Asian Journal of Agriculture and Biology
https://doi.org/10.35495/ajab.2025.297

Original Article

Generating near-infrared imagery of field rice using only UAV visible-light RGB camera

Geng Wei[1,2], Mulan Zou[1,2], Shuyue Wang[3], Bo Liu[1,2*]
[1]School of Surveying and Geoinformation Engineering, East China University of Technology, Nanchang, 330013, China
[2]Jiangxi Key Laboratory of Watershed Ecological Process and Information (Platform No. 2023SSY01051), East China University of Technology, Nanchang, 330013, China
[3]Nanjing Real Estate Registration, Nanjing, 210001, China

**Corresponding author's email:* liubo@ecut.edu.cn
Received: 05 November 2025 / Revised: 01 March 2026 / Accepted: 09 March 2026 / Published Online: 14 March 2026

Abstract

Generating Near-Infrared (NIR) imagery from RGB spectrum offers a low-cost alternative to dedicated multispectral sensors. To investigate whether NIR imagery can be generated from standard visible-light RGB camera, visible-light RGB and multispectral NIR images over a rice field were captured by an unmanned aerial vehicle (UAV). Color space features (including HSV and CIELAB) and texture features (TF) were incorporated, and optimal model inputs were identified through feature screening method. Subsequently, four distinct models were developed for NIR image generation. Results showed that performance of generating NIR images improves substantially when HSV, CIELAB, and TF were included, and further gains were obtained after input feature selection. Pix2Pix achieved the best performance on the test dataset, with the highest determination coefficient (R^2) of 0.78 and the lowest normalized Root Mean Square Error (nRMSE) of 6.41%, and the generated NIR images reached the highest Structural Similarity (SSIM) of 0.85 and Peak Signal-to-Noise Ratio (PSNR) of 28.48 dB. Moreover, feature importance analysis highlighted V, a, b, red-band contrast, green-band mean and variance as key predictors for NIR image generation. This study demonstrates a practical, low-cost approach to produce NIR imagery from standard visible-light RGB cameras, potentially reducing reliance on dedicated multispectral sensors.

Keywords: NIR generation, UAV, RGB spectra, Color space model, Texture feature

How to cite this article:
Wei G, Zou M, Wang S and Liu B. Generating near-infrared imagery of field rice using only UAV visible-light RGB camera. Asian J. Agric. Biol. 2026: e2025297. DOI: https://doi.org/10.35495/ajab.2025.297

Introduction

Faced with a growing global population and increasingly variable climate, Precision Agriculture (PA) has become essential for safeguarding food security (Crookston, 2006). Spectral data of crops — particularly near-infrared (NIR) information — provide a vital window into plant health and dynamics, and Unmanned Aerial Vehicle (UAV) offers a flexible, high-resolution platform for acquiring such data (Sophie et al., 2017). However, a practical trade-off of sensors exists: visible-light (RGB) cameras are inexpensive and offer high spatial resolution but lack direct sensitivity to chlorophyll, whereas multispectral (MS) sensors capture chlorophyll-sensitive NIR bands and thus support more robust vegetation indices (Zhou et al., 2022; Su et al., 2024). Higher cost and operational complexity, however, often put MS sensors beyond the reach of many farmers and smallholders.

This accessibility gap has driven research into methods for generating NIR information from RGB sensors. First approach is hardware modifications, such as specialized filter arrays or internal filter removal, proved technically niche and risky for the equipment (Lu et al., 2009; Rabatel et al., 2011). A simpler, data-centric path exploits statistical correlations between RGB and NIR bands, though early demonstrations based on limited data and linear models suggested limited generalizability (Kohei et al., 2016). More recently, Machine learning (ML) techniques offer a pragmatic balance between improved accuracy and manageable computational cost (de Lima et al., 2019; Lucas et al., 2020; Lavinia et al., 2024). Deep learning (DL), particularly models based on conditional Generative Adversarial Networks (cGAN) like Pix2Pix, pushes performance further at the expense of greater complexity and resource demands (Davidson et al., 2022; Farooque et al., 2023), and several enhanced GAN variants have also been proposed (Suárez et al., 2021; Zhang et al., 2022; Zhao et al., 2024) to generate NIR imagery through RGB sensors.

A common limitation in much of those works is that the RGB inputs are derived from multispectral sensors rather than from standard visible-light sensors. In practice, models are frequently trained by narrow-band RGB inputs extracted from MS sensors, and later applied to broad-band RGB images captured by visible-light sensors to produce so-called "visible-light based" NIR outputs. Although such models report promising results, this pipeline may risk a spectral domain mismatch: standard visible-light sensors lack the narrow-band spectral resolution of MS sensors, so the statistical relationship (from RGB to NIR) learned on MS-derived RGB may not hold for RGB captured by real visible-light sensors. Consequently, model performance and robustness in practical deployments may be overestimated when using MS-derived RGB information as inputs. The impact of this training–application mismatch on synthesis quality and generalization has received limited attention and therefore warrants focused investigation.

In this study, we acquired spectral images of field rice across key growth stages using a UAV, and we emphasize that the RGB images captured from standard visible-light camera rather than multispectral sensors were set as inputs to generate NIR images. The concrete aims were to:

(1) evaluate the performance of NIR generation through only visible-light RGB information and the necessity of introducing additional variables as inputs;

(2) identify the optimal input variables for each growth stage of field rice through a variable-selection method, and test whether feature screening could improve NIR generation performance;

(3) build and compare four candidate models, and determine the optimal model using different evaluation metrics;

(4) apply an interpretable variable-importance method to visualize and quantify the contribution of different predictors to NIR generation.

Material and Methods

Study area

Field experiments were conducted in Xinjizhen, Yizheng City, Jiangsu Province (119.3027°E, 32.3206°N). The rice cultivar used was Ningxiang 9. According to data from the local meteorological bureau (covering 2000–2020), the annual precipitation is approximately 1040 mm, the average annual temperature is 15.4°C, and the annual sunshine duration is about 2100 hours. We selected a rectangular plot (approximately 110 × 80 meters) for all measurements (Figure 1). This plot exhibits typical characteristics of surrounding farmland: mechanized operations, centralized seedling cultivation, precision irrigation, and green pest control, making it a convenient and representative sample site for this study.

https://doi.org/10.35495/ajab.2025.297

Figure-1. Study area.

Data collection

The rice spectral data were acquired with a DJI Mavic 3 Multispectral (Shenzhen, China). The UAV was equipped with a 20-megapixel visible camera and four 5-megapixel multispectral cameras (green, red, red-edge and near-infrared); the NIR sensor had a central wavelength of 860 nm ± 26 nm. Flight campaigns were carried out around noon (between 11 a.m. and 1 p.m.) from July to September 2025 on clear or partly cloudy days with low wind speeds (below 1.5 m/s). Missions flew at a 50 m altitude, with a horizontal speed of 5 m/s and a forward overlap of 75% between successive images. The rice growth stage corresponding to each flight mission is listed in Table 1.

Table-1. UAV flight mission and rice growth stage.

Flight date	Growth stage
July 26, 2025	Tillering
August 19, 2025	Jointing
September 9, 2025	Heading

RGB images were resampled using bilinear interpolation to maintain spatial resolution consistent with NIR imagery (2.5 cm/pixel). Manual verification was undertaken to ensure pixel-to-pixel alignment between RGB and NIR images. Multispectral calibration plates (with spectral reflectance values of 25%, 50%, and 75%) and DJI TERRA (Shenzhen, China) software were employed to calibrate the UAV-captured images. The Digital Number (DN) values in the raw images were corrected to Reflectance values (%).

Color space model and texture feature

To enhance the generation performance of NIR spectra, two color space models and eight texture features were incorporated into our study.

HSV model was designed to reflect the way people intuitively perceive and describe color (Smith, 1978). It comprises three components — hue (H), saturation (S) and value (V). H indicates the type of color, S measures color purity or vividness and V denotes brightness. Due to its ability to better separate luminance and chrominance components, HSV is widely used in vegetation coverage and vegetation segmentation (Yan et al., 2019). By separating luminance from chromaticity, CIELAB more closely approximates the human eye's perception of color differences (Mojtaba et al., 2025). The three axes of CIELAB are Lightness (L), a (green–magenta axis) and b (the blue–yellow axis). Compared to YCbCr and YUV, which are used for video signal transmission and compression, CIELAB performs better in material

https://doi.org/10.35495/ajab.2025.297

classification (Bello-Cerezo et al., 2016). In this study, HSV and CIELAB images were derived from the RGB images acquired by the UAV's standard visible-light camera.

Texture features (TF) capture crucial visual information in an image, representing the spatial arrangement of objects and their interaction with the environment. Gray-Level Co-occurrence Matrix (GLCM), proposed by Haralick et al. (1973), is one of the most widely used methods. A total of eight texture features were calculated across the three bands of RGB images: Angular Second Moment (ASM), Contrast (Con), Correlation (Cor), Dissimilarity (Dis), Entropy (Ent), Homogeneity (Hom), Mean, and Variance (Var). These resulted in 24 combined texture features (the suffixes '_R', '_G' and '_B' denote the red, green and blue bands, respectively; for example, 'Mean_R' represents the red-band Mean, 'Dis_G' represents the green-band Dissimilarity and 'Ent_B' represents the blue-band Entropy). GLCM parameters adopted in this study include a window size of 3 pixels, a direction set to 0° and a stride of 1, thereby maximally preserving the spatial resolution and maintaining the same dimensions as the original image (Liu et al., 2023).

Table-2. Texture features used in this paper.

Texture Feature	Description	Formula (Liang et al., 2022)
Angular Second Moment	Sum of Squares of Matrix Elements	$\mathrm{ASM} = \sum_{i,j=1}^{G} (P_{i,j})^2$
Contrast	The degree of distribution of matrix values deviates from the main diagonal	$\mathrm{Con} = \sum_{i,j=1}^{G} (i-j)^2 P_{i,j}$
Correlation	The degree of linear dependence between the row elements and column elements of a matrix	$\mathrm{Cor} = \sum_{i,j=1}^{G} \frac{(i,j)P_{i,j} - \mu_x \mu_y}{\sigma_x \sigma_y}$
Dissimilarity	The expected value of the absolute difference in grayscale values within the matrix	$\mathrm{Dis} = \sum_{i,j=1}^{G} P_{i,j} \lvert i-j \rvert$
Entropy	Complexity and randomness of the matrix	$\mathrm{Ent} = -\sum_{i,j=1}^{G} P_{i,j} \log (P_{i,j})$
Homogeneity	The degree of concentration of matrix values near the main diagonal	$\mathrm{Hom} = \sum_{i,j=1}^{G} \frac{P_{i,j}}{1+(i-j)^2}$
Mean	Arithmetic mean of pixel grayscale values	$\mathrm{Mean} = \sum_{i,j=1}^{G} i P_{i,j}$
Variance	Variance of pixels	$\mathrm{Var} = \sum_{i,j=1}^{G} (i-u)^2 P_{i,j}$

Note: p (i, j) represents the value of the (i, j) th entry in GLCM; μ_x and μ_y represent the mean value of x and y rows; σ_x and σ_y represent standard deviations of x and y rows.

Variable selection

Using only one predictor is often too simplistic for crop prediction tasks, but throwing in every available variable risks overfitting (Cheng et al., 2025). To isolate the variables with the strongest predictive power, Recursive Feature Elimination with cross-validation (RFECV)—which evaluates model performance across folds and automatically identifies the optimal number of features—was adopted (Wu et al., 2022). In this study, we ran RFECV with 10-fold cross-validation and used mean squared error (MSE) as the selection metric. Partial Least Squares Regression (PLSR) was selected as the base estimator to reduce potential multicollinearity among the 33 input variables.

Model construction

Four models were employed in this study to build 'RGB to NIR' conversion: Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Pix2Pix.

RF was introduced by Breiman (2001) and is an ensemble learning method based on decision trees. It consists of many independently constructed trees that vote (classification) or average (regression) to make predictions. SVM is a supervised learning technique applicable to classification and to regression (Support Vector Regression, SVR). By mapping input data into a higher-dimensional feature space via a kernel function, SVR often achieves strong generalization and can also mitigate multicollinearity (Cheng et al., 2017). ANN is inspired by biological neural systems and is capable of learning hierarchical features to solve nonlinear problems efficiently (Salvador et al., 2025). Here we tuned ANN hyperparameters by random search combined with 5-fold cross-validation. GAN, as proposed by Goodfellow et al. (2014), primarily consist of a generator (G) and a discriminator (D). The generator learns to produce synthetic data that mimics real data, while the discriminator learns to distinguish between real and synthetic samples (Aggarwal et al., 2021). Pix2Pix is a prominent cGAN-based model specifically designed for image-to-image translation tasks (Isola et al., 2017). A summary of the hyperparameters for the four models is presented in Table 3.

Table-3. Hyperparameters summary of 4 models.

Model		Key parameters setting
RF		n_estimators = 100/200/500
SVR		Radial Basis Function kernel (C = 10, ε = 0.1 and gamma = scale) with a 5-fold CV to tune the hyperparameters
ANN		Number of hidden layers: 2–4 Neurons per layer: 64/128/256 Dropout rate: 0/0.2/0.4 Learning rate: 0.01/0.005/0.001 Activation function: ReLU/SELU Batch size: 16/32/64 Epochs: 50/100/ 200 Random search: 5-fold CV
Pix2Pix	Generator	Architecture: U-Net with 8 downsampling and 8 upsampling blocks Convolution (downsampling): 4×4, stride 2, Instance Normalization (except first block), LeakyReLU (slope=0.2) Transposed convolution (upsampling): 4×4, stride 2, Instance Normalization, ReLU Dropout: Optionally applied to first few decoder layers Skip connections: Between corresponding encoder and decoder layers Input/Output range: Scaled to [−1, 1]; final layer uses tanh activation
	Discriminator	Architecture: 70×70 PatchGAN

https://doi.org/10.35495/ajab.2025.297

	Convolutional layers : 5 layers Layer 1: Conv + LeakyReLU (slope=0.2) Layers 2–5: Conv + Instance Normalization + LeakyReLU Output: Patch-wise prediction map (Sigmoid applied depending on loss function)
Training	Optimizer: Adam ($\beta_1 = 0.5$, $\beta_2 = 0.999$) Initial learning rate: 2×10^{-4} Epochs: 200 (first 100 fixed LR, last 100 linear decay to 0)
Input Processing	Tiling strategy: 270 patches (18 × 15 grid) Patch size: 256 × 256 pixels Overlap strategy: 25% overlap, stride = 128 (used in both extraction and inference)

Note: All models were executed on a workstation equipped with an Intel Core i9-14900K CPU, an NVIDIA GeForce RTX 4090 GPU, and 64 GB of RAM. For evaluation, the dataset was randomly split, with 70% of the samples allocated for training and the remaining 30% held out as an independent test set.

Evaluation metrics

To evaluate how well the models generate NIR imagery, we used multiple complementary metrics: Pixel-wise accuracy is quantified by the coefficient of determination (R^2) and normalized Root Mean Square Error (nRMSE), which measure how well predicted values match reference values in terms of variance explained and average error magnitude, respectively; Image quality is assessed using Structural Similarity (SSIM) and Peak Signal-to-Noise Ratio (PSNR): SSIM captures local structural and perceptual similarity (luminance, contrast, structure) while PSNR, derived from mean squared error, summarizes the overall signal-to-noise level of reconstructions (reported in dB). The formulas of 4 metrics are given below:

$$\mathbf{R^2} = \mathbf{1} - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (1)$$

$$\mathbf{nRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}}{x_{max} - x_{min}} \times 100\% \quad (2)$$

$$\mathbf{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

$$\mathbf{MSE} = \frac{i}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \quad (4)$$

$$\mathbf{PSNR} = 20Log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \quad (5)$$

where x_i is the observed value for sample i, y_i r is the model prediction for that sample, and x_{max} and x_{min} are the maximum and minimum observed values, respectively; μ_x and μ_y denote the means of x and y respectively, σ_x and σ_y are their standard deviations, and σ_{xy} is the covariance; C_1 and C_2 are small constants included to stabilize the denominator; Log_{10} denotes the base-10 logarithm and MAX_f is the maximum possible signal value in the reference image and MSE is the mean squared error between the images.

Figure 2 summarizes the overall workflow of this study, covering data collection, model development and result evaluation.

https://doi.org/10.35495/ajab.2025.297

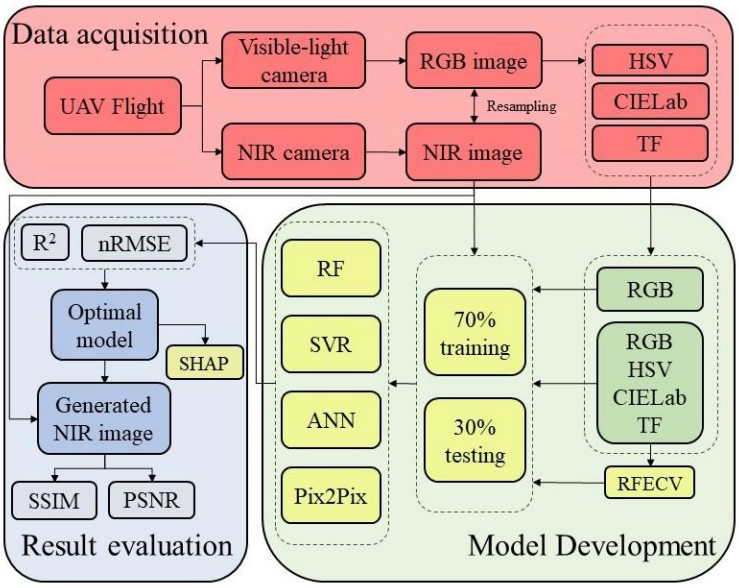

Figure-2. Workflow overview.

Results

Generating NIR images using only visible-light RGB spectra

First, NIR images were generated using only visible-light RGB images of rice fields at three growth stages as input. Table 4 shows the performance of different models on the test dataset. Pix2Pix model performed best in comparison, followed by ANN model, while RF and SVR showed comparable performance to each other. However, the overall performance of all four models was mediocre, with the highest R^2 of 0.56 and the lowest nRMSE of 9.20%. Therefore, using only visible-light RGB images as input for generating NIR images results in poor performance and necessitates considering additional input variables to enhance model accuracy.

Table-4. Test dataset validation results for models trained solely on visible-light RGB datasets.

Model	Evaluation Metrics	Tillering	Jointing	Heading
RF	R^2	0.48	0.46	0.45
	nRMSE(%)	10.04	10.68	10.92
SVR	R^2	0.47	0.44	0.46
	nRMSE(%)	10.23	10.41	10.43
ANN	R^2	0.52	0.49	0.50
	nRMSE(%)	9.81	9.97	9.86
Pix2Pix	R^2	**0.56**	**0.54**	**0.52**
	nRMSE(%)	**9.20**	**9.30**	**9.62**

Variable selection results via RFECV

To identify the input variables that contribute most to predicting the target variable (NIR), a 10-fold RFECV method was used to screen all input features independently across different rice growth stages. Figure 3 shows the MSE results for varying numbers of input variables at each growth stage, with the retained variables listed in Table 5. RFECV retained 20–24 variables from all input features. The variables retained across all three growth stages were: H, V, a, b, Con_R, Con_G, Cor_G, Cor_B, Mean_R, Mean_G, Mean_B, Var_R, Var_G and Var_B. Variables retained in two of the three growth stages included: L, Con_B, Cor_R, Dis_B, Hom_G and Hom_B. Among

the visible-light RGB spectral bands, only G and B were retained—each in only one growth stage (at the tillering and jointing stages, respectively). Therefore, according to RFECV, the raw visible-light RGB bands were considered "unimportant variables" compared to the derived color space and texture features.

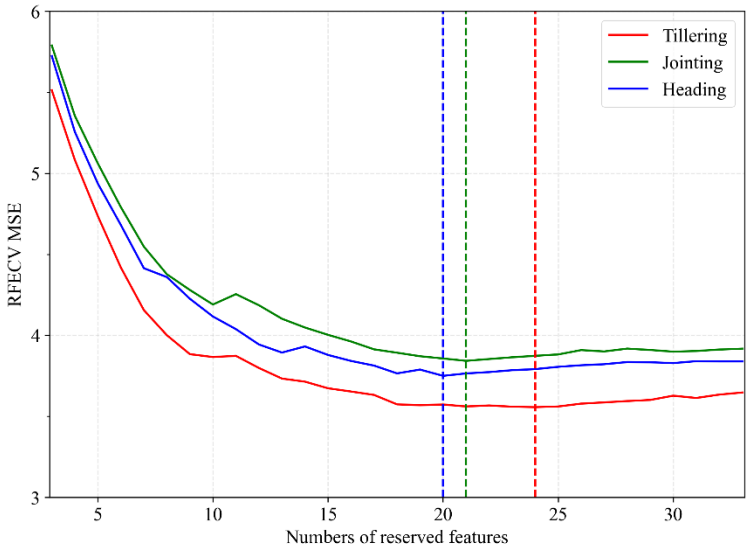

Figure-3. MSE with Different Numbers of Input Variables (dashed line indicates the number of variables corresponding to the minimum MSE value).

Table-5. Retained variables after RFECV screening during different rice growth stages.

Growing Stage	Number	Reserved features
Trilling	24	G, H, V, a, b, ASM_B, Con_R, Con_G, Con_B, Cor_R, Cor_G, Cor_B, Dis_R, Dis_B, Ent_R, Ent_B, Hom_R, Hom_B, Mean_R, Mean_G, Mean_B, Var_R, Var_G, Var_B
Jointing	21	B, H, V, L, a, b, Con_R, Con_G, Con_B, Cor_R, Cor_G, Cor_B, Dis_G, Ent_G, Hom_G, Mean_R, Mean_G, Mean_B, Var_R, Var_G, Var_B
Heading	20	H, S, V, L, a, b, ASM_R, Con_R, Con_G, Cor_G, Cor_B, Dis_B, Hom_G, Hom_B, Mean_R, Mean_G, Mean_B, Var_R, Var_G, Var_B

Model performance before and after variable selection

To assess how RFECV-based variable selection affected model performance, we first trained a set of NIR-generation models using the full variable dataset. The test dataset validation results are summarized in Table 6. Overall, adding color-space information and texture features produced clear improvements across all models. Pix2Pix achieved the best performance on the test dataset, with R^2 values of about 0.71–0.76 and nRMSE in the range 6.58%–7.56%. ANN ranked second, while SVR performed worst. When examined by growth stage, NIR generation was most successful at the tillering stage; performance declined during the jointing and heading stages.

https://doi.org/10.35495/ajab.2025.297

Table-6. Validation results of models trained with full-variety dataset as input on the test dataset.

Model	Evaluation Metrics	Tillering	Jointing	Heading
RF	R^2	0.65	0.58	0.60
	nRMSE(%)	8.10	9.62	9.14
SVR	R^2	0.58	0.56	0.55
	nRMSE(%)	9.73	9.81	9.93
ANN	R^2	0.68	0.66	0.65
	nRMSE(%)	7.82	8.21	8.26
Pix2Pix	R^2	**0.76**	**0.71**	**0.72**
	nRMSE(%)	**6.58**	**7.56**	**7.17**

The validation results on the test dataset using the RFECV-selected dataset as input for each model are shown in Table 7. It can be seen that R^2 values of all the models increased (by 0.01 to 0.05), and nRMSE values decreased (by 0.17% to 1.21%). Pix2Pix model achieved the best performance, with R^2 ranging from 0.72 to 0.78 and nRMSE from 6.41% to 7.19%. ANN model performed second best. Although SVR model showed significant improvement after variable selection, it remained the least effective model.

Table-7. Validation results of models trained with selected-variety dataset as input on the test dataset.

Model	Evaluation Metrics	Tillering	Jointing	Heading
RF	R^2	0.66	0.60	0.62
	nRMSE(%)	7.74	8.73	8.35
SVR	R^2	0.62	0.61	0.59
	nRMSE(%)	8.52	8.84	9.21
ANN	R^2	0.70	0.67	0.69
	nRMSE(%)	7.55	7.88	7.71
Pix2Pix	R^2	**0.78**	**0.72**	**0.74**
	nRMSE(%)	**6.41**	**7.19**	**6.98**

Variable importance assessment

The SHAP method was used to identify the 10 most important variables for NIR generation performance at different rice growth stages. The results are shown in Figure 4. Variables such as V, a, b, Con_R, Mean_G and Var_G were found to be highly important, ranking within the top 10 across all three growth stages.

https://doi.org/10.35495/ajab.2025.297

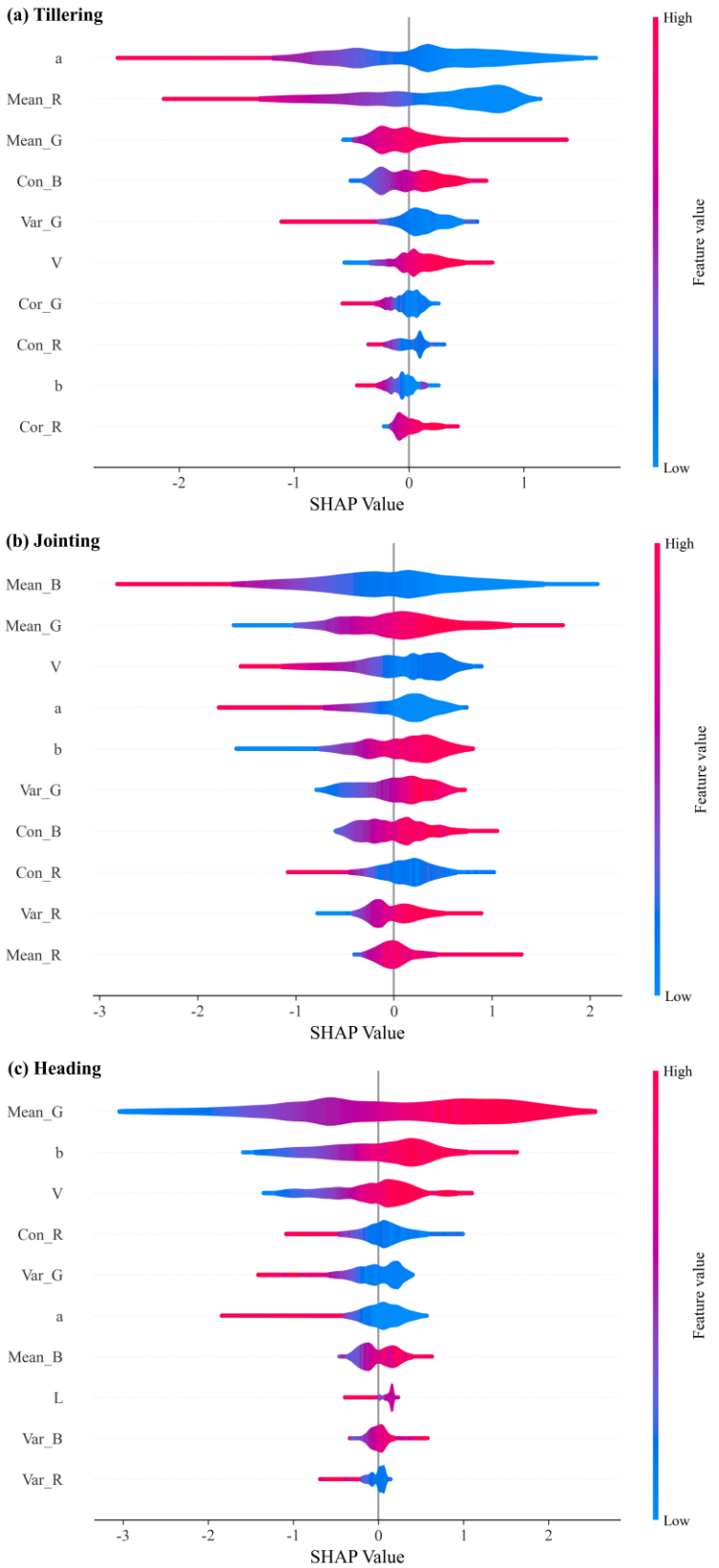

Figure-4. 10 most important variables for generating NIR across different growth stages.

Spatial distribution of generated NIR and difference

Since Pix2Pix model demonstrated the best performance on the test dataset before and after variable selection, NIR spatial distribution images were generated using Pix2Pix model. Figure 5 shows the true NIR images, the generated NIR images and the difference images between true and generated images for each rice growth stage. Figure 6 presents the distribution curve of the errors between predicted and actual NIR values. It can be observed that while Pix2Pix model produces good overall image quality, it exhibits some discrepancies in accurately predicting the extreme high and low values in the true images.

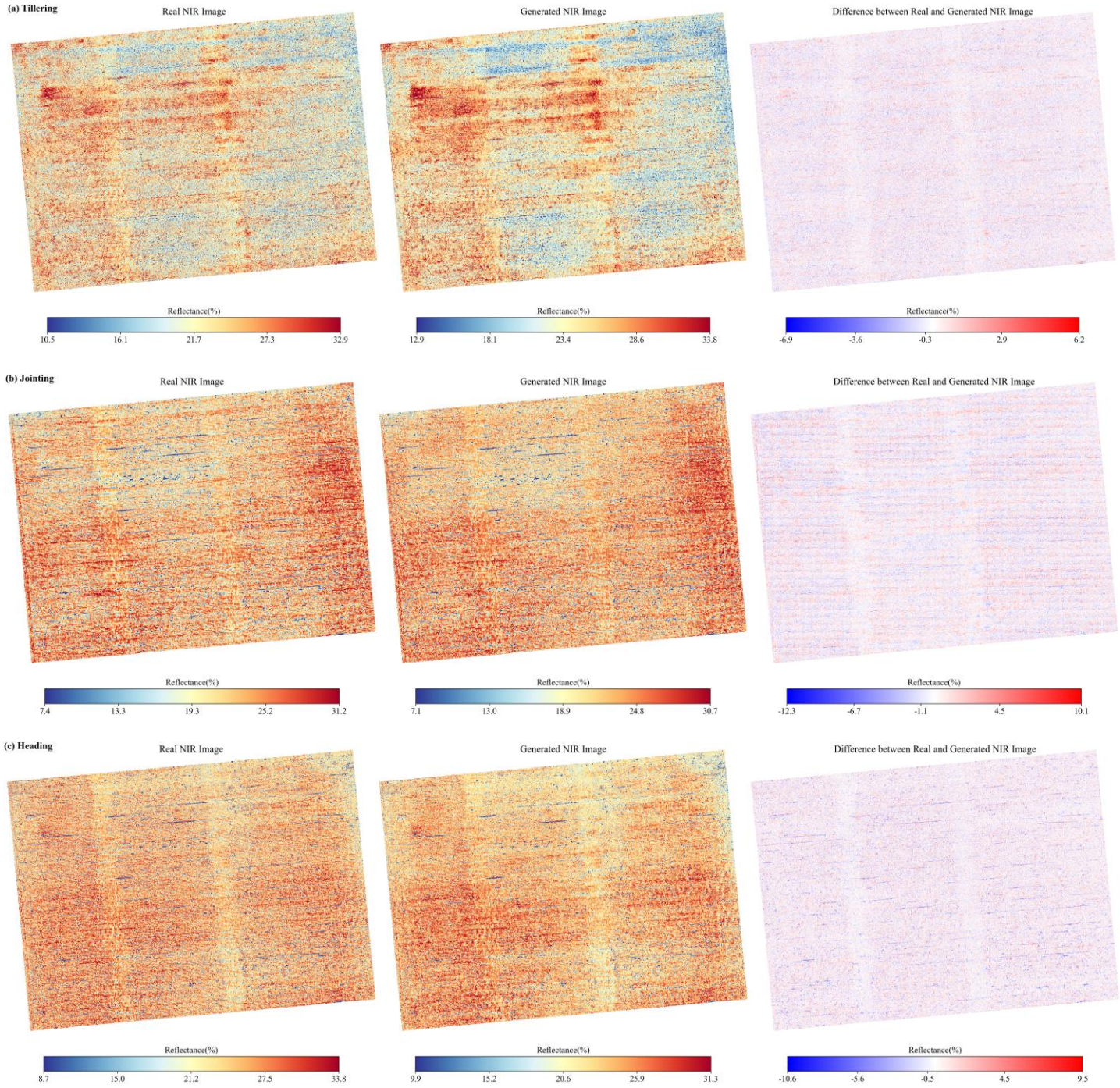

Figure-5. True NIR images, generated NIR images, and difference images for rice at different growth stages.

https://doi.org/10.35495/ajab.2025.297

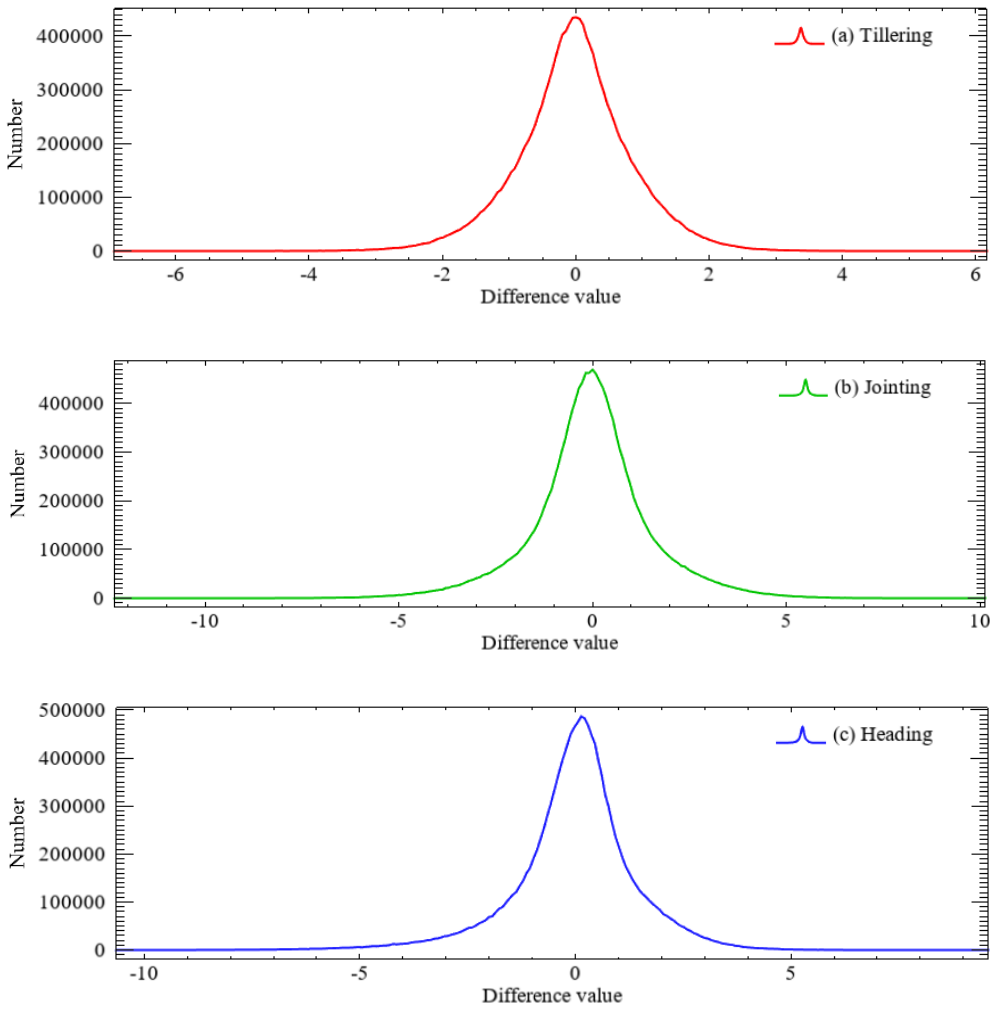

Figure-6. Distribution curve of the difference between actual and predicted NIR values.

Table 8 presents the SSIM and PSNR values of the NIR images generated by the Pix2Pix model at different rice growth stages, the results are generally consistent with the test dataset validation performance. The tillering stage achieved the highest SSIM and PSNR values of 0.85 and 28.48 dB, respectively, while the jointing stage exhibited the lowest values of 0.82 and 25.89 dB, respectively.

Table-8. SSIM and PSNR of generated NIR images at different growth stages.

Rice growth stage	SSIM	PSNR(dB)
Tillering	0.85	28.48
Jointing	0.82	25.89
Heading	0.83	26.12

Discussion

Generating NIR images from more readily available RGB images offers a low-cost, high-precision alternative to meet NIR imaging needs. However, current research on generating NIR images predominantly uses RGB images sourced from MS sensors. Though multispectral RGB images can generate high-quality NIR images, there are discrepancies between visible-light and multispectral

RGB spectra. Consequently, NIR spectrum generation models trained using multispectral RGB spectra as input may not perform optimally when visible-light RGB spectra are used as input for the model. This paper therefore introduces visible-light RGB images with two color space models (HSV and CIELAB) and eight texture features to generate corresponding NIR images.

Performance of visible-light RGB images in generating NIR images

Visible-light and multispectral sensors differ in their data acquisition mechanisms, which in turn affects the quality of their RGB outputs for NIR reconstruction. Specifically, consumer-grade RGB cameras are characterized by their mechanical simplicity and high spatial resolution. However, their R, G, and B channels exhibit broad response curves with significant spectral overlap, thereby limiting their spectral resolution capabilities. In contrast, multispectral cameras employ multiple dedicated optics, each optimized for narrower spectral bands. As Telmo et al. (2017) have noted, broad-band sensors often lack the sensitivity required to detect subtle variations in crop parameters. This limitation highlights the complementary strengths of the two sensor types: while RGB imagery excels in capturing spatial detail, it provides limited spectral information. Conversely, multispectral RGB captures finer reflectance variations directly linked to vegetation health and productivity. Furthermore, multispectral cameras, through radiometric calibration, can effectively mitigate interference from variations in solar illumination and soil reflectance, while also eliminating sensor-induced and illumination-related artifacts (Feng et al., 2020; Juan et al., 2022). As a result, attempts to generate NIR imagery using only the visible-light spectrum are inherently constrained by performance limitations. Our experimental results corroborate this finding: NIR images generated solely from visible RGB data achieved a highest R^2 of 0.56 and a minimum nRMSE of 9.20% on the test set. This significant accuracy gap underscores the necessity of incorporating additional input variables, such as color space transformations and texture features, to overcome the limitations of raw RGB data.

Contributions of color space information and texture features

To address the inherent spectral limitations of raw visible-light RGB data, we leveraged two complementary feature types: color space information and texture features.

HSV color space effectively decouples color attributes (hue and saturation) from brightness (value), thereby demonstrating robustness to illumination variations and sensitivity to subtle color differences (Esmael et al., 2017). The CIELAB color space further enhances this by providing a perceptually uniform representation, which is particularly advantageous for tasks such as color balance correction (Khan et al., 2018). Mechanistically, HSV can disentangle illumination information from color information, and the spectral images transformed into the CIELAB color space exhibit enhanced regularity and uniformity, enabling the extraction of NIR reflectance information that is strongly associated with the biochemical composition and water status of vegetation (Sun et al., 2025). As de Lima et al. (2022) demonstrated that utilizing these transformed color spaces as inputs, rather than raw RGB values, improved NIR reconstruction—a finding that corroborates our approach despite differences in source data (their RGB data originated from a hyperspectral sensor with channel wavelengths of 611nm, 560.2nm, and 453.8nm, respectively).

Complementing color space analysis, we incorporated texture features to quantify spatial patterns by measuring pixel variability within local windows (Zhang et al., 2020). Texture features offer distinct advantages, such as contrast is capable of extracting edge contour information from crops; correlation is effective in quantifying pixel relationships; disparity measures grayscale variations between pixels featuring edge characteristics; entropy assesses the complexity of grayscale distributions; homogeneity identifies pixel features distant from coherent edges through grayscale similarity analysis, whereas mean evaluates the uniformity of grayscale pixel distributions (Xu et al., 2023). These features collectively offer unique contributions to the generation of NIR, potentially enhancing its performance. These characteristics make texture features a powerful complement to spectral data, particularly for addressing limitations such as canopy saturation, as evidenced in studies on biomass estimation and disease detection (Arif et al., 2024; Hu et al., 2025).

The critical role of both feature types was validated through our variable selection process. Notably, raw

https://doi.org/10.35495/ajab.2025.297

RGB spectral bands were consistently ranked as low importance and subsequently excluded across all growth stages. In contrast, features derived from HSV, CIELAB, and texture analysis were consistently retained. This result underscores the importance of perceptual color information and texture features in constructing an effective bridge from visible-light RGB to NIR data. The structural and biochemical characteristics of rice canopies exhibit significant variations across different growth stages. Variable importance analysis results demonstrate that V, a and b, as well as the red-band contrast, the green-band mean and variance consistently rank among the top 10 most influential variables throughout all growth stages. The elevated importance of color space features underscores that the overall color and brightness information of the canopy serve as fundamental variables for predicting NIR reflectance. Notably, regardless of the growth stage, the chlorophyll content and canopy light conditions of rice remain critical factors influencing NIR reflectance. Furthermore, the persistent significance of texture features including Con_R, Mean_G, and Var_G underscores the essential role of canopy structural information throughout the entire growth cycle, a finding consistent with the previously discussed viewpoint that 'texture features can alleviate spectral saturation'.

Feature selection and model performance

Extensive research has demonstrated that variable screening effectively mitigates multicollinearity and minimizes unnecessary redundancy, thereby enhancing overall model performance and reducing computational demands (Cheng et al., 2025). Among various feature selection methodologies, RFE distinguishes itself through its iterative approach, wherein it refines feature sets by repeatedly training a model and discarding the least impactful features until a predetermined number of features—or another stopping criterion—is achieved. When integrated with cross-validation, forming RFECV, this process evaluates feature subsets across cross-validation folds and identifies the optimal number and combination of variables without necessitating manual intervention. This automated approach reduces selection bias stemming from overfitting, rendering RFECV particularly appealing for interpretable, performance-oriented feature selection. However, it is important to note that RFECV—especially when employing numerous cross-validation folds—can impose significant computational and training time requirements; thus, the choice of methodology should account for dataset complexity and available computational infrastructure.

In this investigation, we implemented four distinct models for NIR image generation. Pix2Pix consistently demonstrated superior performance, while ANN ranked second in terms of effectiveness. SVR, however, exhibited relatively weaker performance. A notable trend emerged: models with higher performance tended to exhibit greater complexity and required substantially longer training durations, a pattern particularly pronounced in the case of Pix2Pix. Across all four models, the cumulative training time approximated 300 hours, with Pix2Pix accounting for over half of this duration. ANN, being less complex than Pix2Pix, ranked second in both accuracy and computational cost. In contrast, RF proved to be the simplest and fastest to train. Although SVR incorporated an initial parameter sweep combined with 5-fold cross-validation during tuning, its test performance still fell short of that of RF in our experimental framework. In summary, model performance generally correlates with both model complexity and computational demands. For researchers operating within constrained computational environments, traditional machine learning approaches—such as random forests—may offer a more practical alternative, delivering robust performance with relatively modest computational requirements. Nonetheless, the appropriateness of a given model should be re-evaluated when applied to different data types.

Future Recommendations

We acknowledge that this study has certain limitations. Primarily, due to the lack of a blue-band multispectral camera in the DJI UAV used in this paper, cross-sensor comparisons cannot be conducted. Replacing the equipment could further quantify the differences in performance of different RGB sensors in generating NIR images. Secondly, research centered on a single location and a single rice variety tends to constrain the representativeness and generalizability of the findings. Therefore, it is imperative to undertake further comparative analysis on other natural environments or crops with markedly differing characteristics. Finally, data were solely collected during three critical growth stages in 2025, which may have resulted in insufficient analysis of phenotypic changes within the season and limited

generalization capabilities across different inter-annual climatic variations, cultivation varieties, and multi-site environments. Consequently, the applicability to longer time scales or other production areas needs to be further verified through data from multiple years, and varieties in future work.

Conclusion

In this study, we captured visible-light RGB and corresponding NIR images of field rice using UAV during the tillering, jointing, and heading stages, and investigated the generation of NIR imagery from RGB inputs. The key findings are summarized as follows:

(1) Generating NIR imagery solely from visible-light RGB inputs resulted in suboptimal performance. However, incorporating color space information (HSV and CIELAB) and texture features significantly enhanced the quality of the generated NIR images.

(2) A 10-fold RFECV was employed to select variables independently for each growth stage, with MSE serving as the selection criterion. Notably, visible-light RGB spectral bands were rarely retained and consistently classified by RFECV as "unimportant variables".

(3) Post-variable selection, all models exhibited improved performance. Among the four established models, Pix2Pix demonstrated superior performance, achieving the highest R^2 (0.78) and the lowest nRMSE (6.41%). The NIR images generated by Pix2Pix also displayed high SSIM (0.82–0.85) and PSNR (25.89–28.48 dB) values.

(4) SHAP analysis identified the 10 most influential variables contributing to NIR imagery generation. Among these, V, a, b, red-band contrast (Con_R), and green-band mean and variance (Mean_G and Var_G) were found to be relatively more important.

We believe that if our research can be implemented and adopted in actual agricultural production, particularly if UAVs equipped with low-cost visible-light sensors can generate real-time "RGB to NIR" information or images, it will help mitigate the challenges faced by smallholders and low-income agricultural workers due to limited resources and intensified competition.

Acknowledgement

The authors acknowledge Mr. Ming Zhang for providing the instrumentation and data acquisition, and Prof. Xuejun Liu for his technical guidance on this paper.

Disclaimer: None.

Conflict of Interest: The authors declare that they have no conflict of interest.

Source of Funding: This work was supported by the National Natural Science Foundation of China (No. 42161064).

Use of Generative AI Statement

The authors declare it that AI was only used to translate (from Chinese to English), check grammar, and help proofreading the original text. No content in this paper was generated by AI.

Contribution of Authors

Wei G: Writing – original draft, writing – review & editing, visualization, validation, methodology, data curation and conceptualization.

Zou M: Writing – original draft, writing – review & editing, supervision, methodology and conceptualization.

Wang S: Writing – review & editing, visualization.

Liu B: Writing – review & editing, supervision, funding acquisition, methodology and conceptualization.

References

Aggarwal A, Mittal M and Battineni G, 2021. Generative adversarial network: an overview of theory and applications. Int. J. Inf. Manage. Data Insights. 1(1): 100004. https://doi.org/10.1016/j.jjimei.2020.100004

Arif K, Lilik B, Sahid A, Gunardi S and Chiharu H, 2024. Textural features for BLB disease damage assessment in paddy fields using drone data and machine learning: Enhancing disease detection accuracy. Smart Agric. Technol. 8: 100498. https://doi.org/10.1016/j.atech.2024.100498

Bello-Cerezo R, Bianconi F, Fernández A, González E and Di Maria F, 2016. Experimental comparison of color spaces for material classification. J Electron Imaging. 25: 061406. https://doi.org/10.1117/1.JEI.25.6.061406.

Breiman L, 2001. Random Forests. Mach. Learn. 45: 5–32. https://doi.org/10.1023/A:1010933404324

Cheng K, Lu Z, Zhou Y, Shi Y and Wei Y, 2017. Global sensitivity analysis using support

vector regression. Appl. Math. Model. 49: 587-598. https://doi.org/10.1016/j.apm.2017.05.026

Cheng K, Yan J, Li G, Ma W, Guo Z, Wang W, Li H, Da Q, Li X and Yao Y, 2025. Remote sensing inversion of nitrogen content in silage maize plants based on feature selection. Front. Plant Sci. 16: 1554842. https://doi.org/10.3389/fpls.2025.1554842

Crookston R, 2006. A Top 10 List of Developments and Issues Impacting Crop Management and Ecology During the Past 50 Years. Crop Sci. 46: 2253–2262. https://doi.org/10.2135/cropsci2005.11.0416 gas

Davidson C, Jaganathan V, Sivakumar A, Czarnecki J and Chowdhary G, 2022. NDVI/NDRE prediction from standard RGB aerial imagery using deep learning. Comput. Electron. Agric. 203: 107396. https://doi.org/10.1016/j.compag.2022.107396

Esmael H, Brian Mc G, Martin G and Edward J, 2017. Automatic crop detection under field conditions using the HSV colour space and morphological operations. Comput. Electron. Agric. 133: 97-107. https://doi.org/10.1016/j.compag.2016.11.021

Farooque A, Afzaal H, Benlamri R, Al-Naemi S, MacDonald E, Abbas F, MacLeod K and Ali H, 2023. Red-green-blue to normalized difference vegetation index translation: a robust and inexpensive approach for vegetation monitoring using machine vision and generative adversarial networks. Precis. Agric. 24: 1097–1115. https://doi.org/10.1007/s11119-023-10001-3

Feng A, Zhou J, Vories E and Sudduth KA, 2020. Evaluation of Cotton Emergence Using UAV-Based Narrow-Band Spectral Imagery with Customized Image Alignment and Stitching Algorithms. Remote Sens. 12(11): 1764. https://doi.org/10.3390/rs12111764

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde D, Ozair S, Courville A and Bengio Y, 2014. NIPS'14: Proceedings of the 28th International Conference on Neural Information Processing Systems. 2: 2672 – 2680. https://doi.org/10.5555/2969033.2969125

Haralick R, Shanmugam K and Dinstein I, 1973. Textural Features for Image Classification. IEEE Trans. Syst. Man Cybern. SMC–3(6): 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Hu X, Yang C, Zhang M, Wu F, Wu B and Tian Y, 2025. Superpixel-refined deep learning framework with texture enhancement and multi-layer attention fusion for automatic crop detection in VGI cropland imagery. Comput. Electron. Agric. 238: 110746. https://doi.org/10.1016/j.compag.2025.110746

Isola P, Zhu J, Zhou T and Efros A, 2017. Image-to-image translation with conditional adversarial networks. Proceedings – 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA: 5967-5976. https://doi.org/10.1109/CVPR.2017.632

Juan M, Alfonso L, Luís P and Joaquim J, 2022. Remote sensing image fusion on 3D scenarios: A review of applications for agriculture and forestry. Int. J. Appl. Earth. Obs. 112: 102856. https://doi.org/10.1016/j.jag.2022.102856

Khan U, Akram T, Naqvi S, Haider S, Kamran M and Muhammad N, 2018. Automatic detection of plant diseases; utilizing an unsupervised cascaded design. 2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan: 339-346. https://doi.org/10.1109/IBCAST.2018.8312246.

Kohei A, Kenji G, Osamu S and Yuko M, 2016. Method for NIR Reflectance Estimation with Visible Camera Data based on Regression for NDVI Estimation and its Application for Insect Damage Detection of Rice Paddy Fields. Int. J. Adv. Res. Artif. Intell. 5(11): 17-22. http://dx.doi.org/10.14569/IJARAI.2016.051103

Lavinia M, Luciano O, Federico P, Simone F, Simona V, Catello P, Valerio C, Simone V and Corrado C, 2024. An open-source machine-learning application for predicting pixel-to-pixel NDVI regression from RGB calibrated images. Comput. Electron. Agric. 216: 108536.

https://doi.org/10.1016/j.compag.2023.108536

de Lima D, Saqui D, Ataky S, Jorge L, Ferreira E and Saito J, 2019. Estimating Agriculture NIR Images from Aerial RGB Data. Computational Science – ICCS 2019: 19th International Conference, Part I:562 – 574. https://doi.org/10.1007/978-3-030-22734-0_41

de Lima D, Saqui D, Mpinda S and Saito J, 2022. Pix2Pix Network to Estimate Agricultural Near Infrared Images from RGB Data. Can. J. Remote. Sens. 48(2): 299–315. DOI: https://doi.org/10.1080/07038992.2021.2016056

Liang Y, Kou W, Lai H, Wang J, Wang Q, Xu W, Wang H and Lu N, 2022. Improved estimation of aboveground biomass in rubber plantations by fusing spectral and textural information from UAV-based RGB imagery. Ecol. Indic. 142: 109286. https://doi.org/10.1016/j.ecolind.2022.109286

Liu J, Zhu Y, Song L, Su X, Li J, Zheng J, Zhu X, Ren L, Wang W and Li X, 2023. Optimizing window size and directional parameters of GLCM texture features for estimating rice AGB based on UAVs multispectral imagery. Front. Plant Sci. 14:1284235. https://doi.org/10.3389/fpls.2023.1284235

Lu Y, Fredembach C, Vetterli M and Süsstrunk S, 2009. Designing color filter arrays for the joint capture of visible and near-infrared images. 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 2009: 3797-3800. https://doi.org/10.1109/ICIP.2009.5414324

Lucas C, Leon N and Yiannis A, 2020. A new visible band index (vNDVI) for estimating NDVI values on RGB images utilizing genetic algorithms. Comput. Electron. Agric. 172: 105334. https://doi.org/10.1016/j.compag.2020.105334

Mojtaba N, Maja K, Stacey S, Prasad D and Asim B, 2025. Optimizing image-based soil organic matter prediction: Effects of illumination type and intensity. Smart Agric. Technol. 11: 100922. https://doi.org/10.1016/j.atech.2025.100922

Rabatel G, Gorretta N and Labbé S, 2011. Getting NDVI Spectral Bands from a Single Standard RGB Digital Camera: A Methodological Approach. In: Lozano, J.A., Gámez, J.A., Moreno, J.A. (eds) Advances in Artificial Intelligence. CAEPIA 2011. Lecture Notes in Computer Science 7023. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-25274-7_34

Salvador C, Munera S, Marcelino M, José B, Sergio C and Juan G, 2025. Artificial Neural Networks in Agriculture, the core of artificial intelligence: What, When, and Why. Comput. Electron. Agric. 230: 109938. https://doi.org/10.1016/j.compag.2025.109938

Smith A, 1978. Color gamut transform pairs. SIGGRAPH Comput. Graph. 12(3): 12-19. DOI: https://doi.org/10.1145/965139.807361

Sophie C, Rohan B, Markus G, Michael Y and George V, 2017. Contour Detection for UAV-Based Cadastral Mapping. Remote Sens. 9(2): 171. https://doi.org/10.3390/rs9020171

Su X, Nian Y, Shaghaleh H, Hamad A, Yue H, Zhu Y, Li J, Wang W, Wang H, Ma Q, Liu J, Li X and Alhaj HY, 2024. Combining features selection strategy and features fusion strategy for SPAD estimation of winter wheat based on UAV multispectral imagery. Front. Plant Sci. 15: 1404238. https://doi.org/10.3389/fpls.2024.1404238

Suárez P, Sappa A and Vintimilla B, 2021. Cycle Generative Adversarial Network: Towards A Low-Cost Vegetation Index Estimation. 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA: 2783-2787. https://doi.org/10.1109/ICIP42928.2021.9506746

Sun Y, Huang K, Zhang B, Zhang Y, Qi Y, Yuan Q, Lei X and Lv X, 2025. Improving estimation accuracy of canopy nitrogen content using the fusion of ground-space spectral imagery with machine learning in pear orchards. Comput. Electron. Agric. 239: 111075. https://doi.org/10.1016/j.compag.2025.111075

Telmo A, Jonáš H, Luís P, José B, Emanuel P, Raul M and Joaquim JS, 2017. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for

Agriculture and Forestry. Remote Sens. 9(11): 1110. https://doi.org/10.3390/rs9111110

Wu J, Zheng D, Wu Z, Song H and Zhang X, 2022. Prediction of Buckwheat Maturity in UAV-RGB Images Based on Recursive Feature Elimination Cross-Validation: A Case Study in Jinzhong, Northern China. Plants. 11(23): 3257. https://doi.org/10.3390/plants11233257

Xu T, Wang F, Shi Z, Xie L and Yao X, 2023. Dynamic estimation of rice aboveground biomass based on spectral and spatial information extracted from hyperspectral remote sensing images at different combinations of growth stages. ISPRS J. Photogramm. Remote Sens. 202: 169-183. https://doi.org/10.1016/j.isprsjprs.2023.05.021

Yan G, Li L, Coy A, Mu X, Chen S, Xie D, Zhang W, Shen Q and Zhou H, 2019. Improving the estimation of fractional vegetation cover from UAV RGB imagery by colour unmixing. ISPRS J. Photogramm. Remote Sens. 158: 23-34. https://doi.org/10.1016/j.isprsjprs.2019.09.017

Zhang C, Huang C, Li H, Liu Q, Li J, Bridhikitti A and Liu G, 2020. Effect of textural features in remote sensed data on rubber plantation extraction at different levels of spatial resolution. Forests. 11(4): 399. https://doi.org/10.3390/f11040399

Zhang Y, Yang W, Zhang W, Yu J, Zhang J, Yang Y, Lu Y and Tang W, 2022. Two-step ResUp&Down generative adversarial network to reconstruct multispectral image from aerial RGB image. Comput. Electron. Agric. 192: 106617. https://doi.org/10.1016/j.compag.2021.106617

Zhao G, He Y, Wang Z, Wu H and Cheng L, 2024. Generation of NIR Spectral Band from RGB Image with Wavelet Domain Spectral Extrapolation Generative Adversarial Network. Comput. Electron. Agric. 227(1): 109461. https://doi.org/10.1016/j.compag.2024.109461

Zhou J, Lu X, Yang R, Chen H, Wang Y, Zhang Y, Huang J and Liu F, 2022. Developing Novel Rice Yield Index Using UAV Remote Sensing Imagery Fusion Technology. Drones. 6(6): 151. https://doi.org/10.3390/drones6060151