Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Digital image-based characterization of lint color and uniformity in naturally colored cotton

Gabriel Aragão Fernandes
gabrielaragaoagro@gmail.com
Universidade Federal de Uberlândia
Filipe Inácio Matias
Universidade Estadual Paulista (Unesp)
Isabela Lorrane Abreu Dias
Universidade Federal de Uberlândia
Ana Luiza Gonçalves Costa
Universidade Federal de Uberlândia
Marcello Matheus Corrêa
Universidade Federal de Uberlândia
Maria Luísa Vieira Borges
Universidade Federal de Uberlândia
Larissa Barbosa de Sousa
Universidade Federal de Uberlândia

Research Article

Keywords: Gossypium hirsutum, CIELAB, chroma, image-based phenotyping, color uniformity, cross-validation

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11721581/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Digital image analysis can make lint color characterization more objective, but its effectiveness depends on the trait assessed and the ability to distinguish similar shades. This study compared digital descriptors for characterizing lint color, predicting visual shade and color uniformity classes, and quantifying variation between growing seasons. Five genotypes were evaluated in a randomized complete block design (RCBD) with four replicates over two growing seasons (2023/2024 and 2024/2025). Images of 15 bolls per plot were acquired in a lightbox, calibrated using a ColorChecker, and processed to extract RGB, HSV, and CIELAB descriptors. Two raters assessed shade and color uniformity, and Random Forest models were evaluated using leave-one-plot-out cross-validation.

Results

Genotype color did not change significantly between growing seasons, whereas interannual color differences within genotypes ranged from ΔE00 = 2.64 to 7.96. The L*, a*, and b* coordinates characterized the genotypes without information loss. Principal component analysis yielded loadings of − 0.588 for L*, + 0.574 for a*, and + 0.569 for b*, defining a CIELAB pigmentation gradient for brown cotton lint. Interrater agreement was 72.5% for both shade and color uniformity, indicating differences in visual perception between raters. The highest accuracies of digital predictions against visual ratings were 70.0% and 80.0% for shade and 75.0% and 55.0% for uniformity for raters A1 and A2, respectively.

Conclusions

Calibrated CIELAB coordinates (L*, a*, and b*) provide an objective basis for lint shade screening in naturally colored cotton, whereas reliable prediction of color uniformity requires spatial dispersion metrics beyond plot-level means.

1 Introduction

Naturally colored cotton (*Gossypium hirsutum* L.) accumulates pigments in the fiber lumen and secondary cell wall. These pigments consist mainly of proanthocyanidins and condensed tannins derived from the flavonoid pathway in brown cultivars, and suberin-associated polymers and caffeic acid derivatives in green cultivars (Naoumkina et al. 2024).

Fiber pigmentation reduces water and energy demand and chemical effluent discharge during cotton dyeing (Naoumkina et al. 2024). Despite these environmental benefits, commercial expansion of naturally colored cultivars is constrained by lower color stability under varying soil and climatic conditions and by subjective visual bale grading. Grading requires simultaneous assessment of the

predominant shade, discrimination of similar hues, and quantification of spatial heterogeneity within samples.

In fiber processing and plant breeding, digital phenotyping using RGB images has emerged as a low-cost, nondestructive alternative to High Volume Instrument (HVI) systems and benchtop spectrophotometers (Gonçalves et al. 2015; Fernandes et al. 2026).

Fernandes et al. (2026) demonstrated that empirical spectral indices derived from RGB images, including the brightness index (BI), saturation spectral slope index (SI), and normalized green–red difference index (NGRDI), can quantify color intensity and its variation with boll position across canopy thirds.

However, empirical indices based on RGB band ratios depend on sensor radiometric characteristics and lack direct perceptual uniformity. A psychophysical approach standardized by the Commission Internationale de l'Éclairage (CIE) transforms these signals into the CIELAB color space, which separates color into lightness (L*, ranging from 0 for black to 100 for perfect white) and the green–red (a*) and blue–yellow (b*) opponent coordinates, from which chroma (C*ab) and hue angle (h°ab) are derived (Xu et al. 1998; Matusiak 2015; Kulsum et al. 2024). In textile research, CIELAB coordinates and machine learning models have been widely applied to predict yarn and fabric color changes after dyeing and chemical finishing (Balcı and Tuğrul Oğulata 2009; Şahin et al. 2023; Chen et al. 2024), confirming the value of CIELAB as a reproducible colorimetric standard.

Distances between CIELAB profiles can also be summarized by total color difference metrics (ΔE76 and CIEDE2000, ΔE00), which quantify perceptual differences among genotypes, lots, or growing seasons (Sharma et al. 2005; Vik et al. 2017). RGB, HSV, Hunter Lab, and CIELAB representations have been used for multivariate cultivar discrimination (Gonçalves et al. 2015) and automated prediction of commercial lint and seed cotton grades using machine learning (Fisher et al. 2023; Li et al. 2025). However, the correspondence between digital descriptors and human visual ratings depends on interrater consistency and the trait assessed.

For color uniformity, previous studies show that plot-level mean reflectance or color coordinates can mask relevant within-sample variation. Cui et al. (2014) and Vik et al. (2017) demonstrated spatial variation within cotton samples assigned similar overall grades, while Heng et al. (2020) and Li et al. (2025) showed that shadows, stains, and impurities affect pixel-level color dispersion.

Unlike previous textile finishing studies that modeled CIELAB shifts in processed yarns and fabrics (Balcı and Tuğrul Oğulata 2009; Şahin et al. 2023; Chen et al. 2024) or cotton imaging studies focused on cultivar identity or white-cotton grading (Gonçalves et al. 2015; Fisher et al. 2023), the novelty of this work lies in integrating radiometrically calibrated lightbox imaging, REML variance decomposition across two growing seasons, and leave-one-plot-out cross-validation (LOOCV) of Random Forest models against independent visual ratings of both lint shade and color uniformity in naturally colored cotton, while explicitly testing whether derived metrics (C*ab and ΔE) add predictive value or introduce redundancy.

2 Materials and methods

2.1 Experimental site, soil and climatic conditions, and design

The experiments were conducted over two growing seasons (2023/2024 and 2024/2025) at the Glória Experimental Farm, Universidade Federal de Uberlândia (UFU), Glória Campus, Uberlândia, Minas Gerais, Brazil (18°56'52.77" S, 48°12'51.97" W; mean elevation, 865 m).

The first experiment ran from January to July 2024, with sowing on January 5 and harvest on July 30, 2024. The second ran from December 2024 to August 2025, with sowing on December 17, 2024, and harvest on August 30, 2025. The regional climate is classified as Aw (tropical savanna) under the Köppen–Geiger system, with a rainy season from October to April and a dry season from May to September. Mean air temperature was approximately 22°C in both growing seasons, while cumulative rainfall was approximately 680 mm in 2023/2024 and 930 mm in 2024/2025 (Fernandes et al. 2026). Rainfall near harvest in 2024/2025 prolonged the time that mature cotton remained in the field.

We used a randomized complete block design (RCBD) with four replicates per growing season, totaling 40 experimental plots across both seasons. Five cotton genotypes were evaluated: four commercial naturally colored cultivars developed by Embrapa Algodão—BRS Jade (light-brown lint), BRS Topázio (light-brown/cream lint), BRS Rubi (dark reddish-brown lint), and BRS Verde (greenish lint) (Carvalho et al. 2011)—and the commercial cultivar Suregrow 821 (SG 821), used as a white-lint control. Each plot comprised four 5.0-m rows spaced 0.9 m apart (18.0 m2). The sampling area consisted of the two central rows, excluding 0.5 m at each end as borders.

2.2 Crop management

Before establishing the experiments, we characterized soil chemical properties and particle-size distribution at depths of 0–20 and 20–40 cm (Table S1). Dolomitic lime was applied at 2.2 t·ha − 1 to correct soil acidity and raise base saturation to 70%. At sowing, we applied 20 kg·ha − 1 nitrogen (N) and 60 kg·ha − 1 P_2O_5. Topdressing supplied 60 kg·ha − 1 N and 30 kg·ha − 1 K_2O in two equal applications at 25 and 45 days after emergence (DAE). Foliar boron (B) was supplied at 1 kg·ha − 1 in four applications starting at 45 DAE, at 20-day intervals.

Plant architecture was managed using chlormequat chloride (2-chloroethyltrimethylammonium chloride), starting at 45 DAE. Subsequent applications depended on canopy height growth, with a target final plant height of approximately 1.30 m to improve aeration and facilitate harvest.

2.3 Image acquisition, digital processing, and visual assessment

At full maturity, we collected 15 representative bolls from the sampling area of each plot, with five bolls from each canopy third (lower, middle, and upper) (Fernandes et al. 2026). Seed cotton locks were

arranged in groups representing five bolls per image, yielding three images per plot, inside a controlled-light chamber (30 × 50 cm lightbox). The chamber had a matte blue background and two neutral-white LED lamps (9 W, approximately 800 lm each), positioned symmetrically on the sides behind tracing-paper diffusers. Images were captured at a fixed vertical distance of 50 cm using the 2× telephoto camera of an iPhone 14 Pro Max smartphone (12 MP, 48-mm equivalent focal length, f/1.78 aperture), with manual exposure and white balance locked.

A Pixel Perfect Camera Color Correction Card (5 × 7 inches) was placed at the base of the sample in each image for radiometric calibration in ImageJ (Schneider et al. 2012). The calibrated images were then processed in R using FIELDimageR (Matias et al. 2020). The blue background was segmented and removed using a saturation index threshold (SI = 0.0), followed by individual visual inspection. Using only fiber pixels, we extracted plot-level means for the R, G, and B channels; the HSV components Hue (0–1), Sat (saturation, 0–1), and Val (brightness recorded on a normalized scale); and the CIELAB coordinates L*, a*, and b*.

During image acquisition, two previously trained raters (A1 and A2) independently inspected the same 40 plot samples under standardized lighting. Visual shade ratings (Ton) followed a nominal scale with five classes: 1 = white, 2 = cream, 3 = brown, 4 = reddish-brown, and 5 = green. Sample color uniformity (Unif) followed a three-class ordinal scale: 1 = uniform, 2 = moderately uniform, and 3 = nonuniform. We modeled each rater's assessments separately (Ton A1, Ton A2, Unif A1, and Unif A2) to preserve interrater variation.

2.4 Statistical analysis and predictive modeling

We fitted a univariate linear mixed model for each of ten color descriptors: R, G, B, Hue, Sat, Val, L*, a*, b*, and chroma (C*ab). Chroma and the complementary hue angle (h°ab) were calculated from the Cartesian coordinates a* and b* according to the CIE standard (CIE 2018):

$$C*ab = \sqrt{\left[(a*)^2 + (b*)^2\right]}$$

Analyses used restricted maximum likelihood (REML) in the lme4 package (Bates et al. 2015) in R. Growing season was treated as a fixed effect, and genotype, genotype × growing season interaction, and block nested within growing season were treated as random effects, assuming independent intercepts and Gaussian errors:

$$Y_{ijk} = \mu + S_i + G_j + GS_{ij} + B_{k(i)} + \varepsilon_{ijk}$$

where Yijk is the observed response for the plot of genotype j in block k within growing season i; μ is the overall mean; Si is the fixed effect of growing season i; Gj is the random effect of genotype j; GSij is the random interaction effect between genotype j and growing season i; Bk(i) is the random effect of block k nested within growing season i; and εijk is the random residual error associated with the plot.

Response variables were standardized before model fitting. Variance component estimates and empirical best linear unbiased predictions (BLUPs) were then converted back to their original scales. The mean growing season effect was evaluated using an F-test with Satterthwaite's approximation for degrees of freedom in lmerTest (Kuznetsova et al. 2017). We applied the Benjamini–Hochberg false discovery rate correction across the ten parameters at a 5% significance level. Adjusted marginal means for each growing season were obtained using emmeans. Residual normality was assessed using the Shapiro–Wilk test and quantile–quantile plots (Supplementary material).

We explored multivariate genotypic variation using principal component analysis (PCA) of predicted genotypic values for the primary coordinates *L**, *a**, and *b**, centered and scaled to unit variance. RGB channels, HSV components, and C^*_{ab} were excluded from the primary PCA to avoid structural collinearity among color systems derived from the same spectral bands.

Color dissimilarity among genotypes was also quantified using Euclidean distances between standardized genotypic values of *L**, *a**, and *b**. These distances were subjected to Ward's agglomerative hierarchical clustering to describe the relationships among color phenotypes.

To quantify the perceptual magnitude of color contrasts between genotype pairs and between growing seasons within each genotype, we calculated the classical Euclidean distance (ΔE76) and CIEDE2000 color difference (ΔE00) using farver. The latter incorporates parametric weighting factors for lightness, chroma, and hue, as well as a rotation term for the blue region (Sharma et al. 2005).

To describe the direction of interannual color shifts, we calculated differences in observed genotype means for *L**, *a**, *b**, and C^*_{ab} between the second and first growing seasons:

$$\Delta\, xg \;=\; \mu\;(g_i) - \;\mu\;(g_{i'})$$

where μgi and μgi' are the observed means of color parameter x for genotype g in growing seasons 2024/2025 (i) and 2023/2024 (i'), respectively.

Interrater agreement (A1 vs. A2) for visual ratings was assessed using exact percentage agreement and Cohen's kappa (*κ*), separately for shade and uniformity.

We then fitted Random Forest classifiers using ranger (Wright and Ziegler 2017) to predict each of the four visual targets (Ton A1, Ton A2, Unif A1, and Unif A2) from nine candidate sets of plot-level observed predictors: (i) L* alone; (ii) a* alone; (iii) b* alone; (iv) C*ab alone; (v) Lab (L*, a*, and b* combined); (vi) Lab + C*ab; (vii) ΔE (ΔE76 and ΔE00 combined); (viii) Lab + ΔE; and (ix) All + ΔE.

Because no external ceramic white standard was available, ΔE_{76} and ΔE_{00} predictors were calculated relative to the CIELAB centroid of only the 39 training plots in each cross-validation fold. The same reference point was then used for the test plot to prevent information leakage.

Predictive performance was estimated using leave-one-plot-out cross-validation (LOOCV) (Roberts et al. 2017). All runs used the same setting of 1 000 decision trees. The 40 out-of-sample predictions were summarized in confusion matrices and evaluated using overall accuracy, balanced accuracy (the arithmetic mean of class sensitivities), macro F1 score, and kappa.

We also compared model performance against a majority-class baseline and extracted mean out-of-bag (OOB) permutation importance from the full model (All + ΔE) to explore the relative contributions of the predictors.

3 Results

3.1 Growing season effect and variance components

The mean growing season effect was not significant for any color parameter (Table 1). The lack of significant differences in marginal means between 2023/2024 and 2024/2025 reflects opposing interannual changes among genotypes (genotype × growing season interaction), which offset one another in the experiment-wide mean.

Table 1
Adjusted marginal means of digital lint color descriptors in the 2023/2024 and 2024/2025 growing seasons and F-test results for the growing season effect.

Parameter	2023/2024	2024/2025
L*	75.28	71.46
a*	1.93	3.25
b*	14.03	14.31
Chroma*	14.39	15.99
R	198.56	189.86
G	183.51	172.35
B	160.22	149.78
Hue	0.151	0.191
Sat	0.161	0.209
Val	0.003 591	0.003 488

Means are expressed on their recorded scales: L* (0–100), a*, b*, and chroma (C*ab) in CIELAB units; R, G, and B on an 8-bit scale (0–255); and Hue and Sat on a [0, 1] scale. Val was recorded on a scale

normalized by 2552; multiply by 255 to obtain brightness V on the [0, 1] scale, equivalent to 0.916 in 2023/2024 and 0.889 in 2024/2025.

REML decomposition of phenotypic variance showed that genotypic variance (σg2) exceeded genotype × growing season interaction variance (σgs2) and residual variance (Table 2). Thus, although genotypes retained their characteristic colors, color magnitude varied to some extent between growing seasons.

Table 2
REML estimates of variance components in squared original units for ten digital lint color descriptors.

Parameter	Genotype	G × growing season	Block/growing season	Residual
L*	63.812 6	5.055 3	0.000 0	1.844 6
a*	29.892 3	5.405 8	0.000 0	1.231 6
b*	66.554 0	10.580 3	1.382 7	6.129 4
C*ab	78.370 4	7.013 7	1.154 4	5.570 6
R	85.650 1	22.962 8	0.000 0	17.325 6
G	632.759 1	48.248 6	0.000 0	14.989 3
B	1 223.569 0	106.898 5	5.109 9	39.327 8
Hue	0.025 3	0.003 1	0.000 0	0.008 6
Sat	0.017 9	0.001 5	0.000 2	0.001 2
Val	4.82e-08	3.59e-09	0.000 0	5.82e-09

L* (0–100), a*, b*, and C*ab are expressed in CIELAB units; R, G, and B on an 8-bit scale (0–255); and Hue and Sat on a [0, 1] scale. Val was recorded on a scale normalized by 2552; multiply by 255 to obtain brightness V on the [0, 1] scale, equivalent to 0.916 in 2023/2024 and 0.889 in 2024/2025.

For L*, a*, b*, and C*ab, genotypic variance estimates (63.81, 29.89, 66.55, and 78.37, respectively) were 5.5–12.6 times the corresponding interaction variances (5.06, 5.41, 10.58, and 7.01).

3.2 Genotypic predictions and CIELAB characterization

Predicted genotypic values separated the cultivars within CIELAB space (Table 3). BRS Rubi was distinct because it combined the lowest lightness with the highest a* and b* values.

Values were obtained by adding the weighted marginal mean across growing seasons to the corresponding genotypic BLUP. Chroma (C^*_{ab}) was fitted in a separate linear mixed model using plot-level C^*_{ab} values; h°_{ab} is the corresponding polar angle in degrees [atan2(b^*, a^*)].

At the opposite extreme, the white-lint control SG 821 had the highest lightness (L^* = 82.20), a slightly negative a^* coordinate (− 1.22), and low residual yellowness (b^* = 2.64; C^*_{ab} = 3.82; $h°_{ab}$ = 114.8°). BRS Rubi (− 12.57 for L^*, + 9.11 for a^*, and + 10.05 for b^*) and SG 821 (+ 8.82 for L^*, − 3.82 for a^*, and − 11.52 for b^*) had 95% prediction intervals that did not cross the zero reference line for any of the three CIELAB coordinates (Fig. 1). In contrast, the intermediate-shade genotypes BRS Jade, BRS Topázio, and BRS Verde had deviations closer to the overall mean, consistent with their narrow lightness range of 72.99–75.62.

The intermediate-shade cultivars illustrate the complementary information provided by CIELAB coordinates. BRS Jade and BRS Topázio had nearly identical lightness (L* = 75.26 and 75.62), but differed in their yellow component (b* = 13.90 and 17.11, respectively) and chroma (C*ab = 14.03 and 17.38, respectively), with greater chroma in BRS Topázio. BRS Verde had L* = 72.99, b* = 12.96, and a slightly negative green–red coordinate (a* = −0.30; h°ab = 91.3°).

Because C*ab is the magnitude of the color vector in the a*–b* plane, it summarizes saturation intensity (e.g., 14.03 for BRS Jade and 12.99 for BRS Verde), whereas joint use of L*, a*, and b* preserves the three-dimensional hue direction distinguishing greenish (a* < 0) from reddish (a* > 0) colors.

3.3 Multivariate distribution and color divergence

In the PCA of genotypic values for L^*, a^*, and b^*, the first component (PC1) explained 92.82% of total variance and the second (PC2) explained 5.41%. Together, they captured 98.23% of the three-dimensional variation in a two-dimensional plane (Fig. 2). PC1 had eigenvector coefficients of similar magnitude for all three coordinates (− 0.588 for L^*, + 0.574 for a^*, and + 0.569 for b^*), representing a latent pigmentation intensity gradient from dark, saturated lint at positive scores (BRS Rubi) to light, achromatic lint at negative scores (SG 821).

PC2 was dominated by the contrast between a^* (+ 0.647) and b^* (− 0.756), with a minor contribution from L^* (− 0.099). This component separated the more yellowish BRS Topázio profile (negative PC2 score) from the subgroup comprising BRS Jade and BRS Verde.

Ward's hierarchical clustering supported the relationships observed in the PCA. BRS Jade and BRS Verde merged at the smallest Euclidean distance, followed by BRS Topázio, forming an intermediate group. SG 821 and, more distinctly, BRS Rubi joined at higher fusion distances (Fig. 3).

3.4 Color differences among genotypes and stability across growing seasons

The matrix of pairwise color differences between predicted genotypic values showed the greatest phenotypic divergence between BRS Rubi and SG 821 (ΔE_{76} = 33.02; ΔE_{00} = 24.10). The smallest perceptual distances occurred between BRS Jade and BRS Topázio (ΔE_{76} = 3.31; ΔE_{00} = 2.04) and between BRS Jade and BRS Verde (ΔE_{76} = 2.79; ΔE_{00} = 2.44) (Table 4).

Table 4
Pairwise colorimetric distances calculated using CIE76 (ΔE_{76}) and CIEDE2000 (ΔE_{00}) from predicted genotypic values of *L**, *a**, and *b**.

Genotype pair	ΔE76	ΔE00
BRS Jade × BRS Rubi	20.73	15.65
BRS Jade × BRS Topázio	3.31	2.04
BRS Jade × BRS Verde	2.79	2.44
BRS Jade × SG 821	13.41	10.10
BRS Rubi × BRS Topázio	19.22	14.92
BRS Rubi × BRS Verde	20.48	15.82
BRS Rubi × SG 821	33.02	24.10
BRS Topázio × BRS Verde	5.33	3.97
BRS Topázio × SG 821	16.17	11.76
BRS Verde × SG 821	13.86	10.23

The two metrics ranked the closest pairs differently: simple Euclidean distance (ΔE76) indicated greater similarity between BRS Jade and BRS Verde (2.79 vs. 3.31), whereas CIEDE2000 (ΔE00) indicated greater perceptual similarity between BRS Jade and BRS Topázio (2.04 vs. 2.44).

Between the 2023/2024 and 2024/2025 growing seasons, BRS Jade showed the smallest overall color shift (ΔE00 = 2.64; ΔE76 = 3.90), followed by BRS Topázio (ΔE00 = 4.89), SG 821 (ΔE00 = 5.39), BRS Verde (ΔE00 = 6.24), and BRS Rubi (ΔE00 = 7.96) (Fig. 4 and Table 5).

Trajectories differed among genotypes (Table 5). From the first to the second growing season, L* increased slightly in BRS Jade (+ 1.81) but decreased by 3.95–6.30 units in the other genotypes. In BRS Rubi, lint darkening (ΔL* = −6.25) accompanied a marked increase in the red component (Δa* = +6.70) and chroma (ΔC* = +1.54). BRS Verde and BRS Topázio showed pronounced increases in the yellow component (Δb* = +6.46 and + 4.56) and chroma (ΔC* = +6.48 and + 4.78), whereas b* decreased in BRS Jade and SG 821 (− 3.34 and − 4.52, respectively).

Table 5
Observed directional differences (Δ = 2024/2025 minus 2023/2024) in L*, a*, b*, and C*ab and overall color distances (ΔE76 and ΔE00) between growing seasons for five cotton genotypes in Uberlândia, Minas Gerais, Brazil.

Genotype	ΔL*	Δa*	Δb*	ΔC*	ΔE76	ΔE00
BRS Jade	1.81	-0.89	-3.34	-3.29	3.90	2.64
BRS Rubi	-6.25	6.70	-1.76	1.54	9.34	7.96
BRS Topázio	-4.41	2.53	4.56	4.78	6.83	4.89
BRS Verde	-6.30	-0.20	6.46	6.48	9.03	6.24
SG 821	-3.95	-1.52	-4.52	-1.47	6.19	5.39

3.5 Agreement between visual ratings and supervised prediction of shade and uniformity

The two trained raters (A1 vs. A2) agreed exactly on 72.5% of both shade and color uniformity ratings (Table 6). Cohen's kappa was 0.629 for shade (substantial agreement) and 0.548 for uniformity (moderate agreement).

Table 6
Interrater agreement (Rater 1 vs. Rater 2) for visual ratings of lint shade and color uniformity (n = 40 plots).

Trait	Agreement /%	Unweighted kappa
Shade	72.5	0.629
Uniformity	72.5	0.548

On the ordinal uniformity scale (1 = uniform, 2 = moderately uniform, and 3 = nonuniform), most disagreements between A1 and A2 involved adjacent classes (1 and 2).

Random Forest performance under LOOCV varied with the trait and rater (Table 7). For Rater 1 shade ratings (Ton A1), a* alone and the multivariate sets Lab + ΔE and All + ΔE achieved the highest overall accuracy (70.0%; 28/40 correct predictions), with balanced accuracy of 67.0%, macro F1 of 0.658, and κ of 0.579–0.581. Lab and Lab + C* followed closely (65.0% accuracy; κ = 0.511), whereas b* alone (45.0%), C* alone (52.5%), and ΔE alone (55.0%) performed less well. For Rater 2 shade ratings (Ton A2), Lab and all its extensions (Lab + C*, Lab + ΔE, and All + ΔE) achieved 80.0% overall accuracy (32/40 correct predictions), balanced accuracy of 72.6%, and kappa of 0.725–0.728, outperforming L* alone (77.5%; BA = 66.0%) and a* alone (62.5%; BA = 50.0%).

Table 7
Predictive performance of Random Forest models under LOOCV (n = 40 plots) using nine digital predictor sets to classify visual lint shade and color uniformity.

Target	Predictors	Accuracy /%	BA /%	F1 macro	Kappa
Ton A1	L*	57.5	53.8	0.531	0.402
Ton A1	a*	70.0	67.0	0.658	0.581
Ton A1	b*	45.0	38.0	0.375	0.230
Ton A1	C*	52.5	46.0	0.455	0.335
Ton A1	Lab	65.0	61.0	0.607	0.511
Ton A1	Lab + C*	65.0	61.0	0.607	0.511
Ton A1	ΔE	55.0	49.9	0.485	0.339
Ton A1	Lab + ΔE	70.0	67.0	0.658	0.581
Ton A1	All + ΔE	70.0	67.0	0.658	0.579
Ton A2	L*	77.5	66.0	0.643	0.696
Ton A2	a*	62.5	50.0	0.508	0.485
Ton A2	b*	57.5	40.7	0.403	0.415
Ton A2	C*	57.5	42.1	0.419	0.421
Ton A2	Lab	80.0	72.6	0.734	0.728
Ton A2	Lab + C*	80.0	72.6	0.734	0.728
Ton A2	ΔE	65.0	46.2	0.449	0.503
Ton A2	Lab + ΔE	80.0	72.6	0.750	0.725
Ton A2	All + ΔE	80.0	72.6	0.750	0.725
Unif. A1	L*	65.0	59.0	0.602	0.356
Unif. A1	a*	67.5	73.0	0.670	0.461
Unif. A1	b*	45.0	51.0	0.491	0.072
Unif. A1	C*	45.0	45.9	0.480	0.047
Unif. A1	Lab	75.0	77.5	0.735	0.565
Unif. A1	Lab + C*	70.0	73.4	0.676	0.483
Unif. A1	ΔE	60.0	56.0	0.598	0.249
Unif. A1	Lab + ΔE	70.0	68.3	0.663	0.468

Target	Predictors	Accuracy /%	BA /%	F1 macro	Kappa
Unif. A1	All + ΔE	70.0	64.2	0.642	0.469
Unif. A2	L*	42.5	40.3	0.404	0.091
Unif. A2	a*	55.0	54.7	0.540	0.288
Unif. A2	b*	37.5	39.2	0.399	0.001
Unif. A2	C*	32.5	29.3	0.296	-0.079
Unif. A2	Lab	52.5	48.9	0.475	0.235
Unif. A2	Lab + C*	47.5	43.0	0.425	0.155
Unif. A2	ΔE	47.5	41.4	0.416	0.131
Unif. A2	Lab + ΔE	50.0	48.8	0.450	0.211
Unif. A2	All + ΔE	55.0	49.9	0.498	0.280

Ton: visual lint shade rating; Unif.: visual lint color uniformity rating; A1: Rater 1; A2: Rater 2; BA: balanced accuracy (mean class sensitivity); macro F1: unweighted harmonic mean of class precision and sensitivity; kappa: Cohen's kappa agreement coefficient; *L**: lightness; *a**: green–red coordinate; *b**: blue–yellow coordinate; *C**: chroma (C^*_{ab}); Lab: *L**, *a**, and *b** combined; Δ*E*: ΔE_{76} and ΔE_{00} calculated relative to the training centroid; All + Δ*E*: all ten descriptors (R, G, B, Hue, Sat, Val, *L**, *a**, *b**, and C^*_{ab}) plus ΔE_{76} and ΔE_{00}.

The Lab confusion matrices (Fig. 5) show the sources of classification errors and differences between raters. For Ton A1, only four classes were observed (1–4), with sensitivities of 84.6% for white (11/13), 64.3% for cream (9/14), 28.6% for brown (2/7), and 66.7% for reddish-brown (4/6). Rater 1 assigned none of the 40 plots to class 5 (green), rating BRS Verde samples mainly as cream (2) or brown (3).

Rater 2 (Ton A2) used all five classes, and Lab model sensitivity was 83.3% for white (10/12), 85.7% for cream (12/14), 33.3% for brown (1/3), 85.7% for reddish-brown (6/7), and 75.0% for green (3/4). Both raters had the lowest sensitivity for the intermediate brown class (3), as light-brown genotypes (BRS Jade and BRS Topázio) and oxidized greenish lint (BRS Verde) were rated mainly as cream (class 2).

For Rater 1 uniformity ratings (Unif. A1), Lab performed best overall, with 75.0% accuracy (30/40 correct predictions), exceeding the majority-class baseline of 55.0% by 20.0 percentage points (balanced accuracy = 77.5%, macro F1 = 0.735, κ = 0.565; Table 7). Sensitivity was 86.4% for uniform samples (19/22), 46.2% for moderately uniform samples (6/13), and 100.0% for nonuniform samples (5/5) (Fig. 5), whereas adding chroma or ΔE to Lab reduced accuracy to 70.0%. For Rater 2 uniformity ratings (Unif. A2), all predictor sets performed substantially less well: the highest overall accuracy was 55.0% (22/40 correct predictions with a* alone and All + ΔE), only three plots above the majority-class baseline

(47.5%), with κ ranging from 0.235 (Lab) to 0.288 (a* alone) and sensitivity of 23.1% for the moderately uniform class (Fig. 5).

In the exploratory OOB permutation importance analysis of the full model (All + ΔE; Supplementary material), G, L*, and B were the leading predictors of shade for A1, whereas B, G, and L* ranked highest for A2. For uniformity, ΔE00, a*, and R had the highest importance for A1, while R, a*, and Val ranked highest for A2.

4 Discussion

4.1 Genotypic control, field fading, and colorimetric divergence

The predominance of genotypic variance over genotype × growing season interaction variance for L*, a*, b*, and C*ab (5.5–12.6 times higher; Table 2) is consistent with the oligogenic control of fiber pigmentation in cotton, including proanthocyanidin regulatory genes such as GhTT2-3A on chromosome A07 (Naoumkina et al. 2024). Nevertheless, the contrasting interannual trajectories observed between 2023/2024 and 2024/2025 (Table 5) reflect interactions between flavonoid biosynthesis and weather conditions. Cumulative rainfall increased from 680 mm in 2023/2024 to 930 mm in 2024/2025, altering cloud cover, water availability, and temperature during secondary cell wall deposition (20–45 days after anthesis) and prolonging open-boll exposure in the field (Naoumkina et al. 2024; Fernandes et al. 2026).

The close colorimetric similarity between BRS Verde and BRS Jade in the PC1–PC2 plane (Fig. 2) and dendrogram (Fig. 3), often reported as a challenge in field phenotyping (Fernandes et al. 2026), has a biochemical explanation. Unlike the stable proanthocyanidin polymers of dark-brown cultivars (which give BRS Rubi its condensed tannins oxidized to quinones), green coloration in G. hirsutum fibers is associated with photosensitive suberin polymers and caffeic acid esters that undergo rapid oxidation and photodegradation when open bolls are exposed to ultraviolet radiation and weathering before harvest (Carvalho et al. 2011; Naoumkina et al. 2024). Consequently, harvested BRS Verde lint loses selective green reflectance (a* = −0.30) and approaches the greenish-cream shade of BRS Jade. Li et al. (2020) similarly reported uneven coloration in green fibers relative to brown and white accessions.

In addition, CIELAB coordinates captured subtle distinctions within cream-colored breeding materials, such as the higher yellowness and chroma of BRS Topázio compared with BRS Jade, while preserving the hue direction (a* < 0 vs. a* > 0) omitted by C*ab alone (Matusiak 2015; Kulsum et al. 2024). The divergent ranking of the closest pairs by ΔE76 and ΔE00 (Table 4) illustrates the psychometric advantage of ΔE00, which assigns greater weight to hue differences (ΔH*) between BRS Verde and BRS Jade than to chroma differences along the b* axis between BRS Jade and BRS Topázio (Sharma et al. 2005; Matusiak 2015). From an industrial perspective, interannual ΔE00 shifts of 4.89–7.96 at the same site indicate that lot-by-lot digital monitoring is valuable for assembling homogeneous batches before spinning.

4.2 Visual agreement, supervised prediction, and predictor redundancy

Although exact interrater agreement was 72.5% for both traits (Table 6), Cohen's kappa was lower for uniformity (0.548) than for shade (0.629) because the three-class uniformity scale concentrated observations in class 1, increasing chance agreement. Rater 1 shade distinctions were more closely associated with the green–red axis (a*), whereas Rater 2 relied more strongly on lightness (L*); nevertheless, the three-dimensional Lab set (L*, a*, and b*) provided the most consistent predictive basis across both raters (70.0% for A1 and 80.0% for A2). These accuracies are comparable to the 80.2–85.3% reported by Fisher et al. (2023) for predicting human visual grades of Egyptian cotton, whereas the 96% accuracy reported by Gonçalves et al. (2015) concerned cultivar-pair discrimination using full histograms rather than visual grading.

Adding C*ab to Lab did not improve shade accuracy and reduced uniformity accuracy by 5.0 percentage points (Table 7), because C*ab is a deterministic nonlinear function of a* and b* and dilutes variable selection at decision-tree splits. Similarly, in the full model (All + ΔE), Random Forest distributed permutation importance across collinear RGB and CIELAB variables (such as G and B alongside L* and b*), confirming informational redundancy without undermining the standardization advantage of CIELAB. Prior textile studies using artificial neural networks and multi-output regression likewise demonstrated that CIELAB coordinates effectively capture dyeing and chemical finishing variations when modeled appropriately (Balcı and Tuğrul Oğulata 2009; Şahin et al. 2023; Chen et al. 2024).

The sharp contrast in uniformity prediction between Unif. A1 (75.0%) and Unif. A2 (52.5–55.0%) demonstrates that plot-level color means do not directly measure within-sample mottling. Moderate prediction of Unif. A1 occurred indirectly because genotypes with stronger canopy-third color gradients, such as BRS Rubi and BRS Verde (Fernandes et al. 2026), were more frequently rated as nonuniform. Direct, transferable quantification of color uniformity requires second-order spatial texture and dispersion metrics—such as within-image standard deviations, pixel interquartile ranges, and segmentation of shadows and impurities (Cui et al. 2014; Vik et al. 2017; Heng et al. 2020; Fisher et al. 2023; Li et al. 2025).

4.3 Limitations of the study

Several limitations should be considered when interpreting these findings. First, the experiment evaluated five genotypes at a single location over two growing seasons (n = 40 plots), so the models serve as a proof of concept rather than a universally calibrated grading system across diverse environments and germplasm. Second, visual reference labels were provided by only two raters whose interrater agreement was moderate to substantial (κ = 0.548–0.629), which caps the achievable accuracy of supervised models and introduces rater-specific bias, particularly for intermediate brown and uniformity classes. Third, because no external ceramic white standard was available during image acquisition, ΔE76 and ΔE00 predictors had to be computed relative to the training-fold CIELAB centroid

rather than an absolute colorimetric standard. Fourth, the smartphone-based protocol was not directly benchmarked against an integrating-sphere spectrophotometer or HVI/DigiEye system under standard illuminant D65, nor were multi-device or multi-operator reproducibility trials conducted (Pabuayon et al. 2019). Finally, extracting only plot-level mean color coordinates precluded direct capture of spatial mottling, which limited color uniformity prediction.

4.4 Complementary applications in the textile industry and plant breeding

Within these boundaries, the tested protocol provides a practical, low-cost screening tool for color phenotyping in naturally colored cotton, with complementary roles for each metric: L* measures lightness, a* and b* define chromatic direction, C*ab and h°ab express polar saturation and hue, and ΔE00 quantifies perceptual distance across genotypes or growing seasons.

For the textile industry and colored-cotton processing facilities, lightbox imaging calibrated with a reference card offers a rapid screening aid at bale intake to monitor shade shifts among lots and growing seasons (Fisher et al. 2023; Li et al. 2025). However, definitive commercial lot certification and blending decisions still require independent validation across devices and operators, defined tolerance limits, and calibration curves traceable to integrating-sphere spectrophotometers or DigiEye systems under standard illuminant D65 (Matusiak 2015; Pabuayon et al. 2019; Kulsum et al. 2024).

In breeding programs, routine recording of CIELAB coordinates can support selection of lines with greater pigment saturation (higher C*ab and lower L* in brown cultivars, or more negative a* and reduced photodegradation in green cultivars) and lower interannual ΔE00 variation. Combined with technological fiber traits such as length, strength, and micronaire (Fernandes et al. 2026), this approach can assist early-generation screening of genotypes combining color stability and fiber quality.

5 Conclusions

Calibrated digital image analysis using the three-dimensional CIELAB coordinates (L*, a*, and b*) provides a parsimonious, physically interpretable, and reproducible framework for characterizing lint color and supporting shade screening in naturally colored cotton without requiring redundant derived metrics such as C*ab or ΔE in Random Forest classifiers. While plot-level mean coordinates effectively capture genotypic divergence and interannual color shifts, reliable prediction of visual color uniformity requires pixel-level spatial dispersion descriptors and standardized multi-rater calibration. As a proof-of-concept screening tool, this smartphone-based protocol can support breeding selection and preliminary lot sorting once validated against benchtop spectrophotometers across multiple devices and environments.

Abbreviations

BA

Balanced accuracy
BI
Brightness index
BLUP
Best linear unbiased prediction
C*ab
Chroma / saturation
CAPES
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CIE
Commission Internationale de l'Éclairage
DAE
Days after emergence
$h°_{ab}$
Hue angle
HSV
Hue, Saturation, Value
HVI
High Volume Instrument
L*, a*, b*
CIELAB color space coordinates (lightness, green–red axis, blue–yellow axis)
LOOCV
Leave-one-plot-out cross-validation
NGRDI
Normalized green–red difference index
OOB
Out-of-bag
PC1 / PC2
First and second principal components
PCA
Principal component analysis
RCBD
Randomized complete block design
REML
Restricted maximum likelihood
RGB
Red, Green, Blue channels
SG 821
Suregrow 821
SI

Saturation spectral slope index
UFU
Universidade Federal de Uberlândia
ΔE_{00}
CIEDE2000 color difference
ΔE_{76}
CIE76 Euclidean distance

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article. The raw data are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported in part by CAPES – Finance Code 001.

Authors' contributions

Fernandes GA, Matias FI, and Sousa LB conceived and designed the experiment. Fernandes GA, Dias ILA, Costa ALG, Corrêa MM, and Borges MLV conducted the field experiments and collected the image and fiber quality data. Fernandes GA, Matias FI, and Sousa LB performed the statistical analyses and predictive modeling using R software, interpreted the data, conducted the literature review, and were the main contributors to the manuscript writing. All authors read and approved the final version of the manuscript.

Acknowledgements

This study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Finance Code 001. The authors thank the members of PROMALG for their support and

collaboration, as well as AMIPA and Minas Cotton Laboratory for technical assistance in fiber analysis. Supplementary materials are available from the corresponding author upon request.

References

1. Balcı O, Tuğrul Oğulata R. Prediction of the changes on the CIELab values of fabric after chemical finishing using artificial neural network and linear regression models. Fibers Polym. 2009;10(3):384–93. https://doi.org/10.1007/s12221-009-0384-2.
2. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48. https://doi.org/10.18637/jss.v067.i01.
3. Carvalho LP, Andrade FP, Silva Filho JL. Cultivares de algodão colorido no Brasil. Rev Bras Oleaginosas Fibrosas. 2011;15(1):37–44.
4. Chen Z, Liu J, Li J, et al. Leveraging multi-output modelling for CIELAB using colour difference formula towards sustainable textile dyeing. Auton Intell Syst. 2024;4:19. https://doi.org/10.1007/s43684-024-00076-8.
5. CIE. Colorimetry. 4th ed. CIE 015:2018. Vienna: International Commission on Illumination. 2018. https://doi.org/10.25039/TR.015.2018
6. Cui X, Cai Y, Rodgers J, et al. An investigation into the intra-sample variation in the color of cotton using image analysis. Text Res J. 2014;84(2):214–22. https://doi.org/10.1177/0040517513490055.
7. Fernandes GA, Medeiros LA, Matias FI, et al. Quantifying naturally colored cotton fiber color intensity using RGB image-based phenotyping. Crop Breed Appl Biotechnol. 2026;26(3):e55682637. https://doi.org/10.1590/1984-70332026v26n3a37.
8. Fisher OJ, Rady A, El-Banna AAA, et al. AI-assisted cotton grading: active and semi-supervised learning to reduce the image-labelling burden. Sensors. 2023;23(21):8671. https://doi.org/10.3390/s23218671.
9. Gonçalves MIS, Vilar WTS, Medeiros EP, Pontes MJC. A novel strategy for the classification of naturally colored cotton fibers based on digital imaging and pattern recognition techniques. Anal Methods. 2015;7(14):5869–75. https://doi.org/10.1039/C5AY01116G.
10. Heng C, Chen L, Shen H, Wang F. Study on the measurement and evaluation of cotton color using image analysis. Mater Res Express. 2020;7(7):075101. https://doi.org/10.1088/2053-1591/aba3e4.
11. Kulsum U, Islam T, Sayed A, et al. Analyzing the differences between spectrophotometer and DigiEye assessments of knit fabric. Int J Text Sci. 2024;13(2):11–21. https://doi.org/10.5923/j.textile.20241302.01.
12. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Softw. 2017;82(13):1–26. https://doi.org/10.18637/jss.v082.i13.
13. Li H, Li Q, Zhou W, et al. Measurement of seed cotton color using RGB imaging and Color-Unet. Agronomy. 2025;15(1):19. https://doi.org/10.3390/agronomy15010019.

14. Li Z, Su Q, Xu M, et al. Phenylpropanoid metabolism and pigmentation show divergent patterns between brown color and green color cottons as revealed by metabolic and gene expression analyses. J Cotton Res. 2020;3:27. https://doi.org/10.1186/s42397-020-00069-x.
15. Matias FI, Caraza-Harter MV, Endelman JB. FIELDimageR: an R package to analyze orthomosaic images from agricultural field trials. Plant Phenome J. 2020;3(1):e20005. https://doi.org/10.1002/ppj2.20005.
16. Matusiak M. DigiEye application in cotton colour measurement. AUTEX Res J. 2015;15(2):77–86. https://doi.org/10.2478/aut-2014-0036.
17. Naoumkina M, Hinchliffe DJ, Thyssen GN. Naturally colored cotton for wearable applications. Front Plant Sci. 2024;15:1350405. https://doi.org/10.3389/fpls.2024.1350405.
18. Pabuayon ILB, Sun Y, Guo W, et al. High-throughput phenotyping in cotton: a review. J Cotton Res. 2019;2:18. https://doi.org/10.1186/s42397-019-0035-0.
19. Roberts DR, Bahn V, Ciuti S, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography. 2017;40(8):913–29. https://doi.org/10.1111/ecog.02881.
20. Şahin C, Balcı O, Işık M, Gökenç İ. Artificial neural networks approach for prediction of CIELab values for yarn after dyeing and finishing process. J Text Inst. 2023;114(9):1326–35. https://doi.org/10.1080/00405000.2022.2124629.
21. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–5. https://doi.org/10.1038/nmeth.2089.
22. Sharma G, Wu W, Dalal EN. The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. Color Res Appl. 2005;30(1):21–30. https://doi.org/10.1002/col.20070.
23. Vik M, Khan N, Yildirim B, Viková M. Non-contact method for measurement of colour variation in a cotton sample. Fibres Text East Eur. 2017;25(2):106–11. https://doi.org/10.5604/12303666.1228180.
24. Wright MN, Ziegler A. ranger: a fast implementation of random forests for high dimensional data in C + + and R. J Stat Softw. 2017;77(1):1–17. https://doi.org/10.18637/jss.v077.i01.
25. Xu B, Fang C, Huang R, Watson MD. Cotton color measurements by an imaging colorimeter. Text Res J. 1998;68(5):351–8. https://doi.org/10.1177/004051759806800505.

Table 3

Table 3 is available in the Supplementary Files section.

Figures

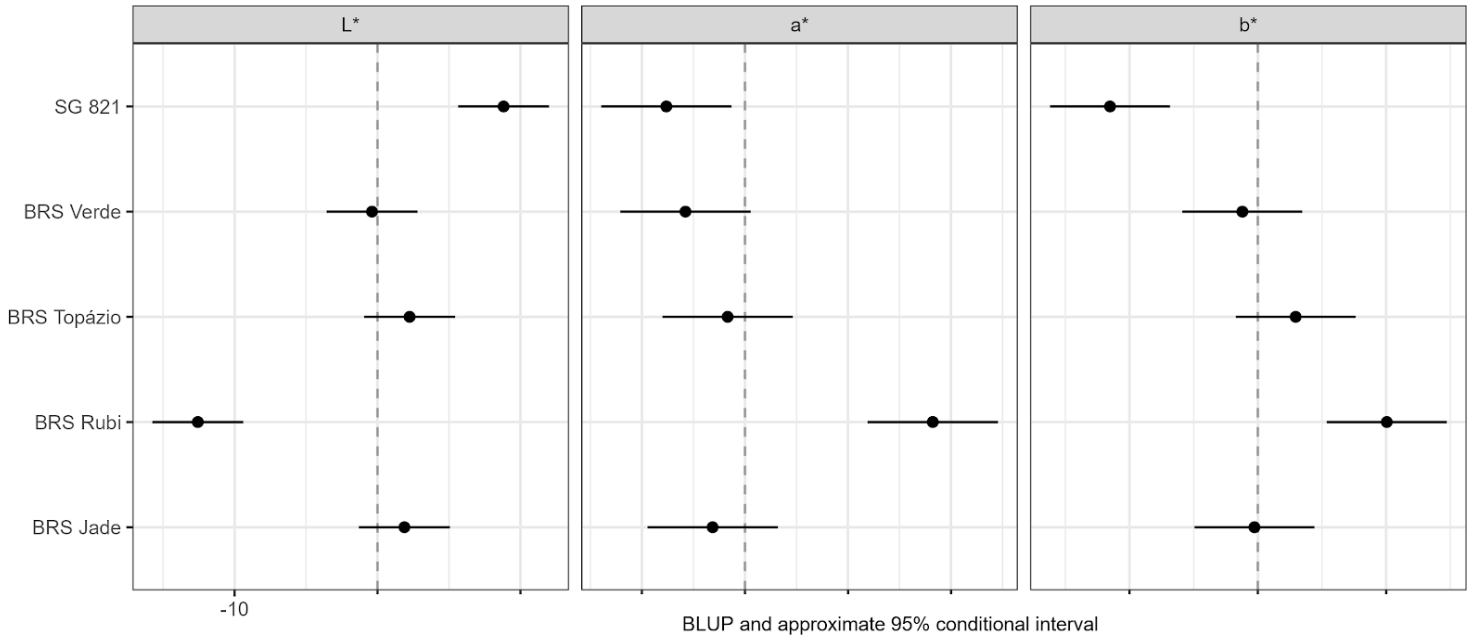


Figure 1

Predicted genotypic deviations (BLUPs) for *L**, *a**, and *b** with 95% conditional prediction intervals. The dashed vertical line indicates zero deviation from the adjusted overall mean.

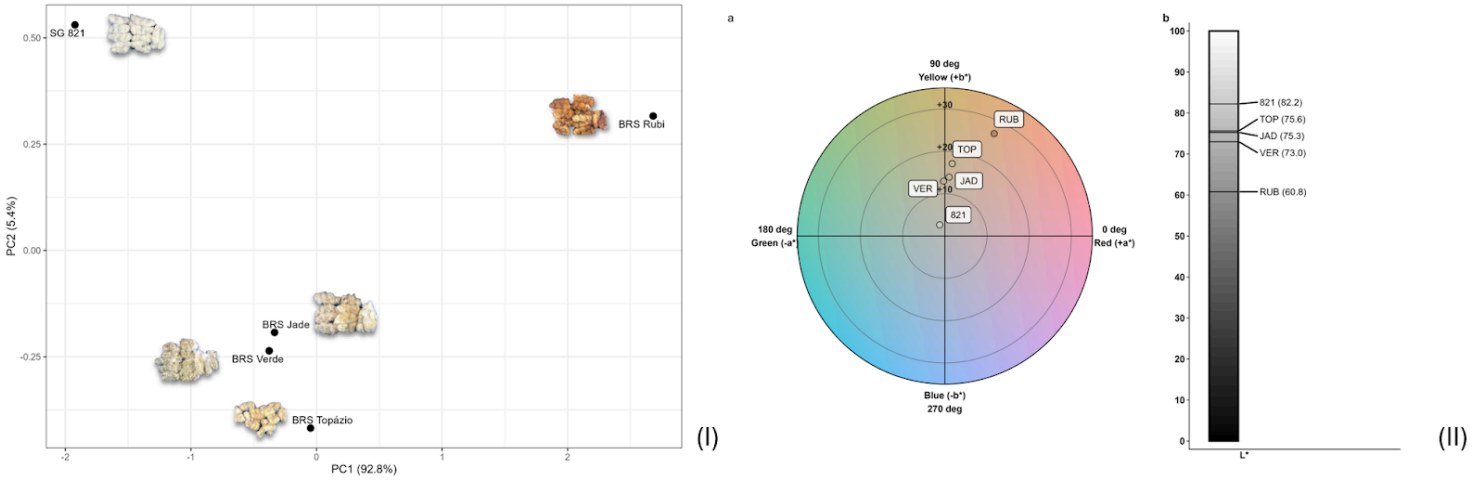


Figure 2

Colorimetric characterization of five cotton genotypes based on predicted genotypic values of L*, a*, and b*. (I) Scores on the first two principal components (PC1 = 92.8%; PC2 = 5.4%) derived from standardized coordinates, illustrated with lint samples from each cultivar. (II) Genotype positions in CIELAB space, showing the a*×b* chromatic plane with chroma isolines (C*ab, +10 to +30) on the left and the vertical lightness scale (L*, 0–100) on the right. 821: SG 821; JAD: BRS Jade; VER: BRS Verde; TOP: BRS Topázio; RUB: BRS Rubi.

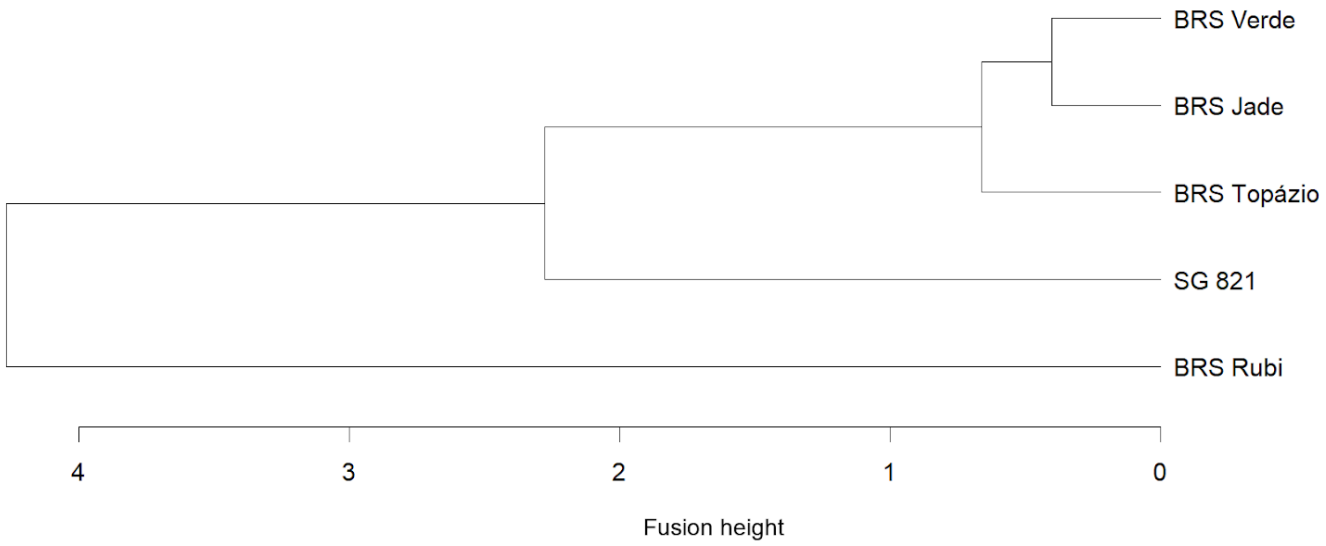


Figure 3

Color dissimilarity dendrogram obtained using Ward.D2 hierarchical clustering of Euclidean distances between standardized predicted genotypic values of *L**, *a**, and *b**.

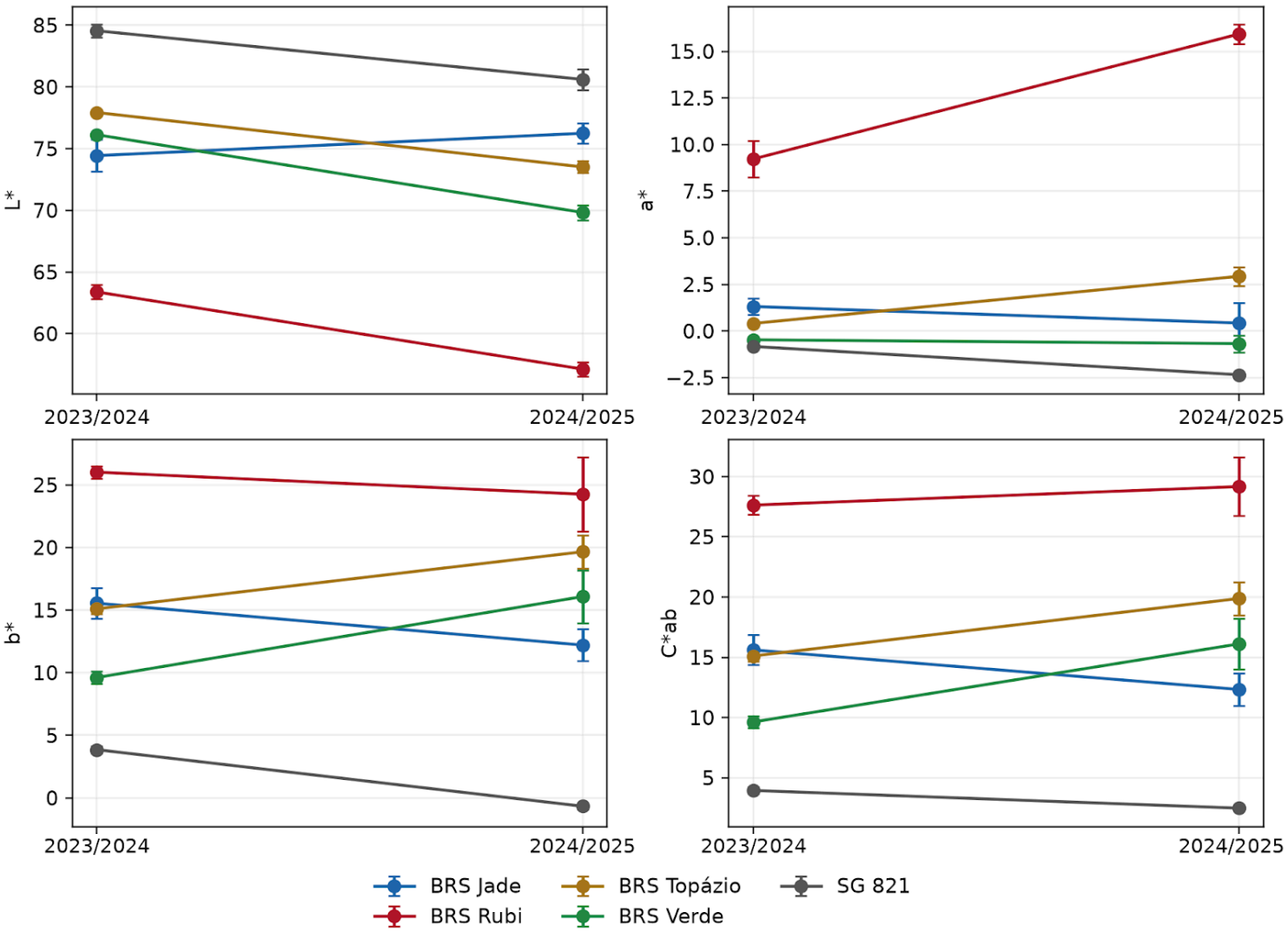


Figure 4

Interannual trajectories of observed means for L*, a*, b*, and chroma (C*ab) in five genotypes between the 2023/2024 and 2024/2025 growing seasons. Vertical bars represent the standard error of the mean (n = 4 plots per genotype per growing season).

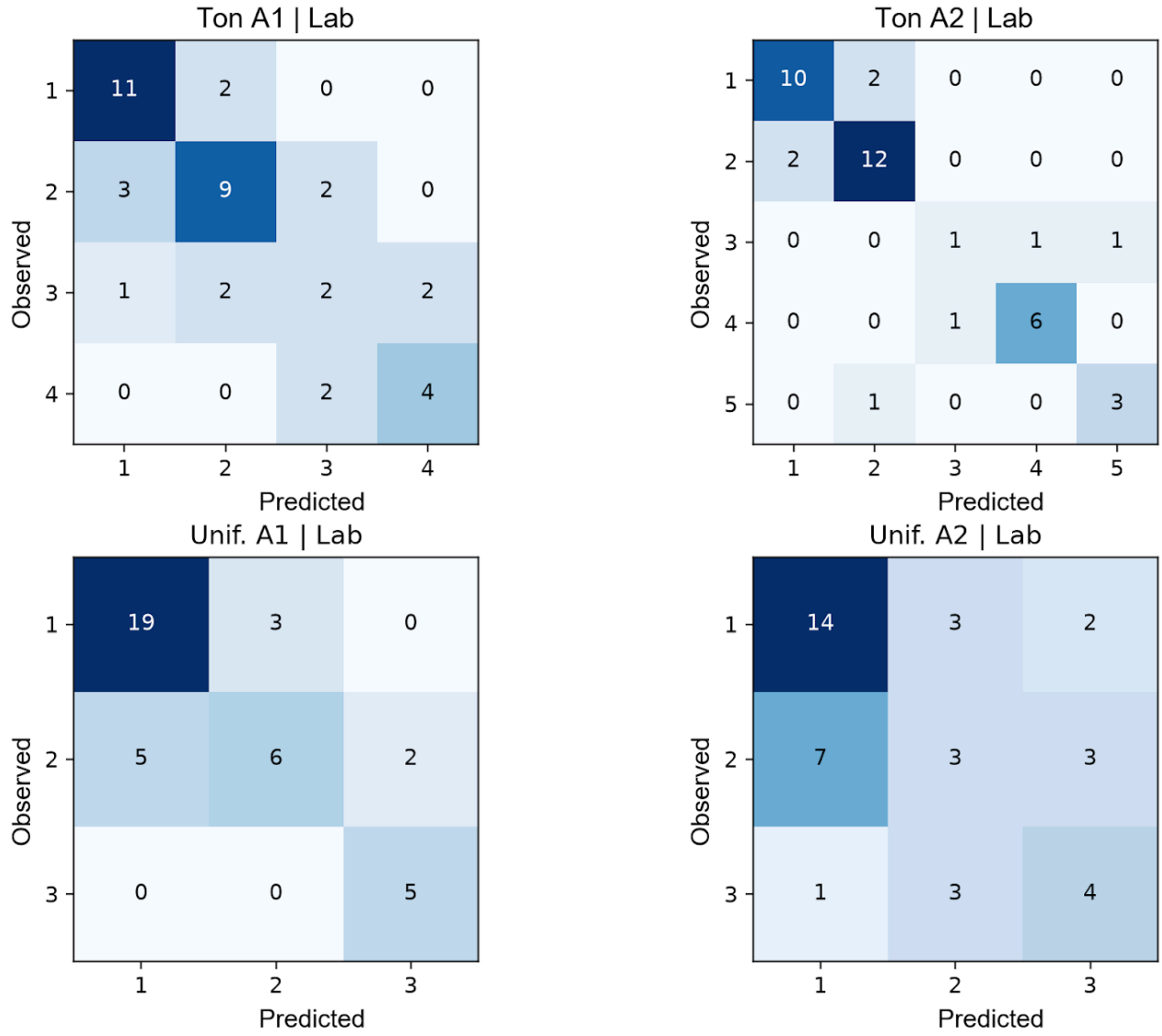


Figure 5

Confusion matrices from the Random Forest model using Lab predictors (*L**, *a**, and *b**) under LOOCV for shade (Ton A1 and Ton A2) and uniformity (Unif. A1 and Unif. A2). Shade classes: 1 = white, 2 = cream, 3 = brown, 4 = reddish-brown, and 5 = green (not assigned by A1). Uniformity classes: 1 = uniform, 2 = moderately uniform, and 3 = nonuniform. Rows represent observed visual classes and columns represent predicted classes (n = 40 plots).

Supplementary Files

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

- Table3.docx
- JCRGAFsupplementarmaterial.docx