

Assessing image based high throughput phenotyping for accurate characterization of soybean seed morphology

Shreyash Gijare

Agharkar Research Institute

Arathi S

Agharkar Research Institute

Preetkumar Trivedi

Agharkar Research Institute

Vaibhav Bhangare

Agharkar Research Institute

Aditya Gobade

Agharkar Research Institute

Jayshree Sarode

Agharkar Research Institute

Vijay Khade

Agharkar Research Institute

Manoj D Oak

Agharkar Research Institute

Deepak Pawar

Agharkar Research Institute

Abhinandan S. Patil

`apatil@aripune.org`

Agharkar Research Institute

Research Article

Keywords: Soybean, Seed morphometrics, High-throughput phenotyping, PLIMAN, Image analysis, Bland–Altman analysis, Vernier caliper

Posted Date: September 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10685107/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Assessing image based high throughput phenotyping for accurate characterization of soybean seed morphology

Shreyash Gijare^{1†}, Arathi S^{1†}., Preetkumar Trivedi¹, Vaibhav Bhangare¹, Aditya Gobade¹, Jayshree Sarode¹, Vijay Khade¹, Manoj. D. Oak¹, Deepak Pawar¹ and Abhinandan S. Patil^{1*}

[†]Contributed equally

¹Genetics and Plant Breeding Group, Agharkar Research Institute, Pune, Maharashtra 411004, India

* To whom correspondence should be addressed: Abhinandan S. Patil

Tel: +91–8983546613; E-mail: apatil@aripune.org

Abstract

Accurate assessment of seed morphometric traits is essential for soybean germplasm characterization and breeding. Image-based phenotyping offers a high-throughput alternative to conventional manual measurements; however, its accuracy and calibration requirements remain insufficiently evaluated. In this study, seed morphometric variation was assessed in 192 soybean accessions grown in two contrasting environments using the PLIMAN-based method and compared with conventional Vernier calliper (VC) measurements. Twelve image-derived traits describing seed size and shape were quantified and evaluated using descriptive statistics, genetic parameters, correlation analysis, principal component analysis (PCA), clustering, regression, and Bland–Altman analysis. Image-based phenotyping revealed substantial variation among the accessions and revealed significant environmental effects on seed morphology. Several traits, including seed area, radius, and diameter, exhibited high broad-sense heritability (>90%), indicating strong genetic control. PCA revealed that the first two principal components explained 80.5% of the total variation, with seed size-related traits contributing predominantly to PC1 and shape-related traits to PC2. PLIMAN classified a

greater proportion of accessions into distinct seed-size categories than VC, indicating greater sensitivity in detecting phenotypic variation. However, direct comparisons between methods showed weak linear relationships ($R^2 = 0.003\text{--}0.069$) and substantial systematic bias. Bland–Altman analysis revealed mean differences of 107.10 mm² for the seed area and 113.37 mm for the seed perimeter, indicating that PLIMAN consistently produced larger estimates than VC. These differences reflect the distinct measurement principles of the two approaches rather than methodological error. Overall, PLIMAN provides a rapid and reproducible platform for large-scale soybean seed phenotyping. However, standardized image acquisition and calibration protocols are needed before image-derived measurements can be directly compared with conventional methods.

Keywords: Soybean, Seed morphometrics, High-throughput phenotyping, PLIMAN, Image analysis, Bland–Altman analysis, Vernier caliper

1. Introduction

Soybean (*Glycine max* [L.] Merr.) is among the most important legume crops worldwide because of its high protein and oil contents. The morphometric characteristics of seeds, including length, width, thickness, area, and shape, are important determinants of seed quality and are widely used in studies of varietal identification, breeding, and crop improvement (Ogunniyan et al. 2021; Çetin 2022). These traits are associated with agronomic performance, seedling establishment, storage behavior, market preference, and processing efficiency.

Seed size and shape are major domestication-related traits that influence germination, moisture absorption, and the quality of soy-derived products. Large-seeded soybean genotypes are generally preferred for Edamame, tofu, soy milk, and other value-added products because of their favorable hydration and processing properties (Ray et al. 2022). Seed size is a complex quantitative trait controlled by multiple genetic factors that regulate endosperm and embryo development, with variation often linked to the timing of endosperm cellularization (Xu and

Zhang 2023). Therefore, classifying soybean accessions into distinct groups on the basis of seed size provides a useful framework for assessing genetic diversity, environmental responses, and end-use suitability.

Traditionally, seed morphometric traits have been measured with Vernier calipers because of their simplicity and accessibility. However, manual measurements are labour-intensive, time-consuming, and generally limited to basic linear dimensions. In addition, measurement accuracy may be affected by operator variation and observer bias, thereby reducing its suitability for high-throughput phenotyping and large-scale evaluations across environments (Parihar and Bhingarde 2024).

Recent developments in digital phenotyping have enabled automated, high-resolution, and nondestructive assessments of plant traits through image-based approaches (Sheikh et al. 2024). Image analysis platforms allow the extraction of multiple morphometric parameters while increasing measurement throughput and reproducibility. Reliable image-based phenotyping depends on standardized acquisition conditions, including uniform backgrounds, controlled illumination, appropriate scale calibration, and consistent object positioning (Tanabata et al. 2012). Several tools have been developed for seed and plant image analysis, including ImageJ (Abramoff et al. 2004), SmartGrain (Tanabata et al. 2012), SeedCounter (Komyshev et al. 2017), MuSeed (Gao et al. 2017), PlantCV v2 (Gehan et al. 2017), and custom workflows implemented in Python and MATLAB (Baek et al. 2020). These platforms support rapid and reproducible extraction of phenotypic traits and offer opportunities for integration with advanced analytical and machine-learning approaches.

Among these tools, the R package PLIMAN (plant image analysis) has emerged as a flexible, open-source platform for extracting morphometric information using image segmentation and object recognition (Olivoto 2022). PLIMAN enables the quantification of both conventional traits, such as length, width, and area, and advanced descriptors, such as

circularity, convex hull area, and aspect ratio. These capabilities make the PLIMAN platform suitable for applications in high-throughput phenotyping, genomic selection, genome-wide association studies (GWASs), and quantitative trait locus (QTL) analysis.

Image-based phenotyping approaches have demonstrated utility across several crops, including SeedExtractor-based seed analysis in rice (Zhu et al. 2021), hyperspectral imaging in maize (Wongchaisuwat et al. 2025), and high-resolution RGB imaging in wheat (Nawaz et al. 2025). Similar advances have also been reported for soybean seed phenotyping (Yang et al. 2021; Ghimire et al. 2023; Miranda et al. 2023). While image-based platforms are being increasingly used for seed phenotyping, relatively few studies have systematically evaluated their ability to characterize soybean seed morphometric variation across contrasting environments by directly comparing image-driven measurements with conventional Vernier calipers (VC) measurements.

Therefore, in the present study, the performance of the PLIMAN package was evaluated against that of conventional VC measurements for quantifying seed morphometric traits across 192 soybean accessions grown in two environments. A total of 12 morphometric traits related to seed size and shape were assessed to compare image-based and manual measurements. We hypothesized that digital image analysis would provide a reproducible and scalable alternative for large-scale phenotyping while maintaining consistency across genotypes and environments. Specifically, this study aimed to (1) compare soybean seed morphometric measurements obtained using the PLIMAN and VC methods and (2) identify and characterize image-processing error sources and potential calibration requirements associated with PLIMAN-based phenotyping. These findings provide insights into the applicability and limitations of image-based phenotyping for soybean seed morphometric evaluation.

2. Materials and Methods

2.1 Plant materials and growing conditions

A total of 192 diverse soybean accessions from our core collection were used in this study (Gobade et al. 2024; Patil et al. 2025). The experiment was conducted in a randomized block design with three replicates across two locations: October 2022 at the Agharkar Research Institute, Pune (designated E1), and February 2023 at Hol Farm (designated E2). Both locations were managed under irrigated conditions, utilizing a standard agronomic package for fertilization, irrigation scheduling, and disease and pest management as described in Gobade et al. 2024. The list of genotypes evaluated is provided in Supplementary Table S1.

Seeds were harvested at full maturity (R8 stage) as per the standard descriptor of soybean growth stages (Fehr and Caviness 1977). After harvest, the seed packets were stored at 37°C to maintain uniform seed moisture. Uniformly sized seeds that were free from physical damage and visible deformities were selected from each genotype for further seed morphometric analysis.

2.2 Seed morphometric phenotyping

The morphometric traits of the seeds of 192 soybean accessions evaluated in the two environments (E1 and E2) were analyzed using the PLIMAN package in R (Olivoto 2022) and compared with measurements obtained using the conventional VC method to determine the most accurate and efficient approach.

In image-based phenotyping, for each genotype in each environment, seeds from the three replications were composited, and 20 seeds were randomly selected. These seeds are then arranged individually on a clean, white background without overlapping. Images were captured using the primary rear camera (50 MP) of a Motorola G64 5G smartphone (Android 15) at 1× magnification from a fixed distance of 50 cm under natural daylight. All the images were acquired under identical imaging conditions and saved in JPEG (.jpg) format at a resolution of

960 × 1280 pixels. Autofocus was enabled to ensure sharpness, while the high dynamic range (HDR) function was disabled to maintain uniform contrast. A metric steel ruler was included in each image as a reference scale for image calibration.

To determine the pixel-to-metric conversion factor, 10 images were randomly selected from the dataset. In each image, the 10 cm (100 mm) reference distance on the ruler was measured using ImageJ (National Institutes of Health, Bethesda, MD, USA). The mean pixel length corresponding to the 100 mm reference distance was 378.2 ± 2.2 pixels ($n = 10$). On the basis of this value, the image conversion factor was calculated to be $0.2644 \text{ mm pixels}^{-1}$ (1 pixel = 0.2644 mm), which was subsequently used to convert all pixel-based measurements to metric units.

Image analysis was performed using the PLIMAN package (Olivoto 2022), which enables (i) extraction of seed morphometric data from digital images, (ii) visualization and manual editing of segmented objects, (iii) statistical analyses, and (iv) compilation of results in tabular format (Claude 2008). Individual seeds were automatically segmented from the background using the package's image segmentation and object recognition functions.

The following 12 seed morphometric traits were extracted from images obtained in both environments: area (mm^2), perimeter (mm), radius (mm), minimum radius (mm), maximum radius (mm), radius ratio, diameter (mm), minimum diameter (mm), maximum diameter (mm), eccentricity, solidity, and circularity. Linear measurements were converted from pixels to millimeters using the conversion factor ($0.2646 \text{ mm pixel}^{-1}$). In contrast, area measurements were converted to square millimeters by multiplying the pixel area by the square of the calibration factor.

In contrast, using the VC method, the length and width of seeds were measured via a VC, specifically Yuzuki, Model VC6 (resolution 0.02 mm/0.001", Yuzuki Precision Instruments, Japan). Each seed was measured only once, and the recorded dimensions were

used to estimate the seed area and perimeter. Seed length was measured along the longest axis, whereas width was measured at the widest point perpendicular to the longitudinal axis. All measurements were performed by the same operator to minimize observer variation. The seed area ($A \text{ mm}^2$) was calculated from the length and width by approximating the seed shape as an ellipse via the following formula: $\text{Area} = \frac{\pi \times L \times W}{4}$

Similarly, the perimeter (mm) of the supposedly elliptical seed is calculated via the following formula: $\text{perimeter} = \pi \times \sqrt{\frac{L^2 + W^2}{2}}$

In both of these formulas, L is the length (in mm), W is the width (in mm), and π is 3.14.

2.3 Seed classification

Seed size classification was performed separately for PLIMAN-based and VC measurements. On the basis of the PLIMAN-derived seed area, the accessions were classified as large ($\geq 100 \text{ mm}^2$), medium ($50\text{--}100 \text{ mm}^2$) or small ($\leq 50 \text{ mm}^2$). VC-based measurements: accessions classified into large ($\geq 50 \text{ mm}^2$), medium ($20\text{--}50 \text{ mm}^2$) and small ($\leq 20 \text{ mm}^2$) categories. The thresholds for each method were determined on the basis of the range of areas obtained by each method. This was necessary because two methods estimate seed area using different measurement principles: PLIMAN calculates the seed area directly from the segmented seed silhouette, whereas the VC method calculates it from manually measured seed length and width, assuming an elliptical seed shape. Therefore, method-specific comparisons of seed size classes were performed by applying different classification limits to each dataset to account for the methodological differences in area estimation.

2.4 Statistical analysis

The mean values of the 10 seeds measured for each genotype in each environment were used for statistical analyses. Descriptive statistics, including the mean, standard error (SE), range,

coefficient of variation (CV), skewness, and kurtosis, were calculated for all the seed morphometric traits. Genetic parameters, including the phenotypic coefficient of variation (PCV), genotypic coefficient of variation (GCV), broad-sense heritability (H^2), and genetic advance as a percentage of the mean (GA%), were estimated following standard quantitative genetic procedures.

Analysis of variance (ANOVA) was performed using a randomized block design (RBD) to evaluate the effects of genotype and environment on the measured traits. Before ANOVA, the assumptions of normality and homogeneity of variance were examined using residual diagnostics. Differences were considered statistically significant at $p < 0.05$.

Linear regression analysis was conducted using the linear model (lm) function in RStudio (R version 4.3.1; R Core Team, 2023) to compare measurements obtained from the PLIMAN method and the VC method. Pearson's correlation coefficients among morphometric traits were calculated and visualized using the corrplot package in R.

Multivariate analyses, including principal component analysis (PCA), hierarchical clustering, constellation analysis, and K-means clustering, were performed using JMP Pro 17 (SAS Institute Inc., Cary, NC, USA). Boxplots illustrating the distribution of seed morphometric traits were generated using BoxPlotR (Spitzer et al. 2014), and Venn diagrams comparing seed size classifications obtained by the PLIMAN and VC methods across environments were generated using EVenn (Yang et al. 2024).

2.5 Bland–Altman analysis

Agreement between the PLIMAN-based measurements and VC measurements for the seed area and perimeter was evaluated using a Bland–Altman analysis. The analysis was performed in RStudio (R version 4.3.1) using the Bland–Altman–Leh package together with base R functions. For each genotype, the mean of the two measurement methods and the corresponding difference between the methods were calculated. The mean difference (bias) and the 95% limits

of agreement (LoA) were estimated as the mean difference $\pm$ 1.96 times the standard deviation of the differences (Bland and Altman, 1986). Bland–Altman plots were constructed by plotting the differences against the corresponding mean values, with horizontal reference lines representing the mean bias and the upper and lower 95% limits of agreement. These analyses were used to assess systematic bias and the level of agreement between the image-based and manual measurement methods.

3. Results

3.1 Phenotypic variability and baseline descriptive statistics

Seed morphometric variability in 192 soybean accessions was evaluated using the ‘PLIMAN’ package and VC across two environments, E1 and E2. The means, descriptive statistics, frequency distribution parameters, and genetic estimates for all evaluated traits are presented in Table 1 and Supplementary Tables S2 and S3.

The PLIMAN-based analysis quantified 12 morphometric parameters related to seed size and shape. Once the pixel-derived measurements were converted to millimeters, significant environmental differences were detected between E1 and E2. For most image-derived traits, the mean values were greater in E2 than in E1. The seed area increased from $125.34 \pm 3.26 \text{ mm}^2$ in E1 to $151.82 \pm 3.75 \text{ mm}^2$ in E2, whereas the seed perimeter increased from $163.49 \pm 2.72 \text{ mm}$ in E1 to $189.18 \pm 3.18 \text{ mm}$ in E2. Similar trends were observed for the radius, minimum radius, maximum radius, diameter, minimum diameter, maximum diameter and circularity. In contrast, eccentricity (0.52 ± 0.01 in E1; 0.56 ± 0.01 in E2) and solidity (0.96 ± 0.01 in E1; 0.95 ± 0.01 in E2) exhibited minimal variation across environments. VC measurements also revealed identical environmental variations between E1 and E2, supporting the PLIMAN-based phenotyping observations. The coefficient of variation (CV) indicated substantial phenotypic variability among the soybean accessions. In PLIMAN analysis, the CV showed a distinct trend; except for the radius ratio, all shape-related traits exhibited lower CVs

in E2 than in E1: eccentricity (20.42 in E1; 17.09 in E2), solidity (12.02 in E1; 10.26 in E2), and circularity (99.94 in E1; 44.64 in E2). For the VC-derived traits, the coefficient of variation (CV) was slightly greater in E2 than in E1. For Area and Perimeter, the CV values were 17.11% and 19.27%, respectively, in E1 and E2.

With respect to the frequency distribution, the PLIMAN-derived size-related traits showed skewness values between -2.2 and -0.2 , indicating moderately to strongly left-skewed distributions. Area and perimeter traits showed weak to moderate negative skewness (area: -0.44 in E1 and -0.57 in E2; perimeter: -0.24 in E1 and -0.60 in E2). Additionally, the radius ratio, eccentricity, and circularity showed positive skewness in both environments. Kurtosis values were positive for all evaluated traits, indicating leptokurtic distributions. The highest kurtosis values were observed for the radius ratio (28.46 in E1 and 36.23 in E2), followed by solidity (20.32 in E1 and 20.15 in E2) and circularity (15.36 in E1 and 39.07 in E2). In contrast, the VC-derived area and perimeter measurements exhibited positive skewness (>0.5) and positive kurtosis (>0), indicating that they are positively skewed and leptokurtic.

The phenotypic coefficient of variation (PCV) was greater than the genotypic coefficient of variation (GCV) for all traits, indicating that environmental factors influence trait expression. In the image-derived dataset, both the PCV and GCV values were generally greater in E1, except for the radius ratio, which exhibited substantially greater variability in E2 (PCV: 174.04%). Broad-sense heritability (H^2) estimates were high for several traits in E2, particularly for seed area (97.04%), mean radius (94.28%), maximum radius (89.37%), mean diameter (94.27%), and maximum diameter (89.37%), suggesting strong genetic control. Similarly, the minimum radius (91.29%) and the minimum diameter (91.21%) indicated strong genetic control, with high H^2 in E1.

Genetic advancement as a percentage of the mean (GA%) was greater in E2 (64.61%) than in E1 (45.38%) for seed area alone; for the other traits, GA was lower in E2 than in E1. Among all the evaluated traits, circularity had the highest estimates for PCV (118.76), GCV (107.18), H^2 (81.46), and GA (199.28) in E1.

Similar trends were observed for the VC-derived measurements, with H^2 and GA estimates higher in E2 for both area ($H^2 = 75.35\%$; GA = 34.18%) and perimeter ($H^2 = 79.98\%$; GA = 16.91%). Furthermore, all measured traits differed significantly ($P < 0.05$) between the phenotyping platforms and environments.

The 100-seed weight was also evaluated for all 192 soybean accessions (Supplementary Table S6). Among these, 27 genotypes had identical 100-seed weight values across E1 and E2, whereas 123 genotypes had higher 100-seed weight in E1 than in E2. These findings suggest that although the projected seed size was greater under E2, the dry matter accumulation was relatively greater in the controlled E1 environment.

3.2 Extreme genotypic cohorts and selection for seed morphometrics

To assess phenotypic variability and evaluate the consistency of high-throughput image-based phenotyping, soybean genotypes exhibiting extreme seed size classes were identified using the PLIMAN-based and VC measurements across two environments (E1 and E2). The genotypes were classified into large- and small-seeded groups on the basis of mean seed area and perimeter, as shown in Table 2 and Fig. 2.

PLIMAN-based analysis revealed substantial variation among the evaluated soybean accessions. Based on the mean seed area, the five largest-seeded genotypes were identified as MACS 288 (247.45 mm²), TS 17 (247.35 mm²), JS (SH) 99-14 (235.90 mm²), MACS 112 (227.35 mm²), and EC 38828 (219.68 mm²). In contrast, the smallest-seeded genotypes identified through image analysis were Green Tur (12.20 mm²), EC 34079A (11.40 mm²), JS (SH) 9337 (11.20 mm²), EC 39795 (10.70 mm²), and EC 39776 (10.50 mm²).

Similarly, seed perimeter-based rankings from the PLIMAN-derived dataset revealed PK 1029 (271.43 mm), MACS 288 (267.00 mm), MACS 37 (257.67 mm), MACS 173 (255.20 mm), and JS (SH) 99-14 (249.39 mm) as the genotypes with the greatest seed perimeter values. In contrast, the lowest perimeter values were recorded for EC 14437 (97.71 mm), EC 39795 (69.97 mm), Kalitur (67.74 mm), EC 39776 (64.57 mm), and MACS 261 (23.09 mm).

The VC-based measurements also clearly distinguished between large- and small-seeded cohorts. Based on the mean seed area, the largest genotypes identified were NRC 150 (53.21 mm²), EC 38828 (44.75 mm²), PS 1370 (43.79 mm²), JS 9305 (43.63 mm²), and EC 14477 (41.36 mm²). Conversely, the smallest-seeded genotypes were IC 13048 (23.17 mm²), IC 13053 (22.50 mm²), VLS 67 (21.63 mm²), MACS 94 (21.53 mm²), and EC 39776 (21.07 mm²).

The perimeter measurements obtained through VC similarly revealed NRC 150 (106.42 mm), EC 38828 (89.51 mm), PS 1370 (87.58 mm), JS 9305 (87.26 mm), and EC 14477 (82.71 mm) as the genotypes with the highest perimeter values. In contrast, IC 13048 (46.33 mm), IC 13053 (45.01 mm), VLS 67 (43.27 mm), MACS 94 (43.07 mm), and EC 39776 (42.15 mm) exhibited the smallest perimeter values.

Notably, EC 38828 was consistently classified as a large-seeded genotype, whereas EC 39776 was consistently classified as a small-seeded genotype across both phenotyping methods. The phenotypic variation in seed size and coat characteristics among representative large- and small-seeded genotypes identified using both approaches is shown in Fig. 2.

3.3 Trait interrelationships and correlation architecture

Linear regression analysis was performed to compare the seed area and perimeter measurements obtained from the PLIMAN-based image analysis and VC methods across both environments (Fig. 3). The linear relationship between the seed area measurements obtained by the two methods was weak, with R² values of 0.069 in E1 and 0.037 in E2 (Fig. 3A, B).

Similarly, perimeter measurements exhibited weak linear associations, with R^2 values of 0.054 for E1 and 0.003 for E2 (Fig. 3C, D).

Pearson's correlation analysis was conducted to examine the relationships among seed morphometric traits (Fig. 4; Supplementary Table S4). The seed area was strongly positively correlated with most size-related traits, particularly the maximum radius ($r = 0.87$), minimum radius ($r = 0.86$), maximum diameter ($r = 0.85$), radius ($r = 0.82$), and diameter ($r = 0.82$). Similarly, the maximum radius was strongly correlated with the maximum diameter ($r = 0.96$), the diameter ($r = 0.91$), the minimum radius ($r = 0.90$), and the minimum diameter ($r = 0.82$). The seed perimeter also exhibited positive correlations with the maximum radius ($r = 0.69$), maximum diameter ($r = 0.69$), area ($r = 0.62$), radius ($r = 0.60$), diameter ($r = 0.60$), minimum radius ($r = 0.57$), and minimum diameter ($r = 0.50$). In addition, perimeter was moderately positively associated with seed solidity ($r = 0.43$).

Among the shape-related traits, circularity was positively correlated with the radius ratio ($r = 0.53$) and eccentricity ($r = 0.50$). In contrast, eccentricity was significantly negatively correlated with several size-related traits, including minimum radius ($r = -0.76$), minimum diameter ($r = -0.69$), solidity ($r = -0.68$), radius ($r = -0.65$), diameter ($r = -0.65$), and area ($r = -0.63$). The radius ratio was also negatively associated with the minimum radius ($r = -0.60$), area ($r = -0.54$), radius ($r = -0.50$) and maximum radius ($r = -0.46$). Circularity showed weak-to-moderate negative correlations with solidity ($r = -0.67$), minimum diameter ($r = -0.55$), minimum radius ($r = -0.54$), radius ($r = -0.44$), diameter ($r = -0.44$), maximum radius ($r = -0.42$) and area ($r = -0.40$).

Overall, the correlation analysis revealed strong positive correlations among seed size-related traits, including area, perimeter, radius and diameter. In contrast, shape-related parameters, such as eccentricity, circularity, and the radius ratio, exhibited distinct correlation patterns, predominantly negative with size-related traits. Most trait associations were

statistically significant ($p < 0.05$), indicating consistent relationships among the evaluated traits within the soybean germplasm panel.

3.4 Seed size classification and comparison of morphometric measurements obtained using PLIMAN and VC

To compare seed morphometric measurements obtained via PLIMAN-based and VC manual approaches, box plots were constructed for the seed area and perimeter using data from the PLIMAN and VC measurements (Fig. 5A). The PLIMAN-derived (R) traits are represented as R_Area and R_Perimeter. In contrast, the manual measurements are represented as VC_Area and VC_Perimeter.

Box plots consistently showed higher means for both area and perimeter in PLIMAN-based measurements than in VC measurements. The median, interquartile range, and whisker limits were consistently greater for PLIMAN-derived traits, indicating that image-based phenotyping yielded larger estimates of seed dimensions than conventional manual measurements. Additionally, the perimeter values exceeded the area values for both methods. As detailed in Table 1, the mean seed area measured using R was $125.34 \pm 3.26 \text{ mm}^2$ in E1 and $151.82 \pm 3.75 \text{ mm}^2$ in E2, whereas the corresponding VC values were substantially lower at $30.97 \pm 0.38 \text{ mm}^2$ and $32.00 \pm 0.44 \text{ mm}^2$, respectively. Similarly, the PLIMAN-based seed perimeter values were greater than those obtained using VC, with mean values of $163.49 \pm 2.72 \text{ mm}$ and $189.18 \pm 3.18 \text{ mm}$ in E1 and E2, respectively, compared with $61.95 \pm 0.76 \text{ mm}$ and $63.99 \pm 0.89 \text{ mm}$ for VC. The broader range and higher values in the PLIMAN-based measurements suggest that the method captures the full extent of seed variability, highlighting its comprehensiveness.

A Venn diagram (Fig. 5B) further illustrates the overlap of seed classifications between environments and methods. This analysis compared four groups based on area measurements: E1-R, E2-R, E1-VC, and E2-VC. The greatest overlap occurred within the same phenotyping

platforms: 154 genotypes were shared between E1-R and E2-R, and 184 were shared between E1-VC and E2-VC. Comparisons across same environment revealed lower overlap: 37 common genotypes between E1-R and E1-VC and eight between E2-R and E2-VC.

With respect to the three-way intersections, five genotypes were shared among E1-R, E2-R, and E1-VC, whereas six genotypes were common among E1-R, E2-R, and E2-VC. Similarly, 37 genotypes overlapped among E1-R, E1-VC, and E2-VC, and six genotypes overlapped among E2-R, E1-VC, and E2-VC. Only two genotypes were consistently identified across all four groups, indicating stable seed size classification irrespective of the environment or measurement platform (Supplementary Table S7).

Based on seed area measurements, the soybean accessions were classified into three seed size categories (Table 3). For the PLIMAN-based method, the seeds were grouped into large (≥ 100 mm²), medium (50–100 mm²), and small (≤ 50 mm²) categories. Under E1, 143 genotypes were classified as small-seeded, 36 as medium-seeded, and 13 as large-seeded. Under E2, the number of small-seeded genotypes increased to 174, whereas only five and 13 genotypes were classified as medium- and large-seeded, respectively.

For the VC measurements, the classification thresholds were adjusted to large (≥ 50 mm²), medium (20–50 mm²), and small (≤ 20 mm²). In E1, most accessions (186 genotypes) were classified as medium-seeded, while only five and one genotype belonged to the small- and large-seeded classes, respectively. Under E2, 188 genotypes were categorized as medium-seeded, whereas two genotypes each were classified as small- and large-seeded. Overall, compared with manual measurements, the PLIMAN-based approach resulted in greater variability in seed size.

3.5 Multivariate analysis of seed morphometric traits: PCA and clustering

Principal component analysis (PCA), hierarchical clustering, constellation analysis, and K-means clustering were performed to assess and group the phenotypic variation among 192 soybean accessions using 12 seed morphometric traits (Fig. 6).

PCA indicated that the first two principal components explained 80.5% of the total variation, with PC1 accounting for 69.6% and PC2 accounting for 10.9%. Most accessions clustered near the origin of the PCA plot, indicating limited overall divergence in seed morphology. In contrast, a subset of accessions was separated along the principal component axes (Fig 6A and Supplementary Table S5).

Trait loadings indicated that perimeter, maximum diameter, area, minimum radius, and solidity contributed most strongly to PC1, suggesting that this axis primarily reflected variation in seed size. In contrast, circularity, the radius ratio, and eccentricity contributed more to PC2, suggesting that this axis captured the variation in seed shape.

The clustering analyses were consistent with the PCA results. Hierarchical clustering grouped the accessions by morphometric similarity (Fig. 6B), whereas constellation analysis illustrated relationships among clusters and identified groups with similar seed characteristics (Fig. 6C). K-means clustering further classified the accessions into three major groups corresponding to small-, intermediate-, and large-seeded classes (Fig. 6D).

Overall, the multivariate analyses revealed seed area and perimeter as the major contributors to seed morphometric variation and as useful descriptors for distinguishing soybean accessions by seed size and shape.

3.6 Error Profiling and Calibration Nuances in Image Extraction

Bland–Altman analysis of the soybean seed perimeter measurements between the VC and PLIMAN-based methods (Fig. 7A) revealed a mean difference (bias) of 113.37 mm, with 95% limits of agreement (LoA) ranging from 42.81 to 183.92 mm, indicating that the seed perimeter

of PLIMAN was approximately 113.37 mm larger than that of VC on average and that for approximately 95% of the observations, the difference between the two methods ranged from 42.81 to 183.92 mm. Similarly, for the seed area measurements (Fig. 7B), the mean difference was 107.10 mm², and the 95% LoA ranged from 24.34 to 189.85 mm². This means that compared with the VC, PLIMAN measured a seed area approximately 107.10 mm² larger on average, and for approximately 95% of the observations, the difference between the two methods ranged from 24.34 to 189.85 mm².

4. Discussion

In this study, the applicability of the PLIMAN platform for the characterization of seed morphometric variation among diverse soybean germplasms grown in different environments was evaluated, and its performance was compared with that of the conventional method. Overall, the results demonstrated that PLIMAN can efficiently extract multiple seed morphometric traits and allow comprehensive characterization of variation among 192 soybean accessions. The growing environment influenced seed morphology, with most size-related traits varying between the E1 and E2 conditions. Compared with the VC method, PLIMAN captured a broader spectrum of morphometric variation and provided more detailed quantitative information on seed size and shape. However, systematic differences between the two methods were observed, as they are based on fundamentally different measurement principles. Collectively, these results emphasise the usefulness of image-based phenotyping for large-scale soybean germplasm characterisation while highlighting the importance of understanding methodological differences when comparing phenotyping platforms.

4.1 Environmental Effects on Seed Traits

A morphometric analysis of seeds from 192 soybean accessions across two contrasting environments, E1 and E2, revealed significant environmental influences on seed traits. For most parameters, including area, perimeter, and diameter, the mean values of the seeds in E2

were greater than those in E1. These differences can be attributed to environmental factors such as light intensity, temperature, soil nutrient availability and water supply, all of which influence seed development and final seed size (Zhou et al. 2022; Zhang et al. 2024).

Eccentricity and solidity were stable between the two environments, indicating that seed elongation and overall outline compactness were not strongly affected by environmental conditions. In contrast, the circularity was greater in E2 than in E1, indicating that the seeds produced in E2 were closer to an ideal circular shape. Taken together, these results suggest that while overall seed geometry is preserved across environments, environmental conditions affect subtle aspects of seed shape, as indicated by circularity. Similar observations have been reported in soybean, where genetic factors and environmental conditions control seed size and shape during seed development (Sun and Yuan 2022; Zhou et al. 2022).

An interesting observation of the experiment was that greater seed dimensions were estimated in E2, contradicting the fact that a large number of genotypes in E1 had 100-seed-weight seeds as compared to E2. A key finding of this study was the differing environmental responses of seed dimensions and weight. These findings indicate that environmental conditions may influence seed expansion and dry matter accumulation differently during seed development, leading to a difference between external seed morphology and final seed mass (Chiluwal et al. 2022; Poudel et al. 2024; Egli and Bruening 2001).

While PCA and hierarchical clustering highlighted distinct environment-specific clustering, with E2 associated with greater seed morphometric traits, correlation analysis confirmed greater trait consistency in E2 than in E1, likely reflecting more favorable growth conditions. These results emphasize the importance of genotype $\times$ environment interactions in determining seed morphology and highlight the need to evaluate germplasms across multiple environments to develop breeding strategies with broad adaptability and stable performance (Zhang et al. 2024).

4.2 Trait variability and potential for genetic improvement

ANOVA and genetic parameters further reinforced the potential for genetic improvement. High phenotypic and genotypic coefficients of variation (PCV and GCV) were observed for several traits, particularly the seed area, radius ratio, and circularity. H^2 values were recorded for traits such as area (97.04% in E2), mean radius (94.28% in E2), and diameter (94.27% in E2), indicating that these traits are predominantly under genetic control and are therefore amenable to selection. Similar relationships between high heritability, high genetic advance, and improved selection efficiency have been reported for soybean seed-related traits (Ray et al. 2022).

The CV analysis indicated substantial variability among the evaluated seed morphometric traits. The circularity and radius ratio displayed the highest CVs (circularity: 99.94% for E1 and 44.64% for E2; radius ratio: 61.51% for E1 and 73.08% for E2), suggesting greater phenotypic variability due to sensitivity to environmental factors. Together, high heritability and genetic gain of area, perimeter, radius and diameter make them promising targets for breeding programs aimed at improving seed size and shape. Similarly, the relatively high heritability of circularity and solidity makes them suitable candidates for improving seed uniformity, compactness and overall seed quality (Ray et al. 2022; Zhang et al. 2024).

The distribution patterns differed between the two measurement methods. The distributions of area and perimeter measurements from the PLIMAN-based were slightly left skewed, particularly in E2, suggesting that most observations were clustered toward relatively large values, with a longer tail toward smaller values. On the other hand, the measurements obtained using the VC showed positively skewed, leptokurtic distributions, indicating greater peakedness. These contrasting patterns suggest that the image-based and VC techniques captured seed morphometric variation differently. The majority of the tested soybean germplasm size-related traits were negatively skewed, suggesting a predominance of larger-

seeded accessions and a small number of genotypes with very small seeds. On the other hand, the positive skewness of circularity, eccentricity, and the radius ratio indicates that variation in seed shape was concentrated around intermediate values, with only a few accessions exhibiting extreme seed characteristics. The positive kurtosis for the evaluated traits also indicated that most accessions were concentrated around the population mean, with a small number showing extreme values, despite significant phenotypic variation. These extreme phenotypes may serve as valuable genetic resources for future studies of the genetic architecture of soybean seed morphology. Such distributions are typical of quantitative traits controlled by many genes and affected by environmental variation (Falconer and Mackay 1996).

Bland–Altman analysis revealed a consistent positive bias in both the seed perimeter and area measurements, with the PLIMAN image-based method systematically yielding larger estimates than the VC method. This systematic difference is likely due to the use of automated edge detection and pixel-based measurements, which more comprehensively capture the entire projected seed boundaries than manual caliper measurements do. The increasing dispersion of differences as the mean perimeter and area values increased indicated proportional bias, with disagreement between the two methods becoming more pronounced for larger seeds. The wide limits of agreement further reinforce poor agreement between the methods, indicating that PLIMAN-derived measurements require appropriate calibration before they can be used interchangeably with VC measurements. Bland–Altman analysis is widely recommended for evaluating the agreement between alternative measurement methods, and similar systematic differences between digital image-based and manual measurements have been reported in plant phenotyping studies (Bland and Altman 1986; Li et al. 2014; Yang et al. 2020).

4.3 Seed categorization and practical implications

The relationship between PLIMAN-based imaging and VC measurements was analyzed using regression and correlation analyses. Traits such as area and diameter presented the highest

correlation values, confirming their robustness across different measurement techniques. Similar advantages of image-based phenotyping over conventional manual measurements have been reported for seed trait characterization and high-throughput plant phenotyping (Hacisalihoglu and Armstrong 2023; Yang et al. 2020).

Seed size plays a crucial role in determining crop establishment, seedling vigor, and ultimately crop productivity (Arshad et al. 2025; Moles et al. 2005). The categorization of seeds on the basis of area measurements obtained using both PLIMAN and VC revealed notable differences in seed size classification between the two environments. Compared with VC and PLIMAN-based measurements, it classified substantially more genotypes as small-seeded, highlighting the sensitivity of digital image analysis for detecting subtle variations in seed morphology. This discrepancy underscores the value of image-based phenotyping for accurate classification and rapid evaluation of large germplasm collections (Hacisalihoglu and Armstrong 2023).

A Venn diagram was used to identify the shared and unique morphometric characteristics among the three seed size categories. The overlapping regions highlighted common traits, such as area and perimeter, whereas the unique regions emphasized traits, such as circularity and eccentricity, that differentiated specific seed-size groups. Such multivariate approaches facilitate the identification of complementary phenotypic traits that can support germplasm characterization and breeding decisions (Yang et al. 2020).

K-means clustering and constellation plots further illustrated the relationships among genotypes based on morphometric traits, highlighting accessions such as JS (SH) 9914 and EC 39776 that consistently exhibit large or small seed areas across environments. The stability of these genotypes across contrasting environments suggests that they may serve as valuable parental material for breeding programs aimed at improving seed size (Zhang et al. 2024).

Both the area and perimeter values obtained using the PLIMAN-based method were consistently greater than those from the VC method, suggesting that digital image analysis captured seed boundary irregularities more comprehensively than geometric estimates from manual measurements. Consequently, image-based phenotyping is better suited to detailed morphometric studies requiring high precision, whereas VC measurements remain useful for rapid, routine measurements when only basic linear dimensions are needed (Hacisalihoglu and Armstrong 2023).

4.4 Evaluation of the PLIMAN package for high-throughput seed morphometrics

PLIMAN has been shown to be an effective platform for high-throughput phenotyping of soybean seeds by automatically extracting multiple size and shape descriptors from a single digital image. PLIMAN quantified 12 morphometric traits simultaneously from images with multiple seeds evaluated in this study, significantly increasing the amount of phenotypic information collected while decreasing the analysis time and manual effort, unlike the VC, which is usually limited to a few linear measurements recorded manually. This high-throughput ability allows for rapid characterization of large germplasm collections that would be much more laborious using conventional measurement approaches (Olivoto et al. 2022; Yang et al. 2020).

The automated image-processing workflow improves the reproducibility of measurements by eliminating observer bias and yielding consistent morphometric data in images captured under standardized conditions. Additionally, digital image acquisition generates a permanent record of the original samples such that archived datasets can be reanalyzed when superior image-processing algorithms or additional morphometric descriptors become available. Reproducibility and data archiving of this kind are difficult to obtain with traditional measurements using a VC, which requires repeated physical measurements of individual seeds (Hacisalihoglu and Armstrong 2023; Olivoto et al. 2022).

Furthermore, since PLIMAN is implemented in the R environment, the extracted morphometric data can be used directly for downstream statistical and genetic analyses, such as descriptive statistics, correlation analysis, principal component analysis, clustering, heritability estimation, and genome-wide association studies (GWASs), without any intermediate conversion. Such features increase analytical efficiency, making PLIMAN especially appropriate for large-scale germplasm characterization, genetic studies and breeding programs that require fast, reproducible and comprehensive seed phenotyping (Olivoto et al. 2022; Yang et al. 2020).

4.5 Methodological differences between PLIMAN and VC

PLIMAN and VC are both used to quantify soybean seed morphometric traits; however, they are based on fundamentally different measurement principles. PLIMAN performs measurements directly from the projected two-dimensional seed silhouette obtained from digital image segmentation, meaning that the entire seed boundary contributes to the calculations of projected area, perimeter, circularity, eccentricity, and solidity (Olivoto et al. 2022). In contrast, the VC method relies on manual measurements of seed length and width, from which the area and perimeter are estimated by geometric equations that approximate the seed as an ellipse. The area and perimeter values based on the geometric equations of the approximated ellipse from the measurement of the VC may not fully represent the actual projected seed outline, particularly for accessions with irregular morphology (Hacisalihoglu and Armstrong 2023), since the shape of soybean seeds is rarely perfectly elliptical.

These methodological differences might explain the weak linear associations between PLIMAN and VC-derived measurements, as well as the systematic bias observed in the Bland–Altman analysis. The differences observed are not due to poor performance of either method but rather to the two approaches quantifying seed morphology differently. The VC method indirectly estimates morphometric traits from two manually measured dimensions. In contrast,

PLIMAN captures the entire projected seed boundary, accounting for subtle edge irregularities and natural deviations from ideal geometric shapes. Therefore, the measurements obtained by the two methods should not be taken as directly interchangeable, especially when quantitative comparisons are needed (Bland and Altman 1986; Olivoto et al. 2022).

Bland–Altman analysis also revealed proportional bias, indicating that the difference between PLIMAN and VC increased with seed size, highlighting the importance of calibration when comparing image-derived measurements with conventionally measured values. Ideally, calibration should account for differences in imaging conditions, camera setups, segmentation parameters, and seed morphology, rather than relying on a single universal correction factor. Thus, the development of standardized calibration protocols would enhance the consistency, reproducibility and comparability of image-based phenotyping across experiments and breeding programs (Yang et al. 2020; Hacisalihoglu and Armstrong 2023, Okyere et al. 2023).

4.6 Limitations and practical considerations for PLIMAN

Although PLIMAN enables rapid, automated image-based phenotyping, several factors can influence measurement accuracy and should be considered when interpreting morphometric data. Whereas VC measurements are influenced primarily by operator-dependent factors such as seed positioning, alignment, pressure application, and measurement consistency, PLIMAN-derived measurements are affected by image quality, background contrast, object orientation, and segmentation accuracy. Consequently, errors introduced during image acquisition or image processing may directly affect the precision of morphometric estimates (Olivoto et al. 2022; Hacisalihoglu and Armstrong 2023; Meraj et al. 2024; Okyere et al. 2023).

Several practical problems were encountered during PLIMAN-based image analysis (Fig. 8). Segmentation failures and erroneous seed counting are often caused by seed overlap or clustering, which hinder object separation. Low contrast between seeds and the background, especially with dark-coated soybean accessions, together with uneven illumination, shadows,

reflections, and background noise, impaired edge detection and, in some instances, led to complete failure to detect edges. In some cases, inconsistent automated detection led to incorrect indexing, duplicate labels or incomplete seed measurements. The above observations suggest that image quality and segmentation performance are major factors in determining measurement accuracy in image-based phenotyping (Okyere et al. 2023; Meraj et al. 2024).

In addition to image quality, several technical factors may affect the PLIMAN measurements. If the camera is not exactly perpendicular to the object, the images will have perspective distortion. Additionally, the edges of smartphone cameras may exhibit optical distortion, which can result in slightly inaccurate estimated dimensions. Additionally, excessive JPEG compression may reduce edge definition compared with lossless image formats, and an improper choice of threshold values may lead to under- or oversegmentation, especially in cases of uneven illumination or for seeds with dark coatings. To minimize these effects, problematic images were reprocessed using optimized segmentation thresholds and improved contrast settings. Optimal segmentation accuracy and reproducible PLIMAN-based measurements therefore require uniform lighting, proper seed spacing, a high-contrast background and standardized imaging geometry (Okyere et al. 2023; Hacisalihoglu and Armstrong 2023).

5. Conclusions

This study demonstrated that PLIMAN is an effective high-throughput platform for characterizing soybean seed morphometric variation across diverse germplasms and contrasting environments. Compared with conventional VC measurements, image-based phenotyping enables the simultaneous extraction of multiple size- and shape-related traits, providing greater phenotypic resolution. Several morphometric traits exhibited high heritability and genetic advancement, indicating their potential utility in soybean improvement programs. Multivariate analyses revealed that seed area and perimeter were the principal contributors to

morphometric variation and effectively differentiated the accessions into distinct seed size groups. Despite these advantages, regression and Bland–Altman analyses revealed poor agreement between the PLIMAN and VC measurements, with the image-based approach systematically overestimated due to differences in measurement principles. Therefore, measurements obtained using the two methods should not be considered directly interchangeable. Standardized image acquisition, segmentation procedures, and calibration protocols are essential to improve reproducibility and facilitate comparisons across studies. Overall, PLIMAN represents a robust, scalable, and cost-effective phenotyping platform for germplasm characterization and breeding applications and provides a valuable foundation for integrating seed morphometric traits into genomic analyses, genome-wide association studies, genomic selection, and crop improvement programs.

Acknowledgments

The authors extend their sincere gratitude to Dr Prashant Dhakephalkar, the director of the Agharkar Research Institute, Pune, for kindly providing the resources and facilities crucial to the successful execution of this study.

Funding

This work is funded by the Ramalingaswami Re-entry Research Program (BT/RLF/Re-entry/01/2021), Department of Biotechnology (DBT), Government of India.

Author information

Authors and Affiliations

Genetics and Plant Breeding Group, Agharkar Research Institute, Pune, Maharashtra 411004, India

Shreyash Gijare, Arathi S., Preetkumar Trivedi, Vaibhav Bhangare, Aditya Gobade, Jayshree Sarode, Vijay Khade, Manoj. D. Oak, Deepak Pawar & Abhinandan S. Patil

Contributions

S.G. performed the data analysis and manuscript preparation. A.S. analyzed the data and wrote the first draft of the manuscript. P.T. performed the Bland–Altman analysis. V.B. contributed to writing the first draft and critically revised the manuscript. A.G. assisted with data acquisition. J.S. assisted with image acquisition. V.K. conducted the field activities. M.O. assisted with R coding for the PLIMAN analysis, review and editing draft. D.P. supported the field operations. A.P. conceived and designed the study, secured funding, supervised the research, monitored the experimental work, and critically revised and approved the final manuscript. All the authors read and approved the final version of the manuscript.

Corresponding author

Correspondence to Abhinandan S. Patil.

ORCID

Abhinandan S. Patil <https://orcid.org/0000-0001-9712-8219>
Preetkumar H. Trivedi <https://orcid.org/0009-0008-6857-6467>
Arathi S. <https://orcid.org/0000-0001-7452-8279>
Vaibhav R. Bhangare <https://orcid.org/0009-0003-3487-4742>
Manoj D. Oak <https://orcid.org/0000-0002-4073-5306>

Ethics Declarations

Ethics approval and consent to participate

This study involved cultivated soybean (*Glycine max* [L.] Merr.) germplasm maintained in the institutional soybean collection and did not involve human participants, animals, or endangered or protected plant species. The collection, maintenance, and use of the soybean plant material complied with applicable institutional guidelines and relevant local and national regulations governing the use of plant genetic resources. As the plant material was cultivated and maintained as part of an institutional germplasm collection, no specific collection permits or licences were required for the research activities described in this study. All plant materials were used in accordance with applicable guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Clinical trial number

Not applicable.

Data availability

All the data supporting the findings of this study are available in the paper and its supplementary materials, which are published online.

References

1. Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with ImageJ. *Biophotonics Int* 11:36-42.
2. Arshad S, Jamil M, Ali A, Saleem Y, Khan UA, Maqsood MF, Zulfiqar U, Ali MF, Binobead MA, AlMunqedhi BM (2025) Exploring multivariate relationships among seed morphometric and yield-related traits in bread wheat (*Triticum aestivum* L.). *Pol J Environ Stud* 34:2003-2010. <https://doi.org/10.15244/pjoes/187779>
3. Baek J, Lee E, Kim N, Kim SL, Choi I, Ji H, Chung YS, Choi MS, Moon JK, Kim KH (2020) High-throughput phenotyping for various traits on soybean seeds using image analysis. *Sensors* 20:248. <https://doi.org/10.3390/s20010248>
4. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327:307-310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)
5. Çetin N (2022) Machine learning for varietal binary classification of soybean (*Glycine max* (L.) Merrill) seeds based on shape and size attributes. *Food Anal Methods* 15:2260-2273. <https://doi.org/10.1007/s12161-022-02286-3>
6. Chiluwal A, Kawashima T, Salmeron M (2022) Soybean seed weight responds to increases in assimilate supply during late seed-fill phase. *J Crop Improv* 36:222-238. <https://doi.org/10.1080/15427528.2021.1943732>
7. Claude J (2008) *Morphometrics with R*. Springer, New York. <https://doi.org/10.1007/978-0-387-77789-4>
8. Egli DB, Bruening WP (2001) Source–sink relationships, seed sucrose levels and seed growth rates in soybean. *Ann Bot* 88:235–242. <https://doi.org/10.1006/anbo.2001.1449>
9. Falconer DS, Mackay TFC (1996) *Introduction to quantitative genetics*, 4th edn. Longman, Harlow
10. Fehr WR, Caviness CE (1977) Stages of soybean development. Special Report 80. Iowa State University Cooperative Extension Service, Ames
11. Gao K, White T, Palaniappan K, Warmund M, Bunyak F (2017) MuSeed: a mobile image analysis application for plant seed morphometry. In: *Proceedings of the 2017*

- IEEE International Conference on Image Processing (ICIP). IEEE, Piscataway, NJ, pp 2826-2830. <https://doi.org/10.1109/ICIP.2017.8296798>
12. Gehan MA, Fahlgren N, Abbasi A, Berry JC, Callen ST, Chavez L, Doust AN, Feldman MJ, Gilbert KB, Hodge JG, Hoyer JS, et al (2017) PlantCV v2: image analysis software for high-throughput plant phenotyping. *PeerJ* 5:e4088. <https://doi.org/10.7717/peerj.4088>
 13. Ghimire A, Kim SH, Cho A, Jang N, Ahn S, Islam MS, Mansoor S, Chung YS, Kim Y (2023) Automatic evaluation of soybean seed traits using RGB image data and a Python algorithm. *Plants* 12:3078. <https://doi.org/10.3390/plants12173078>
 14. Gobade A, Arathi S, Gijare S, Pawar D, Patil AS (2025) Evaluating salt tolerance in soybean core collection: germination response under salinity stress. *Genet Resour Crop Evol* 72:2059-2076. <https://doi.org/10.1007/s10722-024-02081-5>
 15. Hacisalihoglu G, Armstrong P (2023) Crop seed phenomics: focus on non-destructive functional trait phenotyping methods and applications. *Plants* 12:1176. <https://doi.org/10.3390/plants12051177>
 16. Komyshev E, Genaev M, Afonnikov D (2017) Evaluation of the SeedCounter, a mobile application for grain phenotyping. *Front Plant Sci* 7:1990. <https://doi.org/10.3389/fpls.2016.01990>
 17. Meraj T, Sharif MI, Raza M, Alabrah A, Kadry S, Gandomi AH (2024) Computer vision-based plants phenotyping: a comprehensive survey. *iScience* 27:108709. <https://doi.org/10.1016/j.isci.2023.108709>
 18. Miranda MCDC, Aono AH, Pinheiro JB (2023) A novel image-based approach for soybean seed phenotyping using machine learning techniques. *Crop Sci* 63:2665-2684. <https://doi.org/10.1002/csc2.21032>
 19. Moles AT, Ackerly DD, Webb CO, Tweddle JC, Dickie JB, Westoby M (2005) A brief history of seed size. *Science* 307:576-580. <https://doi.org/10.1126/science.1104863>
 20. Nawaz M, Safder S, Riaz S, Ali S (2025) High-resolution RGB image dataset for wheat seed varietal identification and purity assessment. *Data Brief* 61:111690. <https://doi.org/10.1016/j.dib.2025.111690>
 21. Li L, Zhang Q, Huang D (2014) A review of imaging techniques for plant phenotyping. *Sensors* 14:20078-20111. <https://doi.org/10.3390/s141120078>
 22. Ogunniyan D, Makinde S, Adetumbi J, Olosoji J (2021) Assessment of variability in grain yield, seed morphometric and vigour traits of early maturing hybrid maize. *J Agric Sci (Belgrade)* 66:105-119. <https://doi.org/10.2298/JAS2102105O>
 23. Okyere FG, Cudjoe D, Sadeghi-Tehran P, Virlet N, Riche AB, Castle M, Greche L, Mohareb F, Simms D, Mhada M, Hawkesford MJ (2023) Machine learning methods for automatic segmentation of images of field- and glasshouse-based plants for high-throughput phenotyping. *Plants* 12:2035. <https://doi.org/10.3390/plants12102035>
 24. Olivoto T (2022) Lights, camera, PLIMAN! An R package for plant image analysis. *Methods Ecol Evol* 13:789-798. <https://doi.org/10.1111/2041-210X.13803>
 25. Parihar NN, Bhingarde MT (2024) Effect of antioxidant foliar application and harvesting stages on seed morphometry of soybean (*Glycine max* (L.) Merr.). *Int J Adv Biochem Res* 8(Suppl 2):35-46. <https://doi.org/10.33545/26174693.2024.v8.i2Sa.523>
 26. Patil AS, Oak MD, Gijare S, Gobade A, Jaybhay S, Surve VD, Patil GS, Salunkhe D, Waghmare BN, Idhol B, Patil RM, et al (2025) Genome-wide exploration of soybean domestication traits: integrating association mapping and SNP × SNP interaction analyses. *Plant Mol Biol* 115:55. <https://doi.org/10.1007/s11103-025-01583-9>
 27. Poudel S, Khatri D, Pun Magar L, KC S, Mukherjee A, Lucas S, Gebremedhin M, Chilawal A (2024) Final seed size in soybean is determined during mid-seed filling stage. *Agronomy* 14:763.

28. Ray DD, Sen S, Bhattacharyya PK, Bhattacharyya S (2022) Study on seed size variation in soybean (*Glycine max* L. Merr.) and its correlation with yield. *Int J Econ Plants* 9:204-209.
29. Sheikh M, Iqra F, Ambreen H, Pravin KA, Ikra M, Chung YS (2024) Integrating artificial intelligence and high-throughput phenotyping for crop improvement. *J Integr Agric* 23:1787-1802. <https://doi.org/10.1016/j.jia.2023.10.019>
30. Spitzer M, Wildenhain J, Rappsilber J, Tyers M (2014) BoxPlotR: a web tool for generation of box plots. *Nat Methods* 11:121-122. <https://doi.org/10.1038/nmeth.2811>
31. Sun L, Yuan Z (2022) Seed morphology of soybean. In: *Advances in Botanical Research*, vol 102. Academic Press, London, pp 349-375. <https://doi.org/10.1016/bs.abr.2022.03.004>
32. Tanabata T, Shibaya T, Hori K, Ebana K, Yano M (2012) SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiol* 160:1871-1880. <https://doi.org/10.1104/pp.112.205120>
33. Wongchaisuwat P, Chakranon P, Yinpin A, Onwimol D, Wonggasem K (2025) Rapid maize seed vigor classification using deep learning and hyperspectral imaging techniques. *Smart Agric Technol* 10:100820.
34. Xu G, Zhang X (2023) Mechanisms controlling seed size by early endosperm development. *Seed Biol* 2:1. <https://doi.org/10.48130/SeedBio-2023-0001>
35. Yang M, Chen T, Liu YX, Huang L (2024) Visualizing set relationships: EVenN's comprehensive approach to Venn diagrams. *iMeta* 3:e184. <https://doi.org/10.1002/imt2.184>
36. Yang S, Zheng L, He P, Wu T, Sun S, Wang M (2021) High-throughput soybean seeds phenotyping with convolutional neural networks and transfer learning. *Plant Methods* 17:50. <https://doi.org/10.1186/s13007-021-00749-y>
37. Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, Xiong L, Yan J (2020) Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol Plant* 13:187–214. <https://doi.org/10.1016/j.molp.2020.01.008>
38. Zhang Y, Bhat JA, Zhang Y, Yang S (2024) Understanding the molecular regulatory networks of seed size in soybean. *Int J Mol Sci* 25:1441. <https://doi.org/10.3390/ijms25031441>
39. Zhou X, Wang D, Mao Y, Zhou Y, Zhao L, Zhang C, Liu Y, Chen J (2022) The organ size and morphological change during the domestication process of soybean. *Front Plant Sci* 13:913238. <https://doi.org/10.3389/fpls.2022.913238>
40. Zhu F, Paul P, Hussain W, Wallman K, Dhatt BK, Sandhu J, Irvin L, Morota G, Yu H, Walia H (2021) SeedExtractor: an open-source GUI for seed image analysis. *Front Plant Sci* 11:581546. <https://doi.org/10.3389/fpls.2020.581546>

Tables

Table 1 Seed morphometric traits and genetic parameters of soybean accessions according to the 'PLIMAN' package and VC

Method	Parameters	Environment	Mean $\pm$ SE	Range	CV (%)	Frequency distribution		Genetic Parameters				F Ratio
						Skewness	Kurtosis	PCV (%)	GCV (%)	Heritability (%)	GA % of the mean	
PLIMAN package	Area	E1	125.34 $\pm$ 3.26	9.80-235.90	36.03	-0.44	0.49	88.15	44.07	24.99	45.38	0.43*
		E2	151.82 $\pm$ 3.75	10.50-309.40	34.20	-0.57	1.79	32.32	31.84	97.04	64.61	0.47*
	Perimeter	E1	163.49 $\pm$ 2.72	23.09-271.40	23.07	-0.24	0.79	26.97	25.11	86.69	48.17	1.41*
		E2	189.18 $\pm$ 3.18	23.09-326.20	23.29	-0.60	1.13	21.93	20.16	84.53	38.18	2.91*
	Radius	E1	21.05 $\pm$ 0.36	1.72-30.90	23.53	-1.53	3.51	24.84	23.78	91.63	46.89	2.94*
		E2	23.11 $\pm$ 0.40	1.72-36.00	24.05	-1.54	3.43	22.01	21.37	94.28	42.75	2.92*
	Min radius	E1	16.88 $\pm$ 0.33	0.19-25.30	27.32	-2.09	5.39	29.11	27.81	91.29	54.74	2.95*
		E2	18.67 $\pm$ 0.37	0.19-29.10	27.21	-2.11	5.64	26.50	24.99	88.91	48.55	2.80*
	Max radius	E1	24.58 $\pm$ 0.38	3.55-37.10	21.69	-0.82	1.76	24.59	22.14	81.06	41.07	2.93*
		E2	30.49 $\pm$ 0.46	3.55-45.70	21.04	-1.72	4.75	21.12	19.96	89.37	38.88	2.73*
	Radius ratio	E1	0.44 $\pm$ 0.02	0.30-2.50	61.51	5.10	28.46	66.78	54.81	67.37	92.68	1.16*
		E2	0.53 $\pm$ 0.04	0.34-3.70	73.08	5.65	36.23	174.04	52.09	8.96	32.12	0.97*
	Diameter	E1	42.10 $\pm$ 0.71	3.44-61.80	23.53	-1.53	3.51	24.80	23.73	91.61	46.79	2.94*
		E2	46.21 $\pm$ 0.80	3.44-71.90	24.05	-1.54	3.43	21.97	21.33	94.27	42.66	2.92*
	Min diameter	E1	34.71 $\pm$ 0.66	0.42-51.60	26.52	-2.09	5.54	30.10	28.90	92.21	57.17	2.95*
		E2	38.14 $\pm$ 0.72	0.42-58.10	26.18	-2.19	6.17	26.50	24.99	88.91	48.55	2.91*
	Max diameter	E1	49.03 $\pm$ 0.77	8.42-74.20	21.74	-0.84	1.81	24.59	22.14	81.06	41.07	2.93*
		E2	60.55 $\pm$ 0.92	5.65-85.30	20.96	-1.57	3.66	21.12	19.96	89.37	38.88	2.73*
	Eccentricity	E1	0.52 $\pm$ 0.01	0.32-1.00	20.42	1.73	4.00	23.14	19.00	67.45	32.15	0.65*
		E2	0.56 $\pm$ 0.01	0.16-1.00	17.09	1.25	6.00	19.50	15.83	65.87	26.46	0.67*
	Solidity	E1	0.96 $\pm$ 0.01	0.19-1.10	12.02	-4.38	20.32	13.68	11.15	66.33	18.70	0.44*
		E2	0.95 $\pm$ 0.01	0.36-1.20	10.26	-4.35	20.15	12.36	9.08	53.95	13.74	0.67*
	Circularity	E1	14.17 $\pm$ 1.02	0.43-106.40	99.94	2.93	15.36	118.76	107.18	81.46	199.28	0.34*
		E2	21.63 $\pm$ 0.70	0.88-106.40	44.64	5.50	39.07	61.56	53.00	74.14	94.01	0.87*
VC	Area	E1	30.97 $\pm$ 0.38	16.74-51.31	17.11	0.61	1.23	18.94	16.18	72.95	28.47	0.60*
		E2	32 $\pm$ 0.44	17.45-55.11	19.27	0.68	0.99	220.2	19.11	75.35	34.18	1.26*
	Perimeter	E1	61.95 $\pm$ 0.76	33.47-102.62	17.11	0.61	1.23	9.54	8.04	71.05	13.96	0.57*
		E2	63.99 $\pm$ 0.89	34.89-110.22	19.27	0.68	0.99	10.26	9.18	79.98	16.91	1.43*

*Significant at the 0.05 probability level; CV: coefficient of variation (%); SE: standard error; E1: environment 1; E2: environment 2; PCV: phenotypic coefficient of variation; GCV: genotypic coefficient of variation (%); H2: broad-sense heritability (%); GA: genetic advancement

Table 2 Genotypes with large and small seeds assessed by the PLIMAN-based method and VC

PLIMAN package											
Large Seeds											
Code	Genotypes	R area E1	R area E2	R area mean	Code	Genotypes	R peri E1	R peri E2	R peri Mean		
ML231	MACS 288	217.60	277.30	247.45	EM 185	PK 1029	271.43	271.43	271.43		
ML144	TS 17	185.30	309.40	247.35	ML231	MACS 288	207.82	326.18	267.00		
	JS (SH) 99										
EM76	14	235.90	235.90	235.90	EM 216	MACS 37	257.67	257.67	257.67		
ML201	MACS 112	180.00	274.70	227.35	EM 237	MACS 173	255.20	255.20	255.20		
ARI 6	EC 38828	216.30	223.06	219.68	EM76	JS (SH) 99 14	249.39	249.39	249.39		
Small Seeds											
MM 93	Green Tur	12.20	12.20	12.20	MM 64	EC 14437	97.71	97.71	97.71		
MM 21	EC 34079A	11.40	11.40	11.40	MM 71	EC 39795	69.97	69.97	69.97		
PC24	JSSH 9337	11.20	11.20	11.20	PC25	KALITUR	67.74	67.74	67.74		
MM 71	EC 39795	10.70	10.70	10.70	MM 70	EC 39776	64.57	64.57	64.57		
MM 70	EC 39776	10.50	10.50	10.50	ML226	MACS 261	23.09	23.09	23.09		
VC											
Large Seeds											
Code	Genotypes	VC area E1	VC area E2	VC area mean	Code	Genotypes	VC peri E1	VC peri E2	VC peri mean		
PC67	NRC 150	51.31	55.11	53.21	PC67	NRC 150	102.62	110.22	106.42		
ARI 6	EC 38828	45.94	43.57	44.75	ARI 6	EC 38828	91.88	87.14	89.51		
EM68	PS 1370	36.87	50.71	43.79	EM68	PS 1370	73.74	101.42	87.58		
PC21	JS 9305	36.27	51.00	43.63	PC21	JS 9305	72.53	101.99	87.26		
EM1	EC 14477	43.52	39.19	41.36	EM1	EC 14477	87.04	78.38	82.71		
Small Seeds											
EM 151	IC 13048	24.55	21.78	23.17	EM 151	IC 13048	49.11	43.55	46.33		
MM 122	IC 13053	18.35	26.66	22.50	MM 122	IC 13053	36.69	53.33	45.01		
ML169	VLS 67	20.66	22.61	21.63	ML169	VLS 67	41.32	45.22	43.27		
EM 219	MACS 94	25.62	17.45	21.53	EM 219	MACS 94	51.24	34.89	43.07		
MM 70	EC 39776	20.38	21.77	21.07	MM 70	EC 39776	40.76	43.54	42.15		

Table 3 Categorization of 192 soybean accessions on the basis of seed size

Environments	Range					
	R			VC		
	Large-sized seeds (≥ 100 mm ²)	Medium-sized seeds (50-100 mm ²)	Small-sized seeds (≤ 50 mm ²)	Large-sized seeds (≥ 50 mm ²)	Medium-sized seeds (20-50 mm ²)	Small-sized seeds (≤ 50 mm ²)
E1	13	36	143	1	186	5
E2	13	5	174	2	188	2

Figures

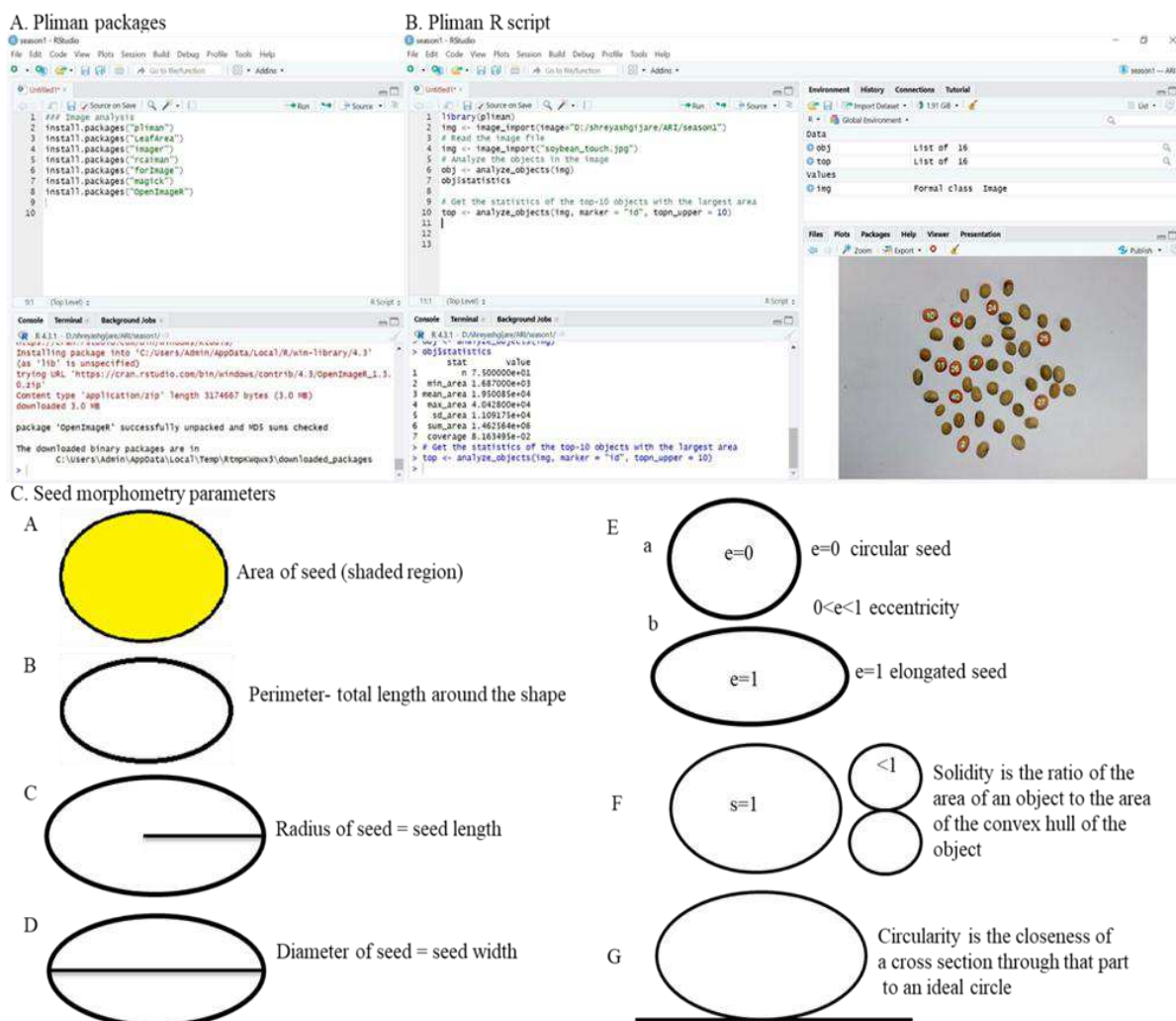
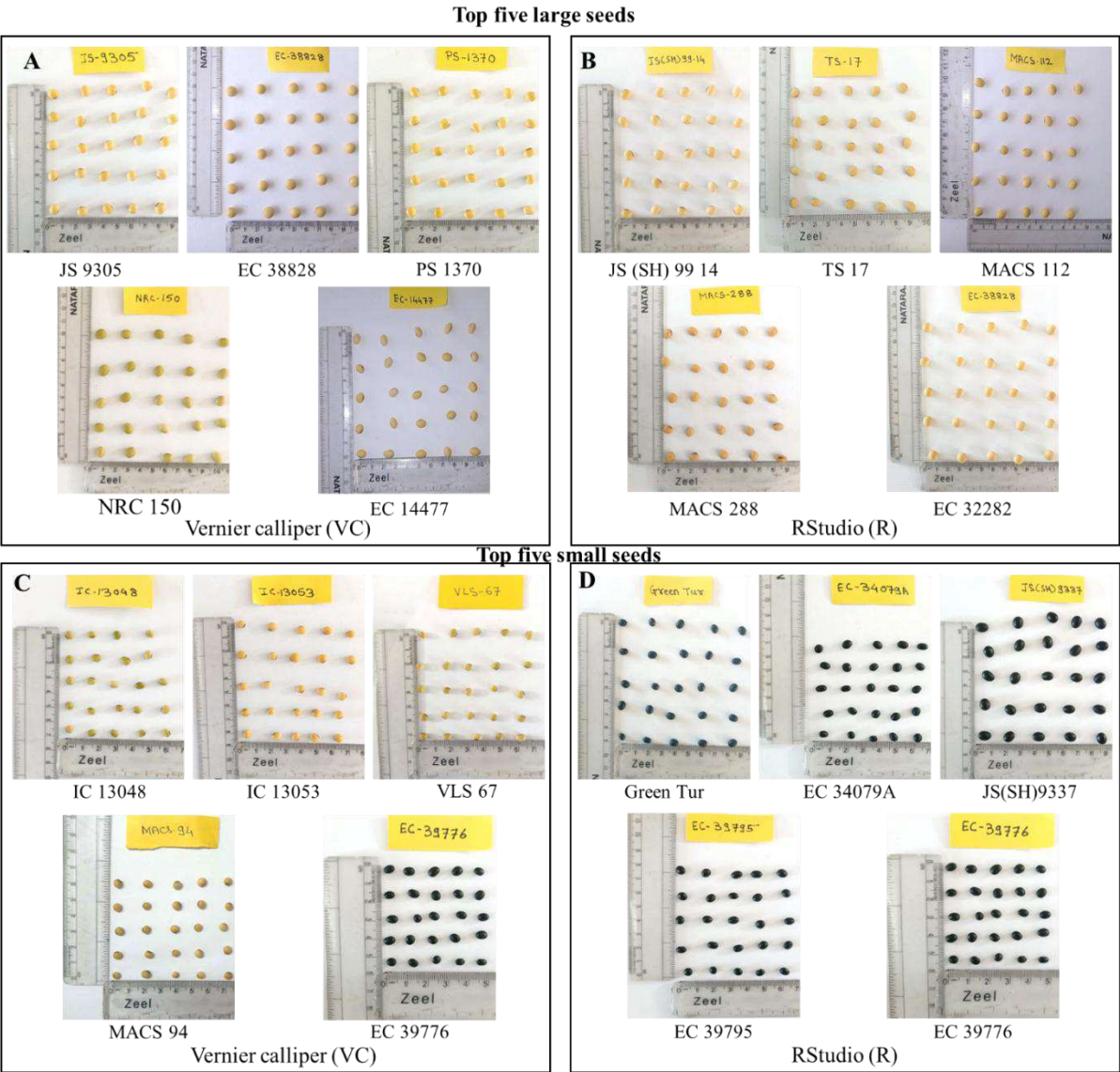


Fig. 1 Image-based soybean seed morphometric analysis using the PLIMAN package in RStudio. (A) Representative RStudio interface showing the PLIMAN workflow, including package loading, image import, object segmentation, extraction of morphometric statistics, and automatic identification and numbering of individual soybean seeds. (B) PLIMAN-extracted quantitative seed morphometric traits, including projected area, perimeter, mean, minimum and maximum radius, mean, minimum and maximum diameter, eccentricity, solidity, and circularity (C) Schematic representation of shape descriptors used in the analysis, illustrating circularity, eccentricity (e), and solidity (s). Eccentricity ranges from $e = 0$ for a perfect circle to $e = 1$ for a highly elongated ellipse, whereas solidity equals $s = 1$ for a convex object and decreases (< 1) as boundary concavities increase.



854

855

856

857

858

859

Fig. 2 Phenotypic variation in seed size and coat color across selected genotypes. High-throughput image analysis revealed the top large-seeded and small-seeded groups. (A) Large seeds by area obtained from VC (B), large seeds by area obtained from R (C), and small-seeded genotypes identified via VC analysis, including (D) small-seeded genotypes identified via R analysis.

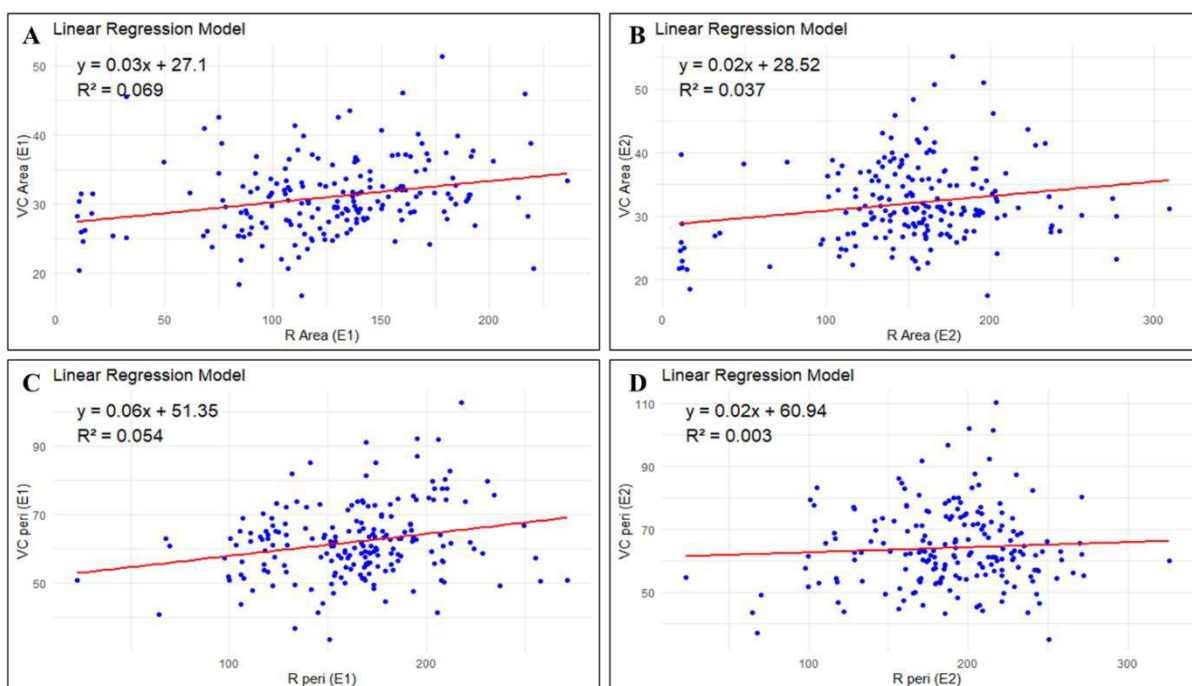


Fig. 3 Regression analysis of the area and perimeter at environments E1 and E2 obtained by R and VC: (A) area (R_area_E1 and VC_area_E1); (B) area (R_area_E2 and VC_area_E2); (C) perimeter (R_peri_E1 and VC_peri_E1); (D) perimeter (R_peri_E2 and VC_peri_E2)

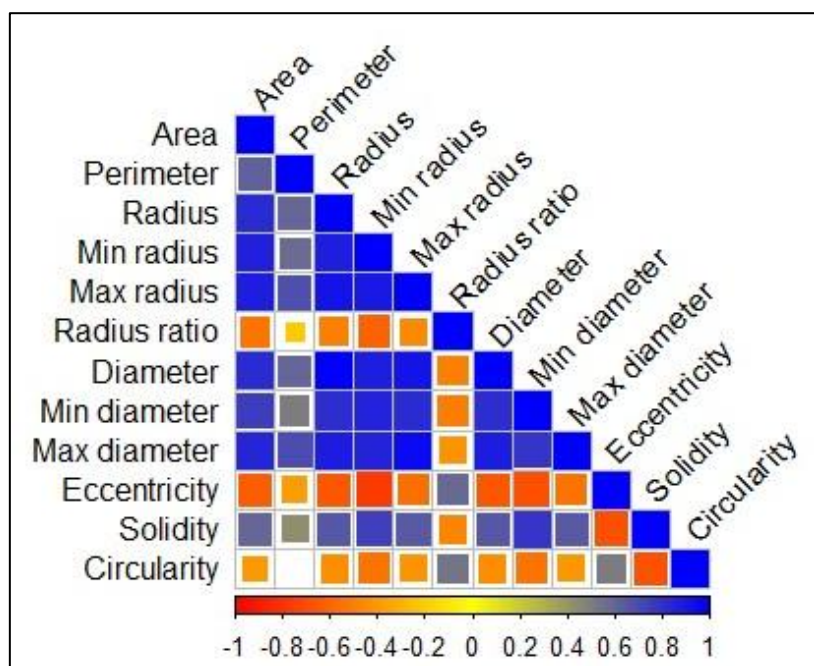
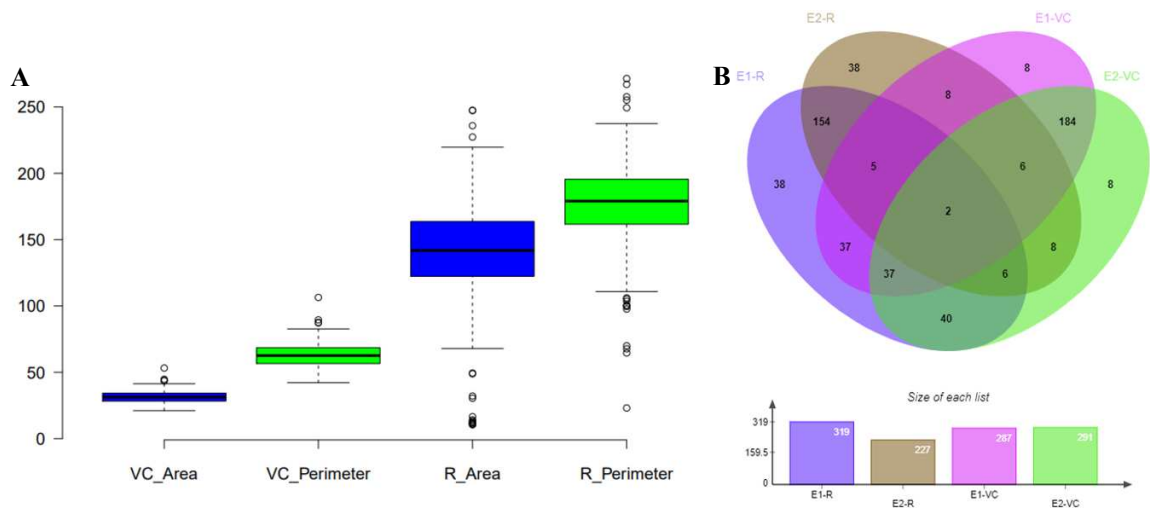


Fig. 4 Correlation matrix heatmap of the morphometric parameters of 192 soybean accessions obtained from PLIMAN-derived

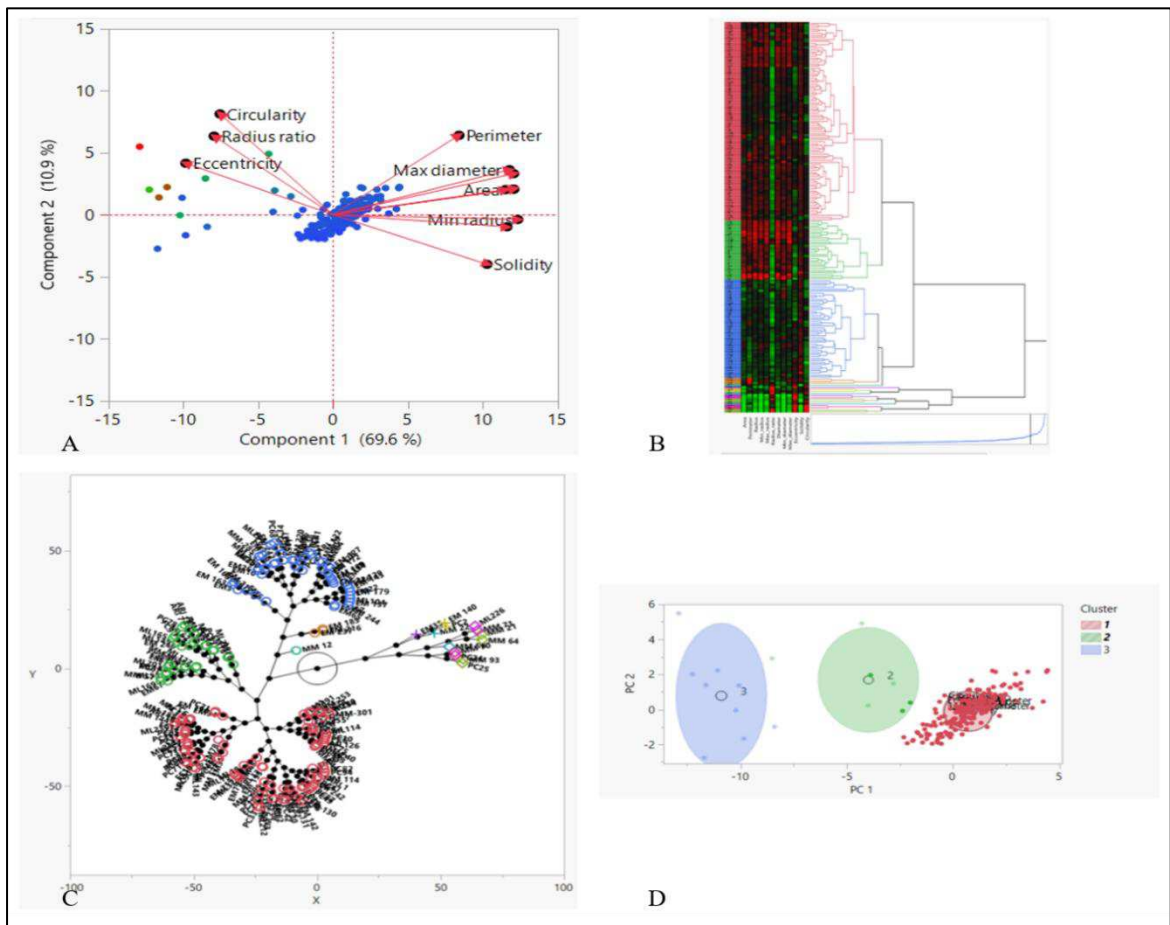
874



875

876 **Fig 5** Comparison and categorization of soybean seed morphometric measurements obtained
877 using the PLIMAN-based and VC methods. (A) Box plot comparing area and perimeter
878 measurements obtained by PLIMAN-derived and VC; (B) Venn diagram showing overlap of
879 genotype classification across E1 and E2 using both methods

880



881

882 **Fig. 6** Multivariate analysis of soybean seed morphometric traits. (A) Principal component
883 analysis (PCA); (B) hierarchical clustering; (C) Constellation plot; (D) K-means clustering of
884 192 soybean accessions based on seed morphometric parameters

885

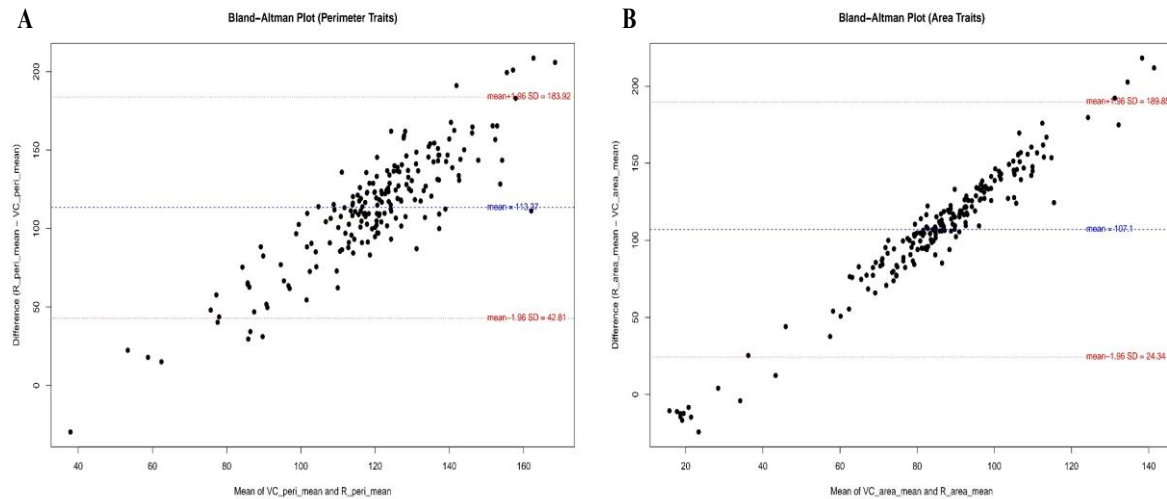


Fig. 7 The Bland–Altman plots illustrate the agreement between the VC and PLIMAN-derived measurements for the following soybean seed traits: (A) perimeter and (B) area. The central dashed line represents the mean difference, and the outer dotted lines indicate the 95% limits of agreement. For both traits, the widening spread of differences at higher mean values suggests a proportional bias between the two methods.

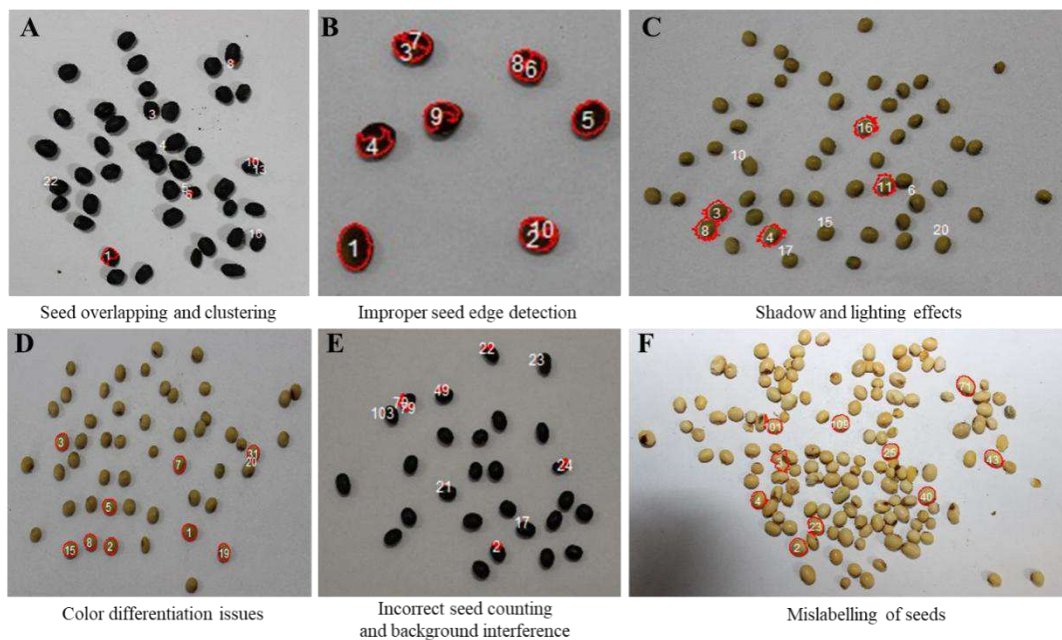


Fig. 8 Common sources of error in the PLIMAN-based soybean seed image analysis. Representative examples of image acquisition and segmentation errors affecting morphometric measurements. (A) Overlapping or clustered seeds causing segmentation failures. (B) Low contrast between the seeds and the background results in poor boundary recognition. (C) Uneven illumination and shadows interfering with edge detection. (D) Reflections and background noise causing incomplete segmentation. (E) Incorrect seed counting due to misidentification of adjacent seeds. (F) Mislabeling or duplicate numbering caused by inconsistent automated detection.

Supplementary Tables (Attached in the Supplementary file)

Supplementary Table S1 Details of the 192 soybeans used for the study.

Supplementary Table S2 Means of 192 genotypes by PLIMAN-based analysis for seed morphometric traits

Supplementary Table S3 Means of 192 genotypes by VC-based analysis for seed morphometric traits

Supplementary Table S4 Correlation coefficients and p values.

Supplementary Table S5 Eigenvalues and eigenvectors of the morphometric parameters from the PLIMAN-based dataset.

Supplementary Table S6 100-seed weights of all 192 genotypes.

Supplementary Table S7 Common genotypes in E1 and E2 (two-way, three-way, and four-way interactions) obtained from PLIMAN-based and VC.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalTableS1S7.xlsx](#)