

Potential of UAV-derived RGB spectral indices for the early selection of cotton genotypes based on fiber quality traits

Gabriel Aragão Fernandes

gabrielaragaoagro@gmail.com

Federal University of Uberlândia

Filipe Inácio Matias

Universidade Estadual Paulista, Produção Vegetal

Isabela Lorrane Abreu Dias

Federal University of Uberlândia

Ana Luiza Gonçalves Costa

Federal University of Uberlândia

Marcello Matheus Corrêa

Federal University of Uberlândia

Michelle Souza Santos

Federal University of Uberlândia

Ana Paula Oliveira Nogueira

Federal University of Uberlândia

Larissa Barbosa de Sousa

Federal University of Uberlândia

Research Article

Keywords: *Gossypium hirsutum* L., high-throughput phenotyping, High Volume Instrument, remote sensing

Posted Date: August 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10733313/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Potential of UAV-derived RGB spectral indices for the early selection of cotton genotypes based on fiber quality traits

Gabriel Aragão Fernandes¹, Filipe Inácio Matias², Isabela Lorrane Abreu Dias¹, Ana Luiza Gonçalves Costa¹, Marcello Matheus Corrêa¹, Michelle Souza Santos¹, Ana Paula Oliveira Nogueira¹, Larissa Barbosa de Sousa¹

¹Federal University of Uberlândia, Uberlândia, Brazil. ²Universidade Estadual Paulista, Produção Vegetal, Via de Acesso Professor Paulo Donato Castellane s/n, 14884-900, Jaboticabal, SP, Brazil.

Corresponding author: Gabriel Aragão Fernandes. gabrielaragaoagro@gmail.com

ABSTRACT

This study evaluated the potential of UAV-derived RGB spectral indices to predict fiber quality traits in cotton genotypes. Nineteen genotypes were assessed under a randomized complete block design, and RGB imagery acquired at full flowering was used to calculate GLI, NGRDI, SCId, and SI. Significant genetic variability and moderate-to-high heritability were observed for both fiber traits and spectral indices. GLI was positively associated with the Spinning Consistency Index, whereas NGRDI was associated with fiber length uniformity. Regression models showed moderate predictive ability (R^2_{LOOCV} between 32.4–35.4%; and accuracy between 0.57–0.60). GLI and NGRDI demonstrated potential as complementary tools for large-scale phenotyping and preliminary genotype selection, although they do not replace conventional fiber quality analyses. Further studies across additional developmental stages are needed to improve prediction accuracy.

Keywords: *Gossypium hirsutum* L.; high-throughput phenotyping; High Volume Instrument; remote sensing.

Background

Cotton (*Gossypium hirsutum* L.) is the world's leading source of natural fiber, with India, China, and Brazil ranking among the largest producers (USDA, 2026). Fiber represents the highest-value product of the cotton crop, and its commercial value is primarily determined by technological properties required

by the textile industry. The most important fiber quality traits include upper half mean length (UHML), fiber length uniformity (UI), fiber strength (Str), micronaire (Mic), fiber elongation (Elg), yellowness (+b), fiber maturity (Mat), Spinning Consistency Index (SCI), and Short Fiber Index (SF), all of which directly influence spinning performance (Belot, 2018). These traits are routinely evaluated after harvest using representative lint samples analyzed in specialized laboratories with the High-Volume Instrument (HVI) system (Sayeed et al., 2021).

In cotton breeding programs, however, fiber quality is generally assessed only in advanced generations because HVI analyses require relatively large amounts of lint. Consequently, early-stage selection is primarily based on agronomic traits, such as seed cotton yield, plant architecture, and lint percentage. The availability of reliable methods for the early screening of genotypes for fiber quality could substantially reduce phenotyping costs while increasing genetic gain through more efficient selection (Vidal Neto and Freire, 2013).

High-throughput phenotyping (HTP) has emerged as a promising alternative for this purpose by combining remote sensing technologies with image-based analyses to derive spectral indices associated with plant physiological status. These indices may be correlated with agronomic and technological traits, enabling rapid, non-destructive, and large-scale phenotypic evaluations (Herrmann Júnior et al., 2024).

Among HTP platforms, unmanned aerial vehicles (UAVs) have been widely adopted for developing predictive models in agricultural crops. In cotton, most studies have focused on yield prediction. For example, Feng et al. (2020) reported that plant height and the Cotton Fiber Index (CFI), measured during flowering and nine days before harvest, explained cotton yield with high predictive accuracy ($R^2 = 0.90$). Likewise, in maize breeding, UAV-derived spectral indices showed strong associations with yield-related traits, with regression coefficients of 0.92 for ear width and 0.88 for ear length (Resende et al., 2024), showing that different application of HTP on crop evaluation.

Despite the increasing application of UAV imagery in HTP across several crops, studies investigating its potential for the indirect estimation of cotton fiber quality traits remain scarce. Predictive

methodologies based on spectral indices could provide valuable support for breeding programs by enabling the early identification of superior genotypes, reducing phenotyping costs, and improving selection efficiency. Although UAV-based phenotyping has been extensively investigated for yield prediction in cotton, its application to the indirect prediction of fiber quality traits remains poorly explored. This knowledge gap limits the use of remote sensing as an early selection tool in cotton breeding programs. Moreover, these approaches may contribute to crop management by facilitating the identification of field areas with greater productive and qualitative potential. Therefore, this study aimed to evaluate the potential of spectral indices derived from UAV-based RGB imagery acquired at the full-flowering stage to predict fiber quality traits in cotton genotypes.

Methods

Local

The experiment was conducted during the 2024/25 growing season at Glória Experimental Farm, Federal University of Uberlândia (UFU), Uberlândia, Minas Gerais, Brazil (18°56'52.77"S, 48°12'51.97"W; 865 m above sea level). Sowing was performed on December 17, 2024, and harvesting took place in August 2025. During the experimental period, the average air temperature ranged from 19 to 23 °C, with cumulative rainfall between 1,000 and 1,050 mm.

Vegetative material and experimental design

The experiment was established in a randomized complete block design (RCBD) with 19 commercial cotton genotypes and three replications. The evaluated genotypes were obtained from Embrapa, Deltapine, Syngenta, FiberMax, IMAmt, and TMG: BRS 286, DP 1857, DP 1866, DP 1949, DP 2077, DP 2176, DP 555, Fabrika, FM 911, FM 912, FM 942, FM 944, FM 970, FM 985, IMA 1327, IMA 243, TMG 31, TMG 44, and TMG 82.

Each experimental plot consisted of two 5 m rows spaced 0.90 m apart. A 0.50 m border was excluded from each end of the rows, resulting in a useful plot area comprising the central portions of both rows. Plant density was approximately 80,000 plants ha⁻¹. At sowing, 20 kg ha⁻¹ of nitrogen and 60 kg ha⁻¹ of phosphorus were applied. Topdressing consisted of 60 kg ha⁻¹ of nitrogen and 30 kg ha⁻¹ of potassium, split into two applications at 25 and 45 days after emergence (DAE). Additionally, 1 kg ha⁻¹ of boron was supplied through four foliar applications, with the first application at 45 DAE and subsequent applications at 65, 85 and 105

DAE. Plant growth was regulated using chlormequat chloride ((2-Chloroethyl)trimethylammonium chloride), with the first application at 45 DAE, and subsequent applications were scheduled according to plant growth to maintain a final plant height of approximately 1.3 m at harvest.

Traits

The following agronomic and fiber quality traits were evaluated. Mean boll weight (MBW, g) was determined from the average weight of 30 bolls collected from each experimental plot. Lint percentage (LP, %) was calculated from the ratio between lint weight and total seed cotton weight after ginning. Fiber quality traits were determined using the High Volume Instrument (HVI) system, including: (i) upper half mean length (UHML, mm); (ii) fiber length uniformity (UI, %); (iii) fiber strength (Str, gf tex⁻¹); (iv) micronaire (Mic); (v) fiber elongation (Elg, %); (vi) yellowness (+b); (vii) fiber maturity (Mat, %); (viii) Spinning Consistency Index (SCI); and (ix) short fiber index (SF).

Flight plan and equipment

RGB images were acquired at the full-flowering stage using a DJI Mavic 3M unmanned aerial vehicle (DJI, Shenzhen, China), when plants exhibited a high number of fully opened flowers. The flight mission was planned to obtain a ground sampling distance (GSD) of approximately 0.55 cm pixel⁻¹, with a flight speed of 1.3 m s⁻¹ and 75% forward and side overlap. Orthomosaics were generated using WebODM software. Subsequently, spectral indices were calculated from the visible bands, including red (620–750 nm), green (495–570 nm), and blue (450–495 nm) (Figure 1).

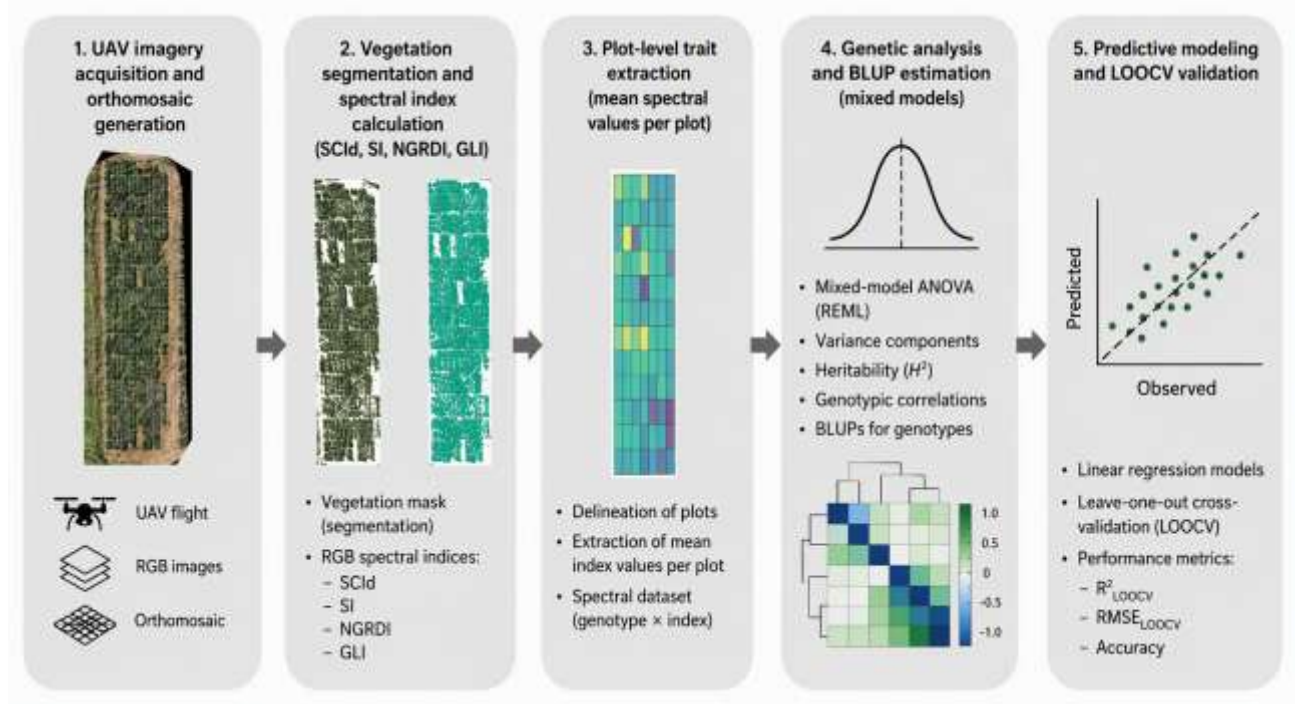


Figure 1. Workflow of UAV-based phenotyping, genetic analysis, and predictive modeling in cotton genotypes. (1) UAV image acquisition and orthomosaic generation; (2) vegetation segmentation and RGB spectral index calculation (SCId, SI, NGRDI, and GLI); (3) plot-level spectral trait extraction; (4) genetic analysis and BLUP estimation using mixed models (REML/BLUP); and (5) predictive modeling and leave-one-out cross-validation (LOOCV) for fiber-quality traits.

Image processing and spectral index extraction were performed in R (R Core Team, 2026) using the FIELDimageR package (Matias et al., 2020). The following RGB-derived indices were calculated: (i) Spectral Color Index (SCId)^[9] (Gitelson et al., 2002); (ii) Saturation Index (SI)^[10] (Bendig et al., 2015); (iii) Normalized Green Red Difference Index (NGRDI)^[11] (Woebbecke et al., 1993); (iv) Green Leaf Index (GLI)^[12] (Louhaichi et al., 2001).

Statistical analysis

Initially, data were subjected to analysis of variance (ANOVA) to determine the existence of significant differences among genotypes, following model $Y_{ij} = \mu + g_i + b_j + \varepsilon_{ij}$, where, y_{ij} = observation of the treatment in block j ; μ = overall mean, g_i = effect of the i -th treatment, b_j = effect of the j -th block, ε_{ij} = random error associated with the observation y_{ij} .

Genetic parameters were estimated from the ANOVA variance components, including genetic variance (σ_g^2), environmental variance (σ_e^2), and genotypic determination coefficient (H^2), following the expressions:

$$\sigma_g^2 = \frac{MS_g - MS_e}{r}$$

$$\sigma_e^2 = MS_{error}$$

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

Where, MS_g is mean square of genotype, MS_e is mean square of error. Genotype means were grouped using the Scott–Knott test at the 5% significance level. Phenotypic and genotypic correlation coefficients were subsequently estimated between fiber quality traits and UAV-derived spectral indices. Confidence intervals were obtained by 5,000 bootstrap resamplings using the GENES software (Cruz, 2013).

Traits showing significant genotypic correlations with spectral indices were further analyzed using univariate linear mixed models to estimate variance components by restricted maximum likelihood (REML) and predict genotypic values using best linear unbiased prediction (BLUP). In these models, block effects were considered fixed, whereas genotype effects were treated as random. Selective accuracy was estimated as the square root of the mean progeny heritability.

To evaluate the predictive ability of spectral indices, simple linear regression models were fitted using BLUP estimates. Owing to the relatively small number of genotypes, model performance was assessed by leave-one-out cross-validation (LOOCV), in which each genotype was sequentially omitted from model fitting and subsequently used for independent prediction. Predictive performance was quantified using the cross-validated coefficient of determination (R^2_{LOOCV}), root mean square error ($RMSE_{LOOCV}$) and mean absolute error (MAE_{LOOCV}).

Results and discussion

Analysis of variance revealed that only MBW, LP, Mat, Rd and +b did not differ significantly among genotypes. In contrast, significant genetic variability was detected for the main fiber quality traits, including SCI, UHML, Mic, Str, UI, SF and Elg. Likewise, significant variation was observed for all evaluated RGB-derived spectral indices, namely SCId, NGRDI, SI and GLI (Table 1).

Table 1. Analysis of variance for yield traits, fiber quality characteristics, and spectral indices of 19 cotton genotypes. Uberlândia, Minas Gerais, Brazil.

Trait	Block	Genotype	Residual	Mean	CV (%)	ϕ_p	ϕ_g	H (%)	CV _g (%)
	2	18	36						
WBM	0.408	0.769 ^{ns}	0.606	4.67	16.64	0.256	0.054	21.14	4.97

LP	8.919	13.563 ^{ns}	8.759	43.67	6.77	4.521	1.601	35.41	2.89
SCI	20.894	206.947**	42.783	147.26	4.44	68.982	14.261	79.32	5.02
Mic	0.018	0.100**	0.038	4.92	3.96	0.033	0.02	61.77	2.91
Mat	0.000007	0.000102 ^{ns}	0.000039	0.87	0.71	0.000034	0.000021	61.92	0.52
UHML	0.21	3.992**	0.391	30.62	2.04	1.33	1.2	90.18	3.57
UI	0.06	1.933*	0.826	84.50	1.07	0.644	0.368	57.25	0.71
SF	0.658	2.223**	0.486	6.36	10.96	0.741	0.579	78.12	11.95
Str	3.496	3.638**	0.8765	33.00	2.83	1.212	0.92	75.91	2.9
Elg	0.001	0.585**	0.105	6.86	4.74	0.195	0.035	81.92	5.82
Rd	1.194	2.313 ^{ns}	1.404	76.58	1.54	0.771	0.302	39.29	0.71
+b	0.666	0.506 ^{ns}	0.514	7.28	9.84	0.168	0.000	0.00	0.00
SCId	0.0019	0.0007**	0.000223	-0.09	16.16	0.0002	0.0001	69.45	14.07
SI	0.0003	0.0008*	0.0003	0.06	29.17	0.0002	0.0001	55.76	18.91
NGRDI	0.001	0.0007**	0.0002	0.092	16.02	0.0002	0.0001	68.5	13.64
GLI	0.002	0.0003**	0.0001	0.12	9.18	0.0001	0.00006	59.99	6.49

Means followed by different letters differ by the Scott-Knott test at 5% probability. ns = non-significant; **, * and indicate significance at 5% and 1% probability, respectively. MBW, mean boll weight; LP, linter percentage; UHML, mean fiber length; UI, fiber length uniformity; Str, fiber strength; Mic, micronaire; Mat, maturity; Elg, fiber elongation; Rd, reflectance; +b, yellowness; SCI, spinning consistency index; SF, short fiber index; SCId, Spectral Color Index; SI, Saturation Index; NGRDI, Normalized Green Red Difference Index; GLI, Green Leaf Index.

The absence of significant differences for MBW, LP, Mat, Rd, and +b likely reflects the relatively narrow genetic background of the commercial cultivars evaluated, which have undergone intensive selection to meet the productivity, yield, and fiber quality standards required by the textile industry. Consequently, these genotypes exhibit similar performance for these traits (Klein et al., 2022).

Broad-sense heritability (H^2) was generally high for fiber quality traits, exceeding 70% for most variables. The highest estimate was observed for UHML (90.18%), indicating that genetic effects accounted for most of the phenotypic variation. These results suggest that fiber quality traits are predominantly under genetic control, supporting the effectiveness of selection in cotton breeding programs (Carvalho et al., 2022).

In contrast, the spectral indices exhibited moderate heritability estimates, ranging from 55.76% for SI to 69.45% for SCId. The genotypic coefficient of variation (CVg) ranged from 0.52 to 18.91%, with the highest values observed for SF, SCId, SI, and NGRDI. From a breeding perspective, higher CVg values indicate greater exploitable genetic variability and, consequently, increased potential for genetic gain through selection (Yan, 2021).

Comparison of genotype means revealed considerable variation in fiber quality traits (Table 2). Among the evaluated genotypes, DP 2077 exhibited the most favorable technological profile, presenting the greatest UHML (34.17 mm) and the lowest SF (4.10%), while also belonging to the group with the highest SCI (167.33) and UI (86.13%). Although its fiber strength (33.40 gf tex⁻¹) was not the highest among the evaluated genotypes, the combined performance of these traits identifies DP 2077 as the most promising genotype in terms of fiber technological quality.

Table 2. Mean values of fiber quality traits and RGB-derived spectral indices for 19 commercial cotton genotypes evaluated in Uberlândia, Minas Gerais, Brazil.

Genótipos	SCI	UHML	UI	SF	Str	Elg	SCId	SI	NGRDI	GLI
BRS 286	138.00b	29.23c	84.10b	7.53a	31.46b	6.96c	-0.1083b	0.0551b	0.1058a	0.1299a
DP 1857	155.00a	31.02b	85.13a	6.00b	33.47a	6.73c	-0.1143b	0.0519b	0.1136a	0.1359a
DP 1866	147.33a	30.981b	83.76b	6.63a	33.96a	7.16c	-0.0856a	0.0664b	0.0863b	0.1170b
DP 1949	151.67a	29.62c	84.93a	6.13a	33.63a	6.67c	-0.1087b	0.0587b	0.1096a	0.1325a
DP 2077	167.33a	34.17a	86.13a	4.10c	33.40a	6.90c	-0.0957b	0.0732b	0.0975a	0.1312a
DP 2176	150.00a	30.78b	85.13a	5.93b	32.67b	6.93c	-0.1046b	0.0758b	0.1031a	0.1379a
DP 555	137.67b	29.31c	84.03b	7.46a	32.10b	6.67c	-0.0759a	0.0680b	0.0789b	0.1076b
Fabrika	132.33b	29.36c	83.30b	7.33a	31.53b	8.27a	-0.0787a	0.0589b	0.0790b	0.1047b
FM 911	148.33a	29.60c	84.60a	6.30a	35.00a	6.53c	-0.0706a	0.1046a	0.0715b	0.1188b
FM 912	153.33a	31.79b	84.80a	5.27b	33.10b	6.50c	-0.0644a	0.1028a	0.0640b	0.1094b
FM 942	145.67a	31.07b	85.26a	5.73b	31.70b	6.80c	-0.0949b	0.0636b	0.0954a	0.1226a
FM 944	151.00a	31.06b	84.10b	6.37a	34.63a	6.63c	-0.0807a	0.0707b	0.0817b	0.1131b
FM 970	142.64b	30.66b	83.76b	7.10a	32.76b	6.27c	-0.0813a	0.0904a	0.0816b	0.1213a
FM 985	137.33b	31.21b	83.06b	6.73a	31.67b	6.76c	-0.0886a	0.0785b	0.0875b	0.1325a
IMA 1327	146.33a	29.97c	84.50b	6.67a	32.93b	6.73c	-0.1085b	0.0507b	0.1100a	0.1297a
IMA 243	155.67a	30.68b	85.36a	5.90b	34.70a	7.13c	-0.1055b	0.0447b	0.1061a	0.1245a
TMG 31	147.67a	30.17c	84.36b	7.20a	33.63a	7.40b	-0.0956b	0.0585b	0.0952a	0.1205a
TMG 44	152.33a	31.32b	85.46a	5.57b	32.76b	7.00c	-0.1162b	0.0476b	0.1162a	0.1328a
TMG 82	138.33b	29.88c	83.76b	6.93a	32.06b	6.33c	-0.0790a	0.0617b	0.0794b	0.1079b

Means followed by the same lowercase letter within a column do not differ according to the Scott–Knott test ($P \leq 0.05$). UHML, mean fiber length; UI, fiber length uniformity; Str, fiber strength; Mic, micronaire; Elg, fiber elongation; SCI, spinning consistency index; SF, short fiber index; SCId, Spectral Color Index; SI, Saturation Index; NGRDI, Normalized Green Red Difference Index; GLI, Green Leaf Index.

Two statistically distinct groups were identified for SCI. Six genotypes (BRS 286, DP 555, Fabrika, FM 970, FM 985, and TMG 82) showed significantly lower SCI values than the remaining genotypes. Most of these cultivars were released between 2008 (DP 555) and 2023 (FM 970). The superior

performance of DP 2077 may reflect recent advances in cotton breeding, suggesting simultaneous genetic gains for UHML, SCI, and UI. Nevertheless, the relationship between cultivar release year and fiber quality is not necessarily linear, as breeding objectives differ among programs.

Other high-performing genotypes included IMA 243, DP 1857, TMG 44, FM 912, and FM 911, all of which exhibited SCI values above 150, UI exceeding 85%, and fiber strength within the standards required by the textile industry. Notably, FM 911 presented the highest fiber strength (35.00 gf tex⁻¹). Conversely, Fabrika, BRS 286, DP 555, FM 985, and TMG 82 consistently showed the lowest SCI values (132.33–138.33), together with shorter fibers and lower fiber length uniformity.

The Spinning Consistency Index (SCI) is a composite parameter that summarizes the industrial performance of cotton fiber by integrating UHML, UI, Str, Mic, Rd, and +b into a single index. Therefore, high SCI values are generally associated with longer, more uniform, and stronger fibers, with adequate fineness and color characteristics, ultimately reflecting superior spinning performance (IMAmt, 2018).

A comparison between fiber quality rankings and spectral indices (SI, NGRDI, and GLI) revealed partial agreement between spectral responses and fiber technological performance. Genotypes with superior fiber quality, such as DP 1857, IMA 243, and TMG 44, exhibited relatively high NGRDI (0.1061–0.1162) and GLI (0.1245–0.1359) values, suggesting that greater green reflectance may be associated with improved fiber quality.

The association between GLI and SCI is biologically plausible because fiber quality results from the interaction of physiological, genetic, and environmental factors. Vegetation indices such as GLI and NGRDI are closely related to chlorophyll content, photosynthetic activity, and canopy vigor ^[18] (Luna Neto et al., 2023), whereas fiber technological properties depend on carbon assimilation and its allocation during boll development.

In cotton, source–sink relationships, particularly those involving the subtending leaf, directly influence carbohydrate supply for fiber elongation, secondary cell wall thickening, and cellulose deposition (Liu et al., 2015). Higher chlorophyll content and biomass accumulation generally promote

better plant development (Fait et al., 2025). However, assimilate-use efficiency, hormonal regulation, and genetic control of fiber development also strongly affect fiber technological properties. Consequently, genotypes with similar spectral responses may exhibit contrasting fiber quality (Basra and Malik, 1984).

This behavior was evident for DP 2077, which exhibited the best overall technological performance despite presenting intermediate NGRDI (0.0975) and GLI (0.1312) values. Likewise, some genotypes with inferior fiber quality displayed spectral responses similar to those of superior genotypes, indicating that RGB-derived indices alone cannot completely discriminate against high-performing cotton genotypes.

To better understand the relationship between spectral indices and fiber quality, phenotypic and genotypic correlation analyses were performed (Table 3). The strongest genotypic association was observed between GLI and SCI ($r = 0.71$), while GLI also showed significant positive correlations with UI ($r = 0.99$) and a moderate negative correlation with SF ($r = -0.56$).

Table 3. Phenotypic (upper diagonal) and genotypic (lower diagonal) correlation coefficients between fiber quality traits and RGB-derived spectral indices in 19 cotton genotypes evaluated in Uberlândia, Minas Gerais, Brazil.

	SCI	UHML	UI	SF	Str	Elg	SCId	SI	NGRDI	GLI
SCI		0.73**	0.85**	-0.87**	0.64**	-0.18	-0.31	0.01	0.33	0.51*
UHML	0.74**		0.54*	-0.84**	0.18	-0.13	-0.05	-0.18	-0.47*	0.27
UI	0.93**	0.61*		-0.81**	0.36	-0.12	-0.48*	-0.15	0.50*	0.60**
SF	-0.92**	-0.88**	-0.93**		-0.33	0.17	0.17	-0.14	-0.19	-0.40
Str	0.65*	0.14	0.49	-0.35		-0.15	-0.15	-0.42	0.14	-0.11
Elg	-0.11	-0.06	-0.05	0.11	-0.09		-0.15	-0.42	0.14	-0.11
SCId	-0.45	-0.09	-0.93**	0.29	0.08	-0.20		0.74**	-0.99**	-0.85**
SI	-0.06	0.20	-0.54	-0.08	0.25	-0.58*	0.82*		-0.75**	-0.31
NGRDI	0.48	0.10	0.96**	-0.31	-0.04	0.18	-0.99**	-0.85*		0.84**
GLI	0.71*	0.36	0.99**	-0.56*	0.12	-0.13	-0.93**	-0.56	0.90**	

**, * and indicate significance at 5% and 1% probability, respectively. UHML, mean fiber length; UI, fiber length uniformity; Str, fiber strength; Mic, micronaire; Elg, fiber elongation; +b, yellowness; Mat, fiber maturity; SCI, spinning consistency index; SF, short fiber index; SCId, Spectral Color Index; SI, Saturation Index; NGRDI, Normalized Green Red Difference Index; GLI, Green Leaf Index; H², genotypic determination coefficient; CVg, genotypic variation coefficient.

In addition, UI exhibited strong genotypic correlations with SCId ($r = -0.93$) and NGRDI ($r = 0.96$). As expected, SCI was positively correlated with UHML ($r = 0.74$), UI ($r = 0.93$), and Str ($r = 0.65$),

while showing a strong negative correlation with SF ($r = -0.92$). These relationships are consistent with the composite nature of SCI, which integrates several technological fiber attributes into a single index.

The positive association between GLI and SCI reinforces the potential of GLI as an indirect predictor of fiber quality. In plant breeding, correlated traits are valuable because indirect selection can substantially improve breeding efficiency when the secondary trait is easier, faster, or less expensive to measure (Loro et al., 2024).

To further investigate these relationships while minimizing environmental effects, best linear unbiased predictions (BLUPs) were estimated for SCI, UI, GLI, and NGRDI. These predicted genetic values were subsequently used to fit simple linear regression models between spectral indices and fiber quality traits, followed by leave-one-out cross-validation (LOOCV) owing to the limited number of evaluated genotypes (Table 4).

Table 4. Simple linear regression models relating GLI and NGRDI to fiber quality traits based on BLUP estimates.

Equation	β	EP(β)	R^2 (%)	R^2_{LOOCV} (%)	RMSE _{LOOCV}	p-value
SCI = 805,48 $\times$ GLI	805.48	238.96	46.6	35.4	5.8281	0.0050
UI = 55,61 $\times$ GLI	55.61	17.77	43.0	32.4	0.4291	0.0080
SF = -67,53 $\times$ GLI	-67.53	26.78	32.9	20.0	0.6523	0.0255
UI = 52,14 $\times$ NGRDI	52.14	15.98	45.0	34.1	0.4237	0.0062

GLI explained 35.4% of the cross-validated variation in SCI ($R^2_{LOOCV} = 35.4\%$), with an RMSE of 5.83, whereas NGRDI explained 34.1% of the variation in UI. These results indicate that both indices exhibit moderate predictive ability for cotton fiber quality. The consistency of these associations was confirmed by LOOCV, which yielded cross-validated coefficients of determination ranging from 32% to 35%.

The relationships between observed and predicted BLUP values are shown in Figure 2. Some genotypes were accurately predicted, whereas others, particularly DP 2077, DP 1857, DP 2176, Fabrika, and FM 985, showed larger prediction errors. Overall, prediction accuracy was

moderate, with correlation coefficients ranging from 0.57 for the UI–NGRDI model to 0.60 for the SCI–GLI model.

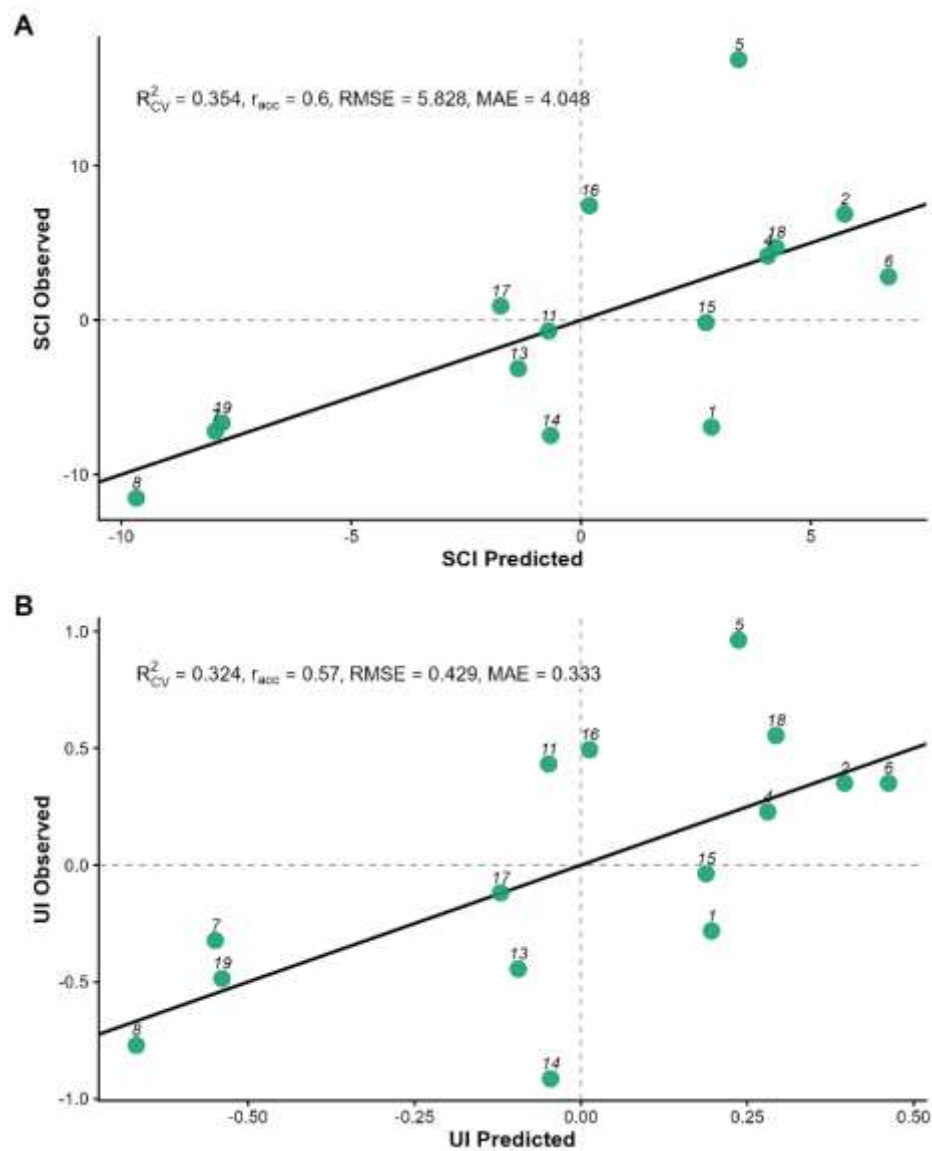
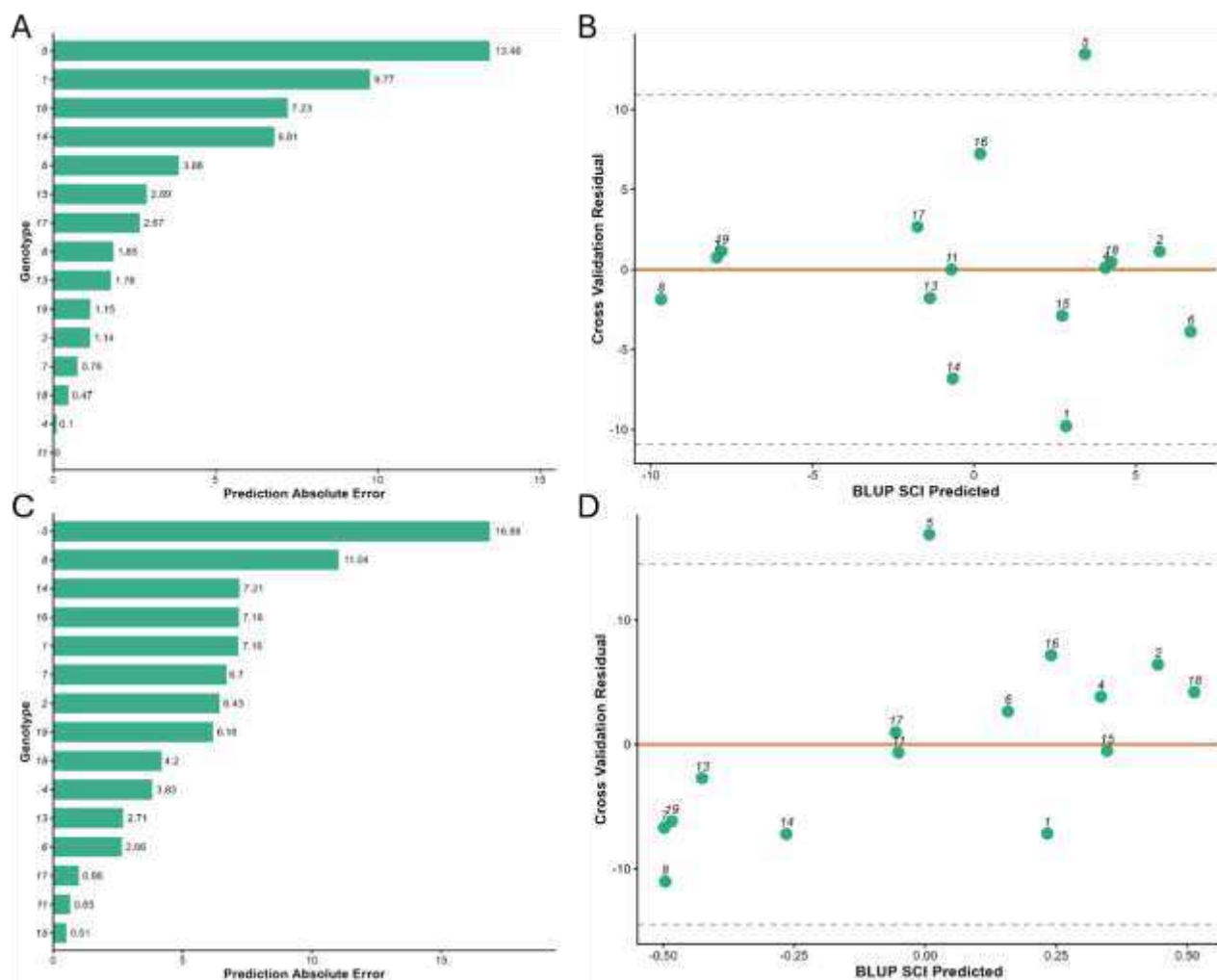


Figure 2. Linear regression between the BLUP-predicted genotypic values plus the overall mean for the Green Leaf Index (GLI) and the Spinning Consistency Index (SCI) (A), and between the Normalized Green Red Difference Index (NGRDI) and fiber length uniformity (UI) (B). Genotypes: 1, BRS 286; 2, DP 1857; 3, DP 1866; 4, DP 1949; 5, DP 2077; 6, DP 2176; 7, DP 555; 8, Fabrika; 9, FM 911; 10, FM 912; 11, FM 942; 12, FM 944; 13, FM 970; 14, FM 985; 15, IMA 1327; 16, IMA 243; 17, TMG 31; 18, TMG 44; and 19, TMG 82. Uberlândia, Minas Gerais, Brazil.

275

276 These results suggest localized prediction bias and heteroscedasticity for specific genotypic
 277 groups, probably reflecting the limited genetic diversity represented in the training dataset. Nevertheless,
 278 residual analyses indicated an approximately random distribution around zero, with no evidence of
 279 systematic bias. Although DP 2077, BRS 286, IMA 243, and FM 985 showed larger residuals, most
 280 genotypes were satisfactorily predicted by the fitted models (Figure 3).



281

282 **Figure 3.** Absolute prediction error and cross-validation residuals for the BLUP-predicted values of the
 283 evaluated cotton genotypes for the Spinning Consistency Index (SCI) and fiber length uniformity (UI).
 284 (A) Absolute prediction error for SCI; (B) cross-validation residuals plotted against the BLUP-predicted
 285 values for SCI; (C) absolute prediction error for UI; and (D) cross-validation residuals plotted against the
 286 BLUP-predicted values for UI. Dashed lines indicate the residual limits, whereas the solid horizontal line
 287 represents a residual value of zero.

288

289 Conversely, Fabrika, DP 555, and TMG 82 exhibited small prediction errors, indicating that the
290 models adequately captured their phenotypic behavior. These findings reinforce that expanding the
291 training population is essential to improve model robustness and increase prediction reliability.

292 Future studies should therefore (i) include a larger and more genetically diverse training
293 population encompassing contrasting fiber quality profiles, (ii) evaluate prediction models incorporating
294 pedigree information or genotype $\times$ environment interactions, and (iii) adopt confidence intervals as an
295 additional criterion for selection decisions, thereby reducing the risk of discarding superior genotypes
296 based on uncertain predictions.

297 Despite their moderate predictive performance, GLI and NGRDI successfully captured part of the
298 genetic variation associated with SCI and UI, respectively. Although these indices cannot replace
299 conventional HVI analyses, they represent promising complementary tools for large-scale phenotyping
300 and early genotype screening in cotton breeding programs, provided that their predictive ability is validated
301 across broader genetic backgrounds and multiple environments.

302 From an economic perspective, HVI fiber-quality analyses cost approximately R\$ 1.60–2.00 per
303 sample for large processing volumes (ABAPA, 2019). In breeding programs evaluating thousands of
304 experimental lines over multiple years, the early elimination of genotypes with low fiber-quality potential
305 could substantially reduce phenotyping costs, particularly during advanced breeding stages and Value for
306 Cultivation and Use (VCU) trials.

307 The RGB-derived spectral indices GLI and NGRDI, obtained from UAV imagery acquired after
308 the full-flowering stage, demonstrated considerable potential for large-scale phenotyping and preliminary
309 genotype selection in cotton breeding programs. Their strong genotypic associations with the SCI and UI
310 enabled the early discrimination of genotypes with inferior fiber quality, highlighting their usefulness as
311 indirect selection criteria.

The association between GLI and SCI may be related to differences in canopy vigor and photosynthetic activity during flowering, which can influence carbohydrate allocation to developing fibers. Similarly, NGRDI may indirectly reflect canopy physiological status associated with fiber development and maturation processes (Taiz et al., 2018)

Our results are consistent with those reported by Xu et al. (2023), who obtained substantially higher predictive accuracies ($R^2 = 0.825, 0.801$ and 0.772 for UHML, UI and Mic, respectively) using RGB and multispectral UAV imagery combined with machine learning and boll segmentation. However, their predictions were performed near harvest, when fiber traits were already expressed in developing bolls. In contrast, the present study aimed to identify genotypes with superior fiber-quality potential at earlier developmental stages, representing a more challenging prediction scenario.

Conclusions

However, despite their moderate predictive performance, these spectral indices do not replace conventional HVI analyses, which remain indispensable for the comprehensive assessment of fiber technological properties. The regression models explained approximately 32–35% of the variation in SCI and UI and achieved moderate prediction accuracies (0.57 – 0.60), indicating that their application should be considered complementary rather than substitutive.

Future studies should evaluate broader genetic populations across multiple environments and crop developmental stages to improve model robustness and to better understand the physiological mechanisms linking canopy spectral responses to fiber quality traits. Such advances will contribute to the development of reliable, low-cost phenotyping tools for early selection in cotton breeding programs.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable. This manuscript does not contain data from any individual person, images, or videos.

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

GAF, FIM, ILAD, ALGC, MSS, MMC, APON and LBS conceived and designed the experiment, conducted the field experiments, and collected the fiber quality data. GAF, FIM and LBS performed the statistical analyses and regression modeling using R software. GAF, FIM, APON, MSS and LBS 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

The authors thank Minas Cotton (AMIPA laboratory, Uberlândia-MG) for the fiber quality analysis, and the members of PROMALG for their support and collaboration. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

358 Associação Brasileira de Produtores de Algodão (ABRAPA). Abapa reduz preço da amostra
359 analisada da fibra de algodão para os associados. Brasília; 2019 [citado 2026 jul 10].
360 Disponível em: abapa.com.br

361 Basra AS, Malik CP. Development of the cotton fiber. *Int Rev Cytol.* 1984;89:65–113.
362 doi:10.1016/S0074-7696(08)61300-5.

363 Belot JL. Indústria têxtil, mercado mundial e qualidade de fibra para o futuro. In: Belot JL,
364 editor. Manual de qualidade da fibra da AMPA. Cuiabá: IMAmt–AMPA; 2018. p. 116–138.

365 Carvalho LP, Farias FJC, Santos RG, Teodoro LPR, Teodoro PE. Genotype selection for fiber
366 quality traits in cotton in the Brazilian Northeast. *Agron J.* 2022;114:3068–3073.
367 doi:10.1002/agj2.21176.

368 Cruz CD. GENES: a software package for analysis in experimental statistics and quantitative
369 genetics. *Acta Sci Agron.* 2013;35(3):271–276. doi:10.4025/actasciagron.v35i3.21251.

370 Fait MER, Rosa M, Marani IKH, Ferro RCAM, Cabral JSR, Rosa RR, et al. Espectros de luz
371 fazem a diferença: o algodoeiro sob LED e luz natural. *Cad Pedagógico.* 2025;22:e18476.
372 doi:10.54033/cadpedv22n9-357.

373 Feng A, Zhou J, Vories ED, Sudduth KA, Zhang M. Yield estimation in cotton using UAV-
374 based multi-sensor imagery. *Biosyst Eng.* 2020;193:101–114.
375 doi:10.1016/j.biosystemseng.2020.02.014.

376 Herrmann Júnior PSP, Bernardi AC de C, Assunção BSB de, Vaz CMP, Mantovani EC,
377 Yassitepe JE de CT, Simões MG, Leite MA de A, Ramos MLG, Inamasu RY, Araujo RF,
378 Oliveira RP de, Santos TT, Ribeiro Junior WQ. 2024. Fenotipagem de plantas de alto
379 rendimento, genotipagem, agricultura de precisão e agricultura digital para uma agricultura
380 sustentável e produtiva. p. 94-102. In: *Agricultura de Precisão*. Editora Cubo.
381 <https://doi.org/10.4322/978-65-86819-38-0.1000087>

382 Instituto Mato-grossense do Algodão (IMAmt). Manual da qualidade – Parte 2: Fibra de
383 algodão. Cuiabá: IMAmt; 2018 [citado 2026 jul 25]. Disponível em: imamt.org.br

384 Klein LM, Etchart VJ, Tcach MA, Gonzales AJ, Dileo PN, Tcach NE, et al. Variabilidad
385 morfológica y molecular en genotipos de algodón (*Gossypium hirsutum* L.) de tres orígenes
386 geográficos diferentes. *Cienc Agronómicas.* 2022;40:e025. doi:10.35305/agro40.e025.

387 Liu J, Meng Y, Lv F, Chen J, Ma Y, Wang Y, et al. Photosynthetic characteristics of the
388 subtending leaf of cotton boll at different fruiting branch nodes and their relationships with
389 lint yield and fiber quality. *Front Plant Sci.* 2015;6:747. doi:10.3389/fpls.2015.00747.

390 Loro MV, Cargnelutti Filho A, Ortiz VM, Reis MB, Andretta JA, Schuller BR. Relações
391 lineares entre caracteres do pendão e da espiga em bases genéticas de milho. *J Environ Anal*
392 *Progress.* 2024;9(2):65–78. doi:10.24221/jeap.9.2.2024.6208.065-078.

393 Luna Neto EV, Silva AV, Silva JHB, Silva CM, Costa TRS, Mielezrski F. Índices espectrais de
394 vegetação no monitoramento do milho (*Zea mays* L.) submetido à bioestimulantes. *Nativa.*
395 2023;11:323–330. doi:10.31413/nat.v11i3.13779.

396 Matias FI, Caraza-Harter MV, Endelman JB. FIELDimageR: An R package to analyze
397 orthomosaic images from agricultural field trials. *Plant Phenome J.* 2020;3:e20005.
398 doi:10.1002/ppj2.20005.

399 Resende EL, Bruzi AT, Cardoso ES, Carneiro VQ, Souza VAP, Barros PHFC, et al. High-
400 throughput phenotyping: Application in maize breeding. *AgriEngineering.* 2024;6(2):1078–
401 1092.

402 Sayeed MA, Schumann M, Wanjura JD, Kelly B, Smith W, Hequet EF. Characterizing the
403 total within-sample variation in cotton fiber length using the High Volume Instrument
404 fibrogram. *Text Res J.* 2021;91(2):175–187. doi:10.1177/0040517520935212.

405 Taiz L, Zeiger E, Moller IM, Murphy AS. *Plant physiology and development.* 6th ed. New
406 York: Sinauer Associates; 2018. □ United States Department of Agriculture (USDA). *Cotton*
407 *Outlook.* Washington: USDA; 2026 [citado 2026 mar 9]. Disponível em: usda.gov

408 Vidal Neto FC, Freire EC. Melhoramento genético do algodoeiro. In: Vidal Neto FC,
409 Cavalcanti JJV, editores. *Melhoramento genético de plantas no Nordeste.* Brasília: Embrapa;
410 2013. p. 49–83.

411 Xu W, Yang W, Chen P, Zhan Y, Zhang L, Lan Y. Cotton fiber quality estimation based on
412 machine learning using time series UAV remote sensing data. *Remote Sens.* 2023;15(3):586.
413 doi:10.3390/rs15030586.

414 Yan W. A systematic narration of some key concepts and procedures in plant breeding. *Front*
415 *Plant Sci.* 2021;12:724517. doi:10.3389/fpls.2021.724517.