. https://doi.org/10.1080/01621459.1967.10482935

Song, P., Wang, J., Guo, X., Yang, W., & Zhao, C. (2021). High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *The Crop Journal*, *9*(3), 633–645. https://doi.org/10.1016/j.cj.2021.03.015

Takahashi, S., Monda, K., Higaki, T., Hashimoto-Sugimoto, M., Negi, J., Hasezawa, S., & Iba, K. (2017). Differential effects of phosphatidylinositol 4-kinase (PI4K) and 3-kinase (PI3K) inhibitors on stomatal responses to environmental signals. *Frontiers in Plant Science*, *8*, 677. https://doi.org/10.3389/fpls.2017.00677

Tao, H., Xu, S., Tian, Y., Li, Z., Ge, Y., Zhang, J., Wang, Y., Zhou, G., Deng, X., Zhang, Z., Ding, Y., Jiang, D., Guo, Q., & Jin, S. (2022). Proximal and remote sensing in plant phenomics: 20 years of progress, challenges, and perspectives. *Plant Communications*, *3*(6), 100344. https://doi.org/10.1016/j.xplc.2022.100344

Tar'an, B., Warkentin, T., Somers, D., Miranda, D., Vandenberg, A., Blade, S., & Bing, D. (2004). Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L.). *Euphytica*, *136*(3), 297–306. https://doi.org/10.1023/b:euph.0000032721.03075.a0

Tefera, A. T., Banerjee, B. P., Pandey, B. R., James, L. E., Puri, R. R., Cooray, O. O., Marsh, J., Richards, M. F., Kant, S., Fitzgerald, G. J., & Rosewarne, G. M. (2022). Estimating early season growth and biomass of field pea for selection of divergent ideotypes using proximal sensing. *Field Crops Research*, *277*, 108407. https://doi.org/10.1016/j.fcr.2021.108407

Vargas, J. J. Q., Zhang, C., Smitchger, J. A., McGee, R. J., & Sankaran, S. (2019). Phenotyping of plant Biomass and performance traits using remote sensing techniques in pea (*Pisum sativum*, L.). *Sensors*, *19*(9), 2031. https://doi.org/10.3390/s19092031

Vélez, S., Martínez-Peña, R., & Castrillo, D. (2023). Beyond vegetation: A review unveiling additional insights into agriculture and forestry through the application of vegetation indices. *J—Multidisciplinary Scientific Journal*, *6*(3), 421–436. https://doi.org/10.3390/j6030028

Volpato, L., Dobbels, A. A., Borém, A., & Lorenz, A. J. (2021). Optimization of temporal UAS-based imagery analysis to estimate plant maturity date for soybean breeding. *The Plant Phenome Journal*, *4*, e20018. https://doi.org/10.1002/ppj2.20018

Wang, J., & Zhang, Z. (2021). GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics & Bioinformatics*, *19*(4), 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Warkentin, T. D., Smýkal, P., Coyne, C. J., Weeden, N., Domoney, C., Bing, D., Leonforte, A., Xuxiao, Z., Dixit, G. P., Boros, L., McPhee, K. E., McGee, R. J., Burstin, J., & Ellis, T. H. N. (2015). Pea. In A. M. De Ron (Ed.), *Handbook of plant breeding* (pp. 37–83). Springer. https://doi.org/10.1007/978-1-4939-2797-5_2

Woebbecke, D. M., Meyer, G. E., Bargen, K. V. O. N., & Mortensen, D. A. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, *38*(1), 259–269. https://doi.org/10.13031/2013.27838

Wu, H., Bulgakov, V. P., & Jinn, T. (2018). Pectin methylesterases: Cell wall remodeling proteins are required for plant response to heat stress. *Frontiers in Plant Science*, *9*, 1612. https://doi.org/10.3389/fpls.2018.01612

Xiao, Q., Bai, X., Zhang, C., & He, Y. (2022). Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review. *Journal of Advanced Research*, *35*, 215–230. https://doi.org/10.1016/j.jare.2021.05.002

Yang, C., Yang, G., Wang, H., Li, S., Zhang, J., Pan, D., Ren, P., Feng, H., & Li, H. (2025). Identifying key traits for screening High-Yield Soybean varieties by combining UAV-Based and field phenotyping. *Remote Sensing*, *17*(4), 690. https://doi.org/10.3390/rs17040690

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Luo, D., Li, Z., Li, H., & Yang, H. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Frontiers in Plant Science*, *8*, 1111. https://doi.org/10.3389/fpls.2017.01111

Zhang, C., McGee, R. J., Vandemark, G. J., & Sankaran, S. (2021). Crop performance evaluation of chickpea and dry pea breeding lines across seasons and locations using phenomics data. *Frontiers in Plant Science*, *12*, 640259. https://doi.org/10.3389/fpls.2021.640259

Zhang, S., Li, X., Ba, Y., Lyu, X., Zhang, M., & Li, M. (2022). Banana fusarium wilt disease detection by supervised and unsupervised methods from UAV-based multispectral imagery. *Remote Sensing*, *14*(5), 1231. https://doi.org/10.3390/rs14051231

Zou, Z. B., Peng, H., & Luo, L. K. (2015). The application of random forest in finance. *Applied Mechanics and Materials*, *740*, 947–951. https://doi.org/10.4028/www.scientific.net/amm.740.947

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Navasca, H., Bazrafkan, A., Dariva, F. D., Kim, J.-H., Worral, H., Johnson, J. P., Acharya, S. R., Piche, L., Ross, A., Raymon, G., Zhang, Q., McGee, R. J., McPhee, K., Coyne, C. J., Flores, P., & Bandillo, N. (2025). Improving estimation of days to maturity in field pea using RGB aerial imagery and machine learning. *The Plant Phenome Journal*, *8*, e70038. https://doi.org/10.1002/ppj2.70038