Pesquisa Agropecuária Brasileira

Genetics/ Original Article

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Phenomic data for pattern recognition in soybean lines

Abstract – The objective of this work was to extract different datasets from soybean leaflets images taking into account shape and texture traits, Fourier descriptors, and vegetative indices. From images of 1,105 lines, 127 attributes were grouped into four sets, as follows: 31 in shape, 13 in texture, 43 in vegetation indices, and 40 in elliptical Fourier descriptors (EFD). Redundancy was evaluated by Pearson correlations and genotypic discrimination by the principal component analysis, the Procrustes and Mantel tests on standardized Euclidean distances, and k-means clustering. The shape set was the most redundant (69 of 465 correlations >0.95), whereas the EFD showed the lowest redundancy and required more components to summarize their variation. Both tests ranked vegetation indices as the set most concordant with shape and texture, and k-means clustering identified five genotype groups across the datasets. The image-based phenomic analysis can support the characterization of breeding lines by recognizing phenotypic patterns and identifying the most informative attributes for prediction and classification in breeding programs.

Index terms: *Glycine max*, shape traits, texture traits, vegetative indices.

Dados fenômicos para o reconhecimento de padrões em linhagens de soja

Resumo – O objetivo deste trabalho foi extrair diferentes conjuntos de dados a partir de imagens de folíolos de soja, ao se considerar os atributos forma e textura, os descritores de Fourier e índices de vegetação. A partir de imagens de 1.105 linhagens, 127 atributos foram agrupados em quatro conjuntos, como a seguir: 31 em forma, 13 em textura, 43 em índices de vegetação e 40 em descritores elípticos de Fourier (DEF). Avaliaram-se redundância por correlações de Pearson e discriminação genotípica por análise de componentes principais, pelos testes de Procrustes e de Mantel sobre distâncias euclidianas padronizadas e pelo agrupamento por médias k. O conjunto forma foi o mais redundante (69 de 465 correlações > 0,95), enquanto os DEF apresentaram menor redundância e exigiram mais componentes para resumir sua variação. Ambos classificaram os índices de vegetação como o conjunto mais concordante com forma e textura, e o agrupamento por médias k identificou cinco grupos genotípicos entre os conjuntos. A análise fenômica baseada em imagens pode ajudar na caracterização de linhagens por meio do reconhecimento de padrões fenotípicos e da identificação dos atributos mais informativos para predição e classificação em programas de melhoramento.

Termos para indexação: *Glycine max*, características de forma, características de textura, índices vegetativos.

Wagner Faria Barbosa
Universidade Federal de Viçosa, Departamento de Biologia Geral, Viçosa, MG, Brazil.
E-mail: wagner.barbosa@ufv.br

Cosme Damião Cruz
Universidade Federal de Viçosa, Departamento de Biologia Geral, Viçosa, MG, Brazil.
E-mail: cdcruz@ufv.br

Ivan Ricardo Carvalho✉
Universidade Regional do Noroeste do Rio Grande do Sul, Departamento de Estudos Agrários, Ijuí, RS, Brazil.
E-mail: carvalhoirc@gmail.com

Jaqueline Piesanti Sangiovo
Universidade Federal de Santa Maria, Departamento de Fitotecnia, Santa Maria, RS, Brazil. E-mail: jpsangiovo@gmail.com

Willyan Júnior Adorian Bandeira
Universidade Federal de Santa Maria, Departamento de Fitotecnia, Santa Maria, RS, Brazil. E-mail: bandeira.wjab@gmail.com

Marco Luís Conde Machado de Souza
Universidade Federal de Viçosa, Departamento de Biologia Geral, Viçosa, MG, Brazil.
E-mail: conde-marco@hotmail.com

✉ Corresponding author

Received
January 19, 2026

Accepted
August 3, 2026

How to cite
BARBOSA, W.F.; CRUZ, C.D.; CARVALHO, I.R.; SANGIOVO, J.P.; BANDEIRA, W.J.A.; SOUZA, M.L.C.M. de. Phenomic data for pattern recognition in soybean lines. **Pesquisa Agropecuária Brasileira**, v.61, e04359, 2026. DOI: https://doi.org/10.1590/S1678-3921.pab2026.v61.04359.

This is an open-access article distributed under the Creative Commons Attribution 4.0 International License

Introduction

Phenomics systematically studies phenotypes across diverse dimensions of an organism biology, which demands an interdisciplinary approach integrating molecular, biochemical, cellular, genetic, and morphological levels of analysis (Jangra et al., 2021). Advances in image acquisition and processing have substantially accelerated this field, allowing of plant material to be characterized without destruction (Gill et al., 2022) and generating datasets of unprecedented scale. Image-based phenotyping stands out for its precision, accuracy, and speed, often outperforming conventional approaches that remain vulnerable to observer bias and higher operational costs (Duddu et al., 2019). By extending evaluation to structural levels, once costly or technically inaccessible, imaging has broadened the scope of what can be measured (Visakh et al., 2024), with applicable techniques ranging from traditional photography and light microscopy to thermography, fluorescence, tomography, and nuclear magnetic resonance imaging (Zhang et al., 2024).

Within agriculture, imaging-based tools are reshaping how decisions are made, adding precision and efficiency to a domain that depends heavily on timely information (Omari et al., 2020; Nguyen et al., 2025). Digital imaging has become progressively embedded in crop breeding, a discipline central to agricultural progress (Cortes et al., 2017; Dipta et al., 2023). Substantial resources have gone into developing high-throughput phenotyping platforms capable of collecting, processing, and storing the image data that plant breeding programs now depend on (Chawade et al., 2019; Jangra et al., 2021).

Among the most economically important crops worldwide, soybean (*Glycine max* (L.) Merr.) has benefited considerably from image analysis across multiple stages of the breeding pipeline (Silva Junior et al., 2018; Alabi et al., 2022). The use of image analysis to characterize the morphology of soybean vegetative structures is now a consolidated practice (Momin et al., 2017; Shilpashree et al., 2021; He et al., 2022; Xing et al., 2022). Beyond distinguishing lines or cultivars, traits such as leaf shape, texture, and color are associated with agronomically relevant attributes, including yield potential, resistance to pest and disease, and adaptability to environmental variation (Ren et al., 2019; Rowland et al., 2020).

Shape-related traits, such as area, perimeter, length, and width of vegetative organs, were traditionally obtained through manual measurement. These traits can now be extracted with accuracy through digital image analysis (Olivoto, 2022). Two-dimensional contours, in turn, can be described mathematically through elliptical Fourier descriptors (EFDs), an approach first proposed by Kuhl & Giardina (1982), which is still applied today to capture shape variation in seeds and leaves across crop species (Khan et al., 2024). Texture and color, though perceptible to the eye, require computational metrics for rigorous analysis. Texture is characterized through the spatial dependency of grayscale values, a principle established by Haralick et al. (1973) still employed today, for instance, in crop classification from remote sensing imagery (Iqbal et al., 2021). Colorimetric traits, meanwhile, can be quantified through vegetative indices (VIs) derived from image bands (for instance, RGB), a strategy that increases sensitivity to the detection of specific traits (Ge et al., 2025).

The objective of this work was to extract different datasets from soybean leaflets images taking into account shape and texture traits, Fourier descriptors, and vegetative indices. Furthermore, it evaluates the ability of these datasets to discriminate and identify distinct patterns among breeding lines under development, providing insights for improving genetic selection and breeding strategies.

Materials and Methods

The experiment with soybean lines in the F_3 generation was carried out in the experimental field of the Universidade Regional do Noroeste do Estado do Rio Grande do Sul (Unijuí), located in the municipality of Ijuí, in the state of Rio Grande do Sul (RS) (28°23'16"S, 53°54'53"W). Each line was cultivated in 2×0.5 m plots, organized into augmented blocks with interspersed four control cultivars. Fifteen blocks were established, each one containing about 100 lines, which resulted in a total of 1,449. Sowing occurred on three dates: lines from 1 to 1,000 were planted in October 14, 2022; lines 1,001 to 1,300, on October 17, 2022; and lines 1,301 to 1,499, on October 18, 2022. Fertilization involved the pre-application of 100 kg per block of organic fertilizer, 8 metric tons of Biogranum,

and 6 kg/block of a 3-21-21 (N-P-K) fertilizer applied in two equal doses of 3 kg each.

The phenotypic analyses were based on RGB images of leaflets collected at the R_5 growth stage. Between 8:00 h and 4:00 h, one leaf from the middle third of three plants was collected from each line. The leaves were transported to the laboratory, detached and arranged on a red cardboard in three horizontal rows. A yellow reference marker with 18 cm^2 was included to convert pixel metrics into standard measurement units. The images were captured using an imaging development systems (IDS) camera controlled by the uEye Cockpit software, positioned 50 cm above the leaflets, which were placed on an LED light pad. Images were acquired at 1,280×1,024 pixels and 96 dpi resolution. All images were systematically named. Due to 23.74% cultivation loss, images were taken for 1,105 lines out of the 1,449 initially planted, resulting in 9,945 observations (9 leaflets per line). This number corresponded to three fully expanded trifoliolate leaves per line, adopted to capture within line variability in leaflet shape, texture, and color.

In the present work, soybean leaflet attributes were grouped into four sets by information type: general shape (area, perimeter, and length), texture (Haralick descriptors), vegetation indices (VIs) derived from the RGB bands, and elliptical Fourier descriptors of the leaflet contour. The four sets comprised 31 shape attributes, 13 texture attributes, 43 vegetation indices, and 40 elliptical Fourier descriptors, totaling 127 attributes subjected to pattern recognition. Each observational unit corresponded to the mean attribute values of the leaflets of a given soybean line. Within each set, redundant traits (Pearson $r > 0.95$), replicated traits ($r = 1$), and highly correlated blocks were identified from correlation networks. The PCA allowed of the evaluation of phenotypic differentiation through scatterplots of scores for the first two principal components, as well as for the dimensionality reduction by determining the minimal number of components needed, and for the relative contribution of each attribute to the total variability of each phenotypic set. Subsequently, the scores from the first two principal components were used in a Procrustes analysis, to assess the concordance of dissimilarity patterns across phenotypic datasets. A Mantel test was also conducted to measure the association between dissimilarity matrices generated by each dataset, using the Pearson correlation coefficients to compare standardized Euclidean distances between genotypes. The standardized Euclidean distances were calculated as

$$D_{i,i'} = \sqrt{\sum_{k=1}^{p} \frac{(x_{ik} - x_{i'k})^2}{s_{kk}}} \ ,$$

where: x_{ik} is the value of trait k for genotype i; $x_{i'k}$ is the value of trait k for genotype i'; s_{kk} is the variance of trait k, and p is the total number of traits.

Scatterplots of pairwise phenomic distance matrices were generated to infer concordance graphically. Points concentrated in quadrants 1 and 3 indicate similarity/dissimilarity agreement, while concentration in quadrants 2 and 4 suggests divergences.

The pattern recognition was performed by k-means clustering for each set, with the number of clusters (k) defined by the elbow method (Bishop, 2006). The concordance between phenotypic information sets in cluster formation was assessed after the clustering process. The Jaccard indices were calculated for each pair of clusters formed by the different phenotypic information sets. Therefore, the concordance proportion (P_{XY}) between two phenotypic information sets X and Y was defined as

$$P_{XY} = \frac{2}{n+m} \sum_{i=1}^{n} \sum_{j=1}^{m} p_{XYij} = \frac{2nm\bar{p}_{XY}}{n+m} \ ,$$

where: n and m are the numbers of clusters formed by X and Y, respectively; $p_{XY_{ij}} = \frac{N_{X_i} \cap Y_j}{N_{X_i} \cup Y_j}$, and it represents the proportion of the intersection over the union of clusters X_i and Y_j, commonly referred to as the Jaccard index; $\bar{p}_{XY}$ represents the arithmetic mean of the Jaccard indexes between clusters formed from the attribute sets X and Y, and it is calculated as

$$\bar{p}_{XY} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} p_{XYij} \ .$$

Data processing was performed using the R version 4.2.3 (https://cran.r-project.org/) integrated with RStudio (https://posit.co/download/rstudio-desktop/). Phenotypic attributes were extracted using the pliman package (Olivoto, 2022), while data manipulation was facilitated by the tidyverse package version 2.0.0. Attributes extracted from each leaflet were averaged to

generate representative values for each line, reducing the 9,945 individual observations to 1,105 summarized records.

Results and Discussion

The shape dataset exhibited the highest internal redundancy, with many attribute pairs showing strong correlations evidenced by the intense red tones (Figure 1). Specifically, 69 out of the 465 calculated correlations in this dataset had Pearson correlation coefficients greater than 0.95 and were classified as redundant. Furthermore, some shape attributes exhibited perfect correlations (Pearson = 1), such as those between average diameter and average radius and between maximum diameter and maximum radius. Despite this redundancy, size, area, and perimeter-related attributes offer practical criteria for differentiation, as they are easy for breeders to perceive and measure.

Conversely, the elliptical Fourier descriptors dataset showed the lowest inter-attribute associations, suggesting that each attribute provides unique information that is not represented by others within the same set. Although Fourier morphological description generates four coefficients per harmonic, which may not be entirely independent (Haines & Crampton, 2000), the limited number of highly correlated attributes observed here can be attributed to the use of average leaflet coefficients to represent each line in the database.

This contrast is expected from how each set is constructed: elliptical Fourier descriptors encode contour geometry as harmonics independently of size (Kuhl & Giardina, 1982), whereas shape metrics such as diameter and radius are near-duplicates by definition, which explains the perfect correlations found here. The practical consequence runs opposite to the redundancy. Although the shape set is the least informative statistically, its size, area, and perimeter descriptors remain the traits breeders measure most readily, and leaf shape alone has predicted cultivar performance in tomato (Rowland et al., 2020).

Beyond characterizing the phenomic datasets and identifying redundant attributes, the PCA was used to explore patterns of similarity among the breeding lines. A significant cluster of lines with high-similarity patterns was concentrated near the origin of the PCA scatter plots. At the same time, some dispersed points indicated distinct lines warranting further investigation into agronomic traits (Figure 2). The PCA is advantageous for summarizing information into a few components when redundancy exists, enabling structural simplification. The redundancy levels observed (in Figure 1) are confirmed (in Figure 3), with about 80% (or more) of the variance explained by two principal components for all datasets, except for the Fourier descriptors. This lack of redundancy highlights the unique potential of Fourier descriptors in pattern recognition or discrimination processes, though caution is advised when structurally simplifying this dataset.

Broadly, attribute importance measures are displayed from each phenomic dataset concerning the first two principal components (Figure 3). Attributes with positive correlations cluster together (forming angles < 90°), while negatively correlated attributes are on opposite sides of the origin (> 90°). A color gradient at the top of each graph indicates the contribution of each attribute, with attributes close to the origin deemed less critical (Kassambara, 2017). A notable number of highly correlated attributes were observed within the shape and vegetative index datasets. Nevertheless, only a few attributes in each dataset showed minimal importance (blue tones) for forming the first two principal components, which captured the most data variance.

Dimensionality reduction using PCA for phenomic data from soybean leaflets proves valuable for statistical studies, particularly where high dimensionality presents challenges. Identifying key and less critical attributes is crucial in breeding studies, where agronomic traits differ in measurement difficulty due to labor, timing, environmental effects, and costs. Discarding redundant or invariant attributes could reduce costs and optimize resources.

In this work, image analysis depended solely on computational tools capable of extracting maximum information from refined images. Even if attribute discarding is unnecessary, (Figure 3) identifies attributes of greater interest for breeders to associate with other agronomic, morphological, or physiological traits. The scores of the first two PCA components also allow of graphical agreement verification between two datasets, by using the Procrustes analysis. The shape and vegetative index datasets showed the highest

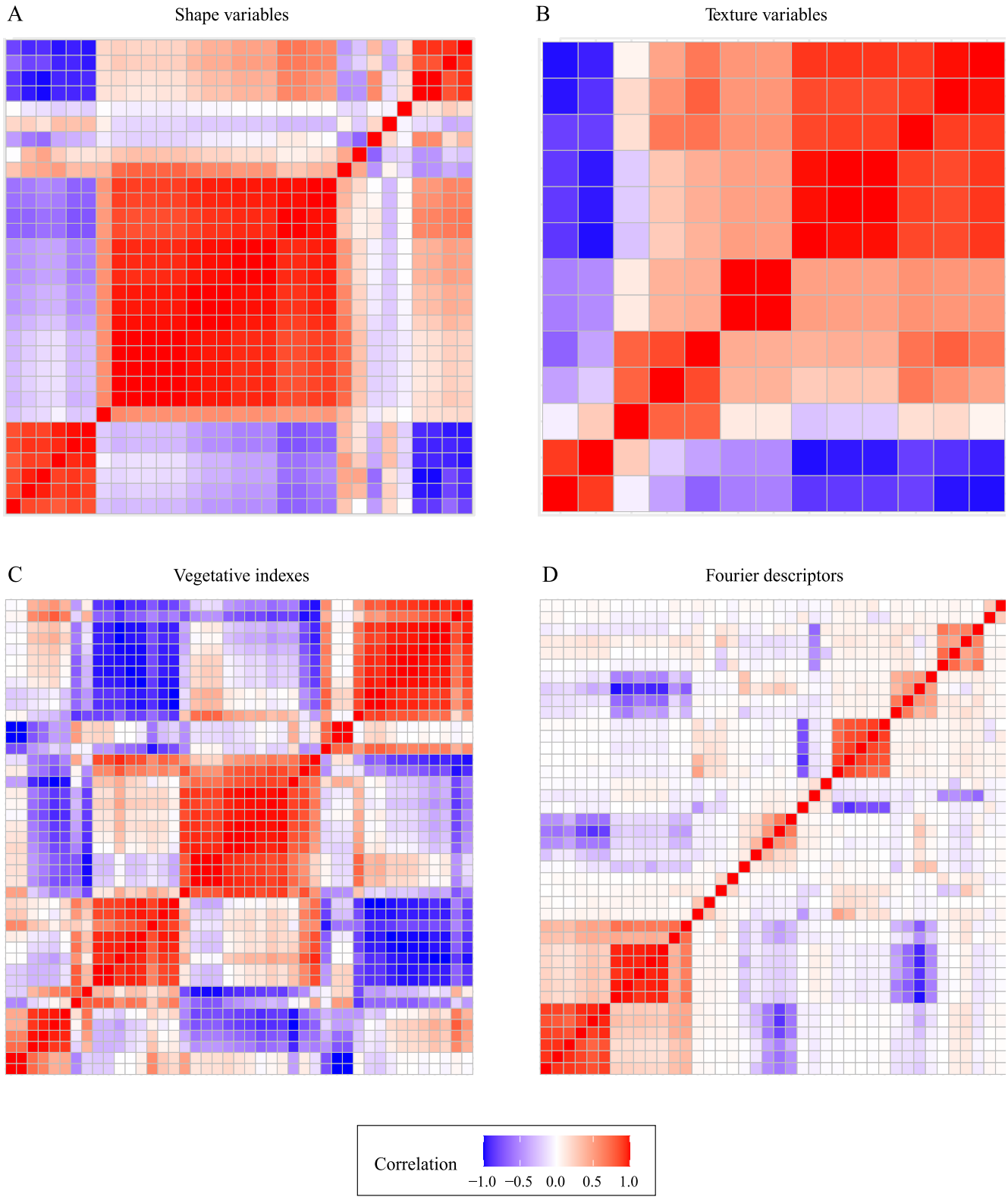

Figure 1. Colorimetric representation of Pearson correlations derived from attributes related to shape variables (A), texture variables (B), vegetative indexes (C), and Fourier descriptors (D) of extracted from leaflets digital images of soybean lines (*Glycine max*).

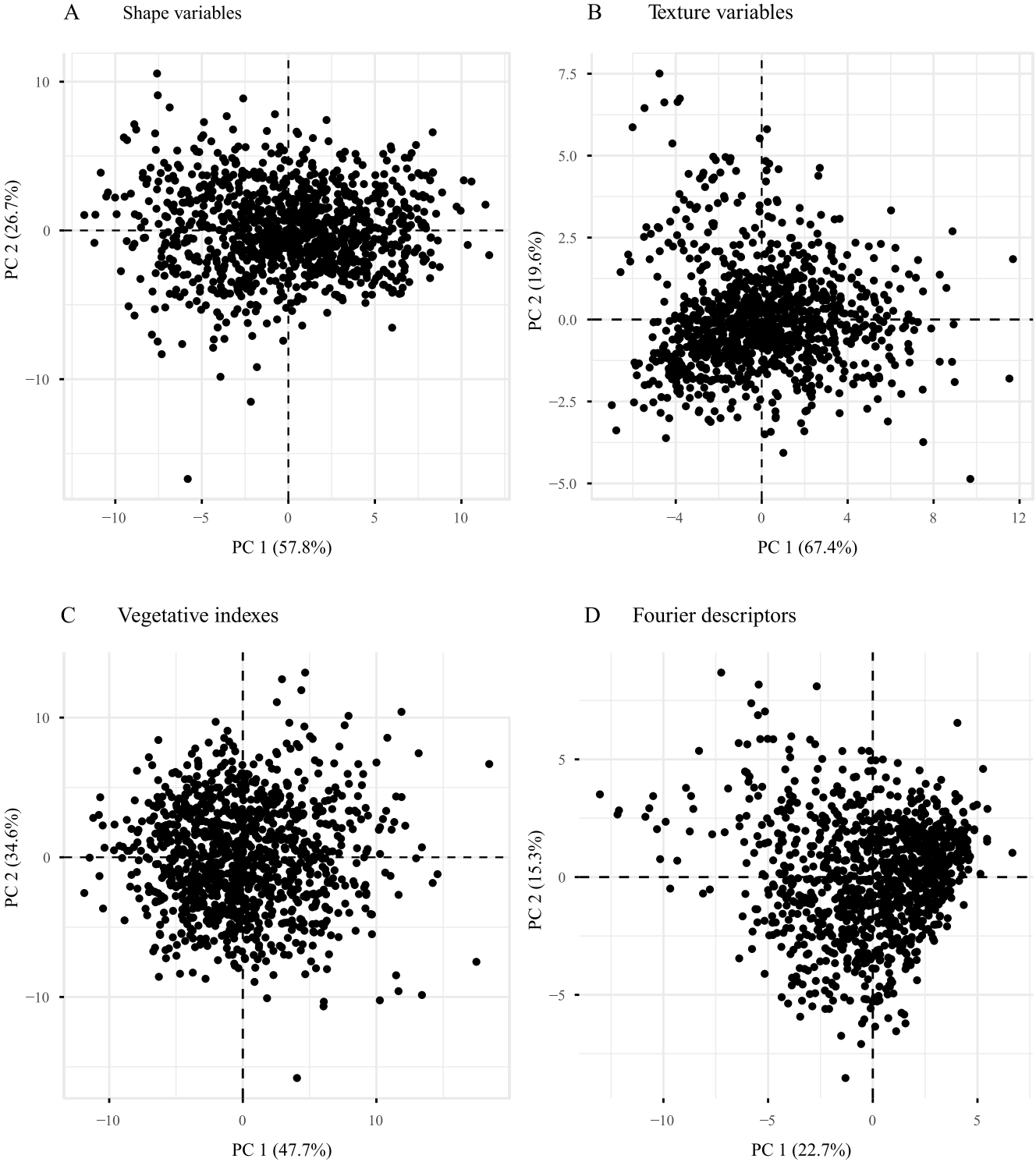

Figure 2. Scatter plot of scores from the first two principal components, obtained from different phenomic information sets of shape variables (A), texture variables (B), vegetative indexes (C), and Fourier descriptors (D) extracted from leaflets digital images of soybean lines (*Glycine max*).

graphical similarity, followed by texture and vegetative indices, showing significant concordance between their respective similarities and dissimilarities.

Alternatively, the concordance among different information sets regarding genotypic dissimilarity can be assessed without structurally simplifying the data. Thus, dissimilarity matrices were generated for each phenomic dataset using Euclidean distances, and their pairwise correlations were tested with Mantel tests. Consistent with the Procrustes analysis results, except for correlation magnitude order, the highest agreement pairs were texture versus vegetative indices, and shape versus vegetative indices. Notably, although significant correlations were shown for all pairs of distance matrices, these significant results are influenced by the high degrees of freedom, as the dataset includes information from 1,105 soybean lines, resulting in 609,960 dissimilarity measures.

The agreement between Procrustes and Mantel tests, for the ranking vegetation indices as the set most concordant with shape and texture, indicates they occupy a central position in the phenomic space, confirming reports that spectral information alone separates soybean genotypes (Silva Junior et al., 2018). The caveat is the mirror image of this centrality. Because vegetation indices overlapped strongly with the other sets, their significant Mantel correlations added little independent discrimination once shape and texture were included, which the inflated degrees of freedom (609,960 measures) further exaggerated.

A graphical inference can also be made, using dissimilarity measures regarding the degree of concordance between pairs of phenomic information sets. In this context, the less elliptical distribution of points predominant quadrants 1 and 3 (Figure 4), in the counterclockwise order, indicates a lower-information overlap among the generated datasets.

Despite their redundancy and genotypic discrimination capabilities, the phenomic datasets also show great potential for genotype pattern recognition

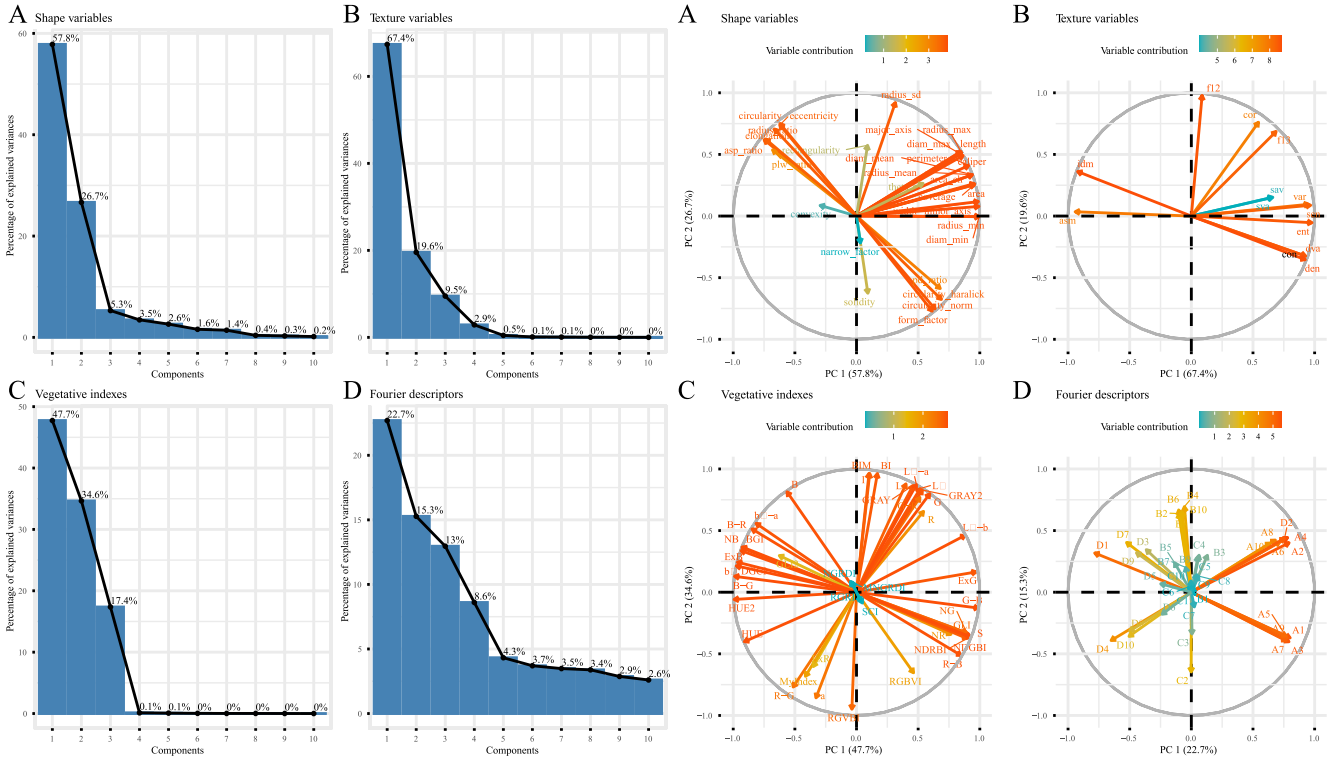

Figure 3. Percentage of variance explained by the first 10 principal components derived from different phenomic information sets of shape (A), texture (B), vegetative indexes (C), and Fourier descriptors (D); and correlation, representation quality, and contribution of attributes to the first two principal components, obtained through principal component analysis, and extracted from digital images of soybean leaflets (*Glycine max*).

– a critical approach in plant breeding. The goal is to identify homogeneous groups without predefined target responses (Bishop, 2006). In the present work, the k-means clustering identified five optimal clusters (Figure 5) across all datasets. Despite the coincidence, the number of clusters was mathematically determined by the most significant perpendicular distance from the intermediate (x, y) points to the line connecting the extreme points on the Elbow curve. Cluster visualization (Figure 6) showed well-defined groupings for shape and texture, while Fourier and vegetative indices exhibited more significant overlaps.

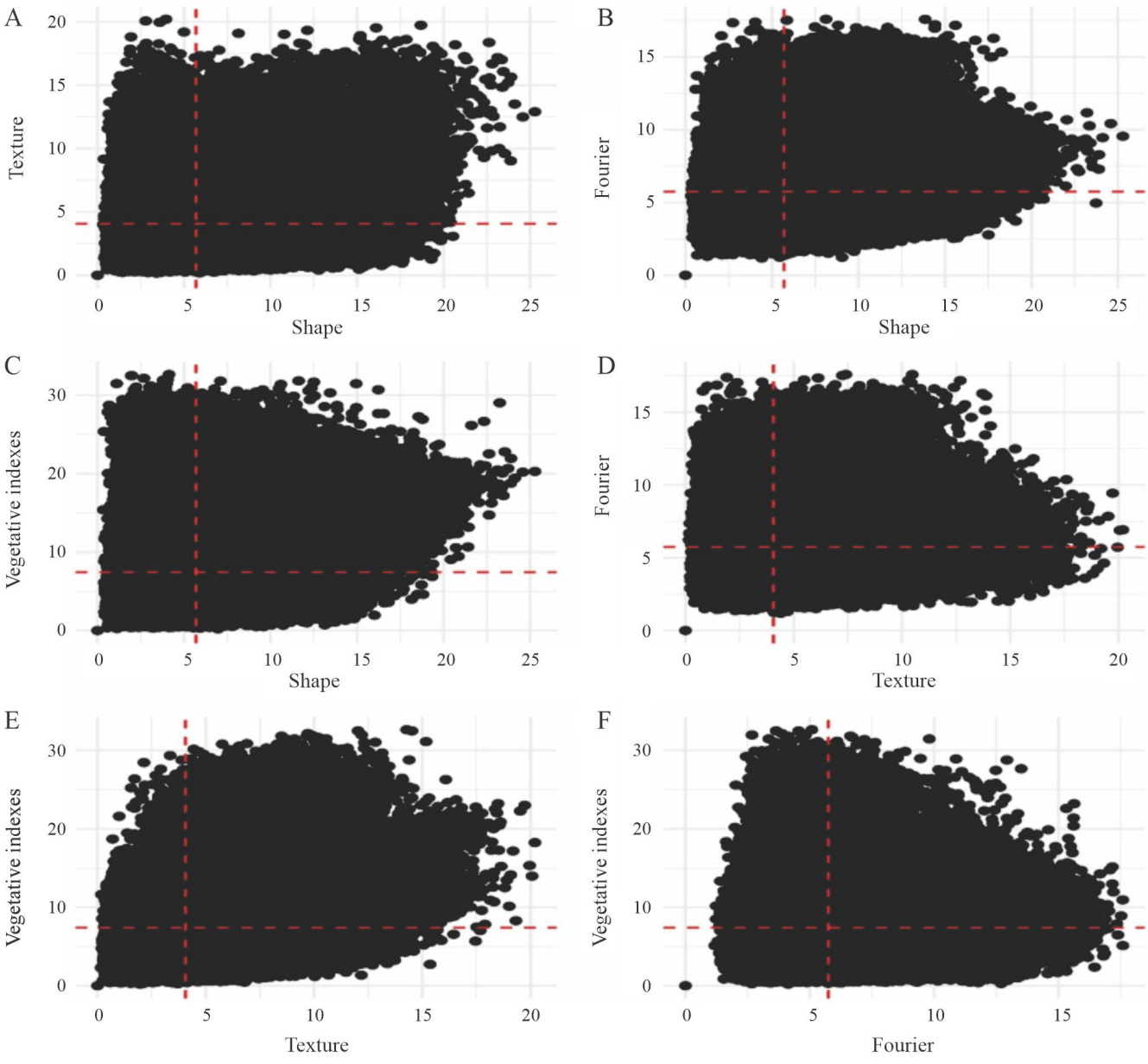

Figure 4. Relationship between dissimilarity distances for pairs of attribute sets (shape, texture, Fourier, and vegetative indexes), extracted from leaflets digital images of soybean lines (*Glycine max*). Red dashed horizontal and vertical lines indicate the medians of dissimilarity distances for each phenomic information set.

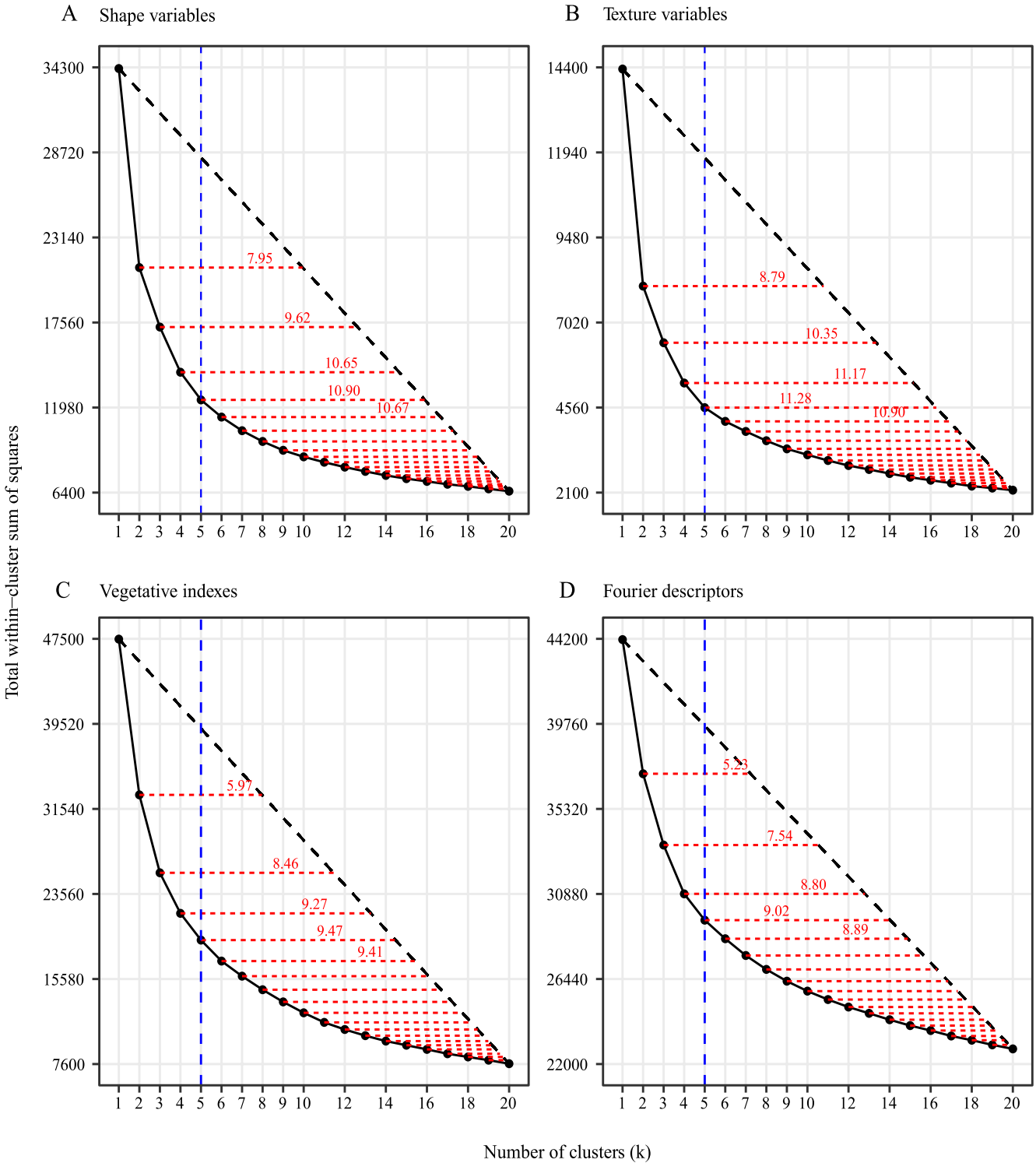

Figure 5. The elbow method determines the number of clusters used in the k-means algorithm from different phenomic information sets of shape variables (A), texture variables (B), vegetative indexes (C), and Fourier descriptors (D), extracted from leaflets digital images of soybean lines (*Glycine max*). Red dashed lines represent the perpendicular distances from each intermediate point (x, y) to the black dashed line connecting the first (x_1, y_1) and last (x_{20}, y_{20}) points on the curve. The vertical blue dashed line indicates the clusters associated with the most considerable perpendicular distance. Red dashed lines are visually represented horizontally due to the compression of the y-axis relative to the x-axis in the graphs.

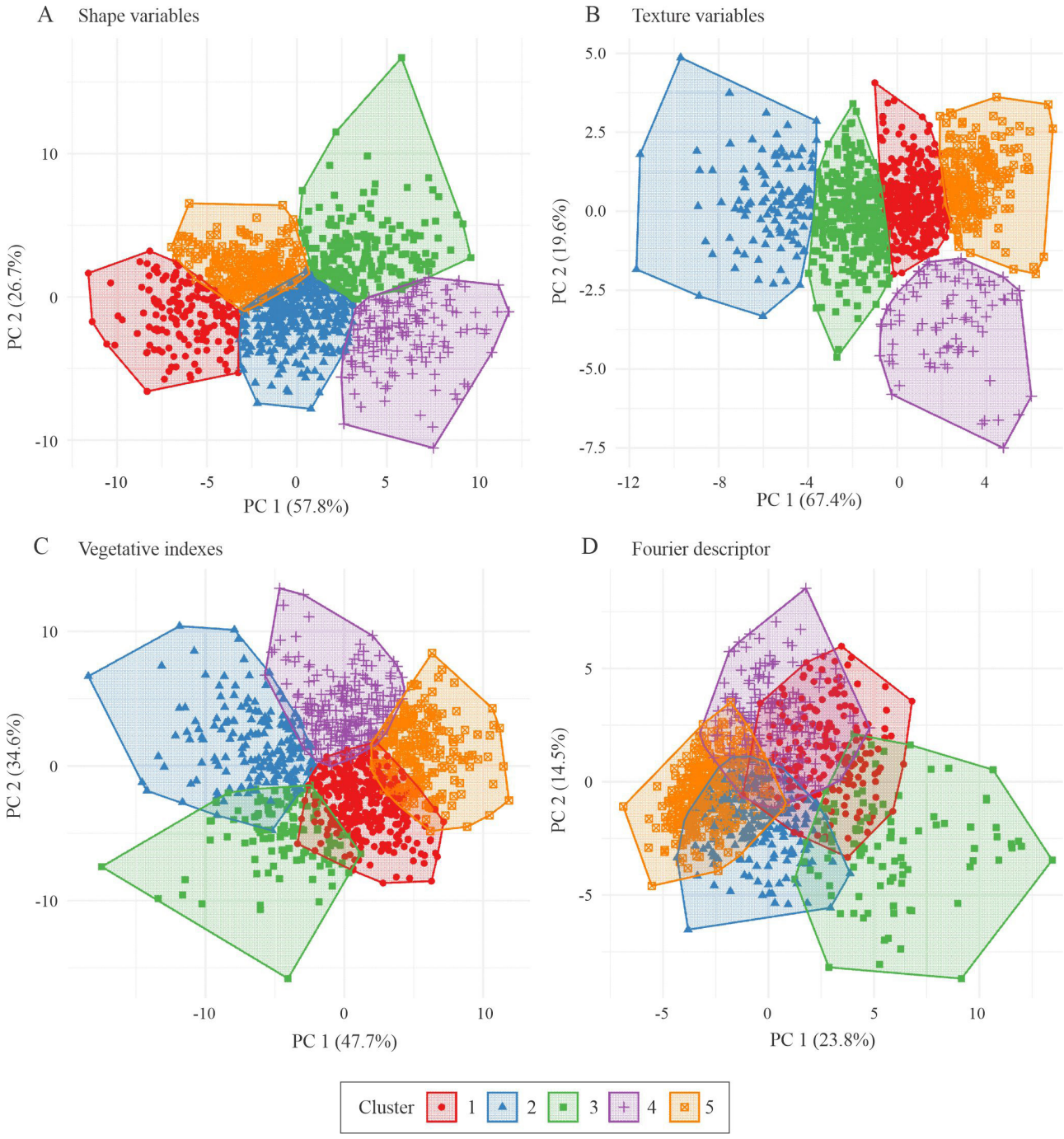

Figure 6. Clusters formed by the k-means algorithm applied to each attribute set of shape variables (A), texture variables (B), vegetative indexes (C), and Fourier descriptors (D), extracted from leaflets digital images of soybean lines (*Glycine max*).

Conclusions

1. Among the four datasets, the elliptical Fourier descriptors show the lowest information redundancy, according to the Pearson correlation criteria.

2. The four datasets of convey complementary information about the breeding lines, with each set contributing nonredundant descriptors to their characterization.

3. Five phenotypic groups of soybean breeding lines were consistently recovered, with concordance above 50% across the datasets.

4. These findings indicate that image-based phenomic analysis can support the characterization of soybean breeding lines, helping to recognize phenotypic patterns, and to identify the most informative attributes for prediction and classification in breeding programs.

References

ALABI, T.R.; ABEBE, A.T.; CHIGEZA, G.; FOWOBAJE, K.R. Estimation of soybean grain yield from multispectral high-resolution UAV data with machine learning models in West Africa. **Remote Sensing Applications: Society and Environment**, v.27, art. 100782, 2022. DOI: https://doi.org/10.1016/j.rsase.2022.100782.

BISHOP, C.M. **Pattern recognition and machine learning**. New York: Springer, 2006.

CHAWADE, A.; van HAM, J.; BLOMQUIST, H.; BAGGE, O.; ALEXANDERSSON, E.; ORTIZ, R. High-throughput field-phenotyping tools for plant breeding and precision agriculture. **Agronomy**, v.9, art.258, 2019. DOI: https://doi.org/10.3390/agronomy9050258.

CORTES, D.F.M.; CATARINA, R.S.; BARROS, G.B. de A.; ARÊDES, F.A.S.; SILVEIRA, S.F. da; FERREGUETTI, G.A.; RAMOS, H.C.C.; VIANA, A.P.; PEREIRA, M.G. Model-assisted phenotyping by digital images in papaya breeding program. **Scientia Agricola**, v.74, p.294-302, 2017. DOI: https://doi.org/10.1590/1678-992X-2016-0134.

DIPTA, B.; SOOD, S.; DEVI, R.; BHARDWAJ, V.; MANGAL, V.; THAKUR, A.K.; KUMAR, V.; PANDEY, N.K.; RATHORE, A.; SINGH, A.K. Digitalization of potato breeding program: improving data collection and management. **Heliyon**, v.9, e12974, 2023. DOI: https://doi.org/10.1016/j.heliyon.2023.e12974.

DUDDU, H.S.N.; JOHNSON, E.N.; WILLENBORG, C.J.; SHIRTLIFFE, S.J. High-throughput UAV image-based method is more precise than manual rating of herbicide tolerance. **Plant Phenomics**, v.2019, art.6036453, 2019. DOI: https://doi.org/10.34133/2019/6036453.

GE, H.; LV, G.; QIN, Y.; SHEN, M. Assessing leaf nitrogen concentration in rice using RGB imaging: a comparative study at leaf, canopy, and plot scales. **Frontiers in Plant Science**, v.16, art.1599177, 2025. DOI: https://doi.org/10.3389/fpls.2025.1599177.

GILL, T.; GILL, S.K.; SAINI, D.K.; CHOPRA, Y.; DE KOFF, J.P.; SANDHU, K.S. A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. **Phenomics**, v.2, p.156-183, 2022. DOI: https://doi.org/10.1007/s43657-022-00048-z.

HAINES, A.J.; CRAMPTON, J.S. Improvements to the method of Fourier shape analysis as applied in morphometric studies. **Palaeontology**, v.43, p.765-783, 2000. DOI: https://doi.org/10.1111/1475-4983.00148.

HARALICK, R.M.; SHANMUGAM, K.; DINSTEIN, I. Textural features for image classification. **IEEE Transactions on Systems, Man, and Cybernetics**, v.SMC-3, p.610-621, 1973. DOI: https://doi.org/10.1109/TSMC.1973.4309314.

HE, H.; MA, X.; GUAN, H. A calculation method of phenotypic traits of soybean pods based on image processing technology. **Ecological Informatics**, v.69, art.101676, 2022. DOI: https://doi.org/10.1016/j.ecoinf.2022.101676.

IQBAL, N.; MUMTAZ, R.; SHAFI, U.; ZAIDI, S.M.H. Gray level co-occurrence matrix (GLCM) texture based crop classification using low altitude remote sensing platforms. **PeerJ Computer Science**, v.7,e536, 2021. DOI: https://doi.org/10.7717/peerj-cs.536.

JANGRA, S.; CHAUDHARY, V.; YADAV, R.C.; YADAV, N.R. High-throughput phenotyping: a platform to accelerate crop improvement. **Phenomics**, v.1, p.31-53, 2021. DOI: https://doi.org/10.1007/s43657-020-00007-6.

KASSAMBARA, A. **Practical guide to principal component methods in R**: PCA, M (CA), FAMD, MFA, HCPC, Factoextra. [S.l.]: Sthda, 2017. v.2.

KHAN, T.; JAMIL, M.; ALI, A.; RASHEED, S.; IRSHAD, A.; MAQSOOD, M.F.; ZULFIQAR, U.; CHAUDHARY, T.; ALI, M.A.; ELSHIKH, M.S. Exploring water-absorbing capacity: a digital image analysis of seeds from 120 wheat varieties. **Scientific Reports**, v.14, art.6757, 2024. DOI: https://doi.org/10.1038/s41598-024-57193-w.

KUHL, F.P.; GIARDINA, C.R. Elliptic Fourier features of a closed contour. **Computer Graphics and Image Processing**, v.18, p.236-258, 1982. DOI: https://doi.org/10.1016/0146-664X(82)90034-X.

MOMIN, M.A.; YAMAMOTO, K.; MIYAMOTO, M.; KONDO, N.; GRIFT, T. Machine vision based soybean quality evaluation. **Computers and Electronics in Agriculture**, v.140, p.452-460, 2017. DOI: https://doi.org/10.1016/j.compag.2017.06.023.

NGUYEN, H.T.; KHAN, M.A.R.; NGUYEN, T.T.; PHAM, N.T.; NGUYEN, T.T.B.; ANIK, T.R.; NGUYEN, M.D.; LI, M.; NGUYEN, K.H.; GHOSH, U.K.; TRAN, L.-S.P.; HA, C.V. Advancing crop resilience through high-throughput phenotyping for crop improvement in the face of climate change. **Plants**, v.14, art.907, 2025. DOI: https://doi.org/10.3390/plants14060907.

OLIVOTO, T. Lights, camera, pliman! An R package for plant image analysis. 2022. Methods in Ecology and Evolution, v.13, p.789-798, 2022. DOI: https://doi.org/10.1111/2041-210X.13803.

OMARI, M.K.; LEE, J.; FAQEERZADA, M.A.; JOSHI, R.; PARK, E.; CHO, B.-K. Digital image-based plant phenotyping: a review. **Korean Journal of Agricultural Science**, v.47, p.119-130, 2020. DOI: https://doi.org/10.7744/kjoas.20200004.

REN, T.; WERADUWAGE, S.M.; SHARKEY, T.D. Prospects for enhancing leaf photosynthetic capacity by manipulating mesophyll cell morphology. **Journal of Experimental Botany**, v.70, p.1153-1165, 2019. DOI: https://doi.org/10.1093/jxb/ery448.

ROWLAND, S.D.; ZUMSTEIN, K.; NAKAYAMA, H.; CHENG, Z.; FLORES, A.M.; CHITWOOD, D.H.; MALOOF, J.N.; SINHA, N.R. Leaf shape is a predictor of fruit quality and cultivar performance in tomato. **New Phytologist**, v.226, p.851-865, 2020. DOI: https://doi.org/10.1111/nph.16403.

SHILPASHREE, N.; DEVI, S.N.; MANJUNATHAGOWDA, D.C.; MUDDAPPA, A.; ABDELMOHSEN, S.A.M.; TAMAM, N.; ELANSARY, H.O.; ZIN EL-ABEDIN, T.K.; ABDELBACKI, A.M.M.; JANHAVI, V. Morphological characterization, variability and diversity among vegetable soybean (*Glycine max* L.) genotypes. **Plants**, v.10, art.671, 2021. DOI: https://doi.org/10.3390/plants10040671.

SILVA JUNIOR, C.A. da; NANNI, M.R.; SHAKIR, M.; TEODORO, P.E.; OLIVEIRA-JÚNIOR, J.F. de; CESAR, E.; GOIS, G. de; LIMA, M.; WOJCIECHOWSKI, J.C.; SHIRATSUCHI, L. Soybean varieties discrimination using non-imaging hyperspectral sensor. **Infrared Physics & Technology**, v.89, p.338-350, 2018. DOI: https://doi.org/10.1016/j.infrared.2018.01.027.

VISAKH, R.L.; ANAND, S.; REDDY, S.B.; JHA, U.C.; SAH, R.P.; BEENA, R. Precision phenotyping in crop science: from plant traits to gene discovery for climate-smart agriculture. **Plant Breeding**, n.0, p.1-29, 2024. DOI: https://doi.org/10.1111/pbr.13228.

XING, Y.; LV, P.; HE, H.; LENG, J.; YU, H.; FENG, X. Traits expansion and storage of soybean phenotypic data in computer vision-based test. **Frontiers in Plant Science**, v.13, art. 832592, 2022. DOI: https://doi.org/10.3389/fpls.2022.832592.

ZHANG, Q.; LUAN, R.; WANG, M.; ZHANG, J.; YU, F.; PING, Y.; QIU, L. Research progress of spectral imaging techniques in plant phenotype studies. **Plants**, v.13, art.3088, 2024. DOI: https://doi.org/10.3390/plants13213088.

Author contributions

Wagner Faria Barbosa: conceptualization, methodology, writing – original draft; **Cosme Damião Cruz**: conceptualization, supervision; **Ivan Ricardo Carvalho**: conceptualization, methodology, investigation, writing – original draft; **Jaqueline Piesanti Sangiovo**: writing – review & editing, investigation; **Willyan Júnior Adorian Bandeira**: methodology, investigation, writing – review & editing. Marco **Luís Conde Machado de Souza**: validation, investigation, data curation.

Chief editor: Edemar Corazza
Edited by: Madalena Rinaldi

Data availability statement

Data available upon request: research data are only available upon reasonable request to the corresponding author.

Declaration of use of AI technologies

No generative artificial intelligence (AI) was used in this study.

Conflict of interest statement

The authors declare no conflicts of interest.

Disclaimer/Publisher's note

The statements, opinions, and data contained in all texts published in Pesquisa Agropecuária Brasileira (PAB) are solely those of the individual author(s) and not of the journal's publisher, editor, and editorial team, who disclaim responsibility for any injury to people or property resulting from any referred ideas, methods, instructions, or products.

The mention of specific chemical products, machines, and commercial equipment in the texts published in this journal does not imply their recommendation by the publisher.