

Bài báo khoa học

## Ước tính hàm lượng diệp lục trong lá lúa bằng chỉ số thực vật chiết xuất từ dữ liệu ảnh UAV đa phổ

Phạm Thị Làn<sup>1\*</sup>, Lê Văn Cảnh<sup>1</sup>

<sup>1</sup> Trường Đại học Mở - Địa chất; phamthilan@humg.edu.vn; levancanh@humg.edu.vn

\*Tác giả liên hệ: phamthilan@humg.edu.vn; Tel.: +84-983321882

Ban Biên tập nhận bài: 10/11/2024; Ngày phản biện xong: 27/12/2024; Ngày đăng bài: 25/5/2025

**Tóm tắt:** Việc ước tính hàm lượng diệp lục của lá là việc rất thiết yếu để giám sát tăng trưởng cây lúa, giúp quản lý phân bón nhằm nâng cao năng suất cây lúa. Trong bài báo này, hàm lượng diệp lục được ước tính bằng các chỉ số thực vật (VIs) tính từ ảnh UAV đa phổ. Chỉ số NDVI (*Normalized Difference Vegetation Index*), OSAVI (*Optimized Soil Adjusted Vegetation Index*) và NSRI (*NIR Shoulder Region Index*) là các chỉ số tối ưu được lựa chọn bằng phương pháp đánh giá tương quan với hàm lượng diệp lục (giá trị SPAD) mà đã được đo đạc trên lá lúa ngoài thực tế. Hàm lượng diệp lục trên lá lúa được ước tính bằng VIs thông qua các mô hình học máy, bao gồm mô hình Hồi quy tuyến tính - *Linear Regression* (LR), cây quyết định ngẫu nhiên - *Random Forest* (RF), hồi quy điểm láng giềng gần nhất - *KNN Regression* (KNN) và hồi quy hỗ trợ vector - *Support Vector Regression* (SVR). Phương án kết hợp chỉ số NSRI và OSAVI cho kết quả tốt nhất và mô hình LR đạt độ chính xác cao hơn so với các mô hình KNN, SVR và RF. Kết quả nhận được cho thấy rằng kết hợp VIs từ ảnh UAV đa phổ bằng mô hình LR đã cải thiện độ chính xác của kết quả ước tính hàm lượng diệp lục trên lá lúa. Kết quả bài báo cũng là cơ sở tin cậy để ứng dụng phương pháp này cho cánh đồng có diện tích lớn trong định hướng sử dụng phân bón nhằm tối ưu năng suất cây lúa.

**Từ khóa:** Hàm lượng diệp lục (giá trị SPAD); Chỉ số thực vật; UAV, Cây lúa.

### 1. Đặt vấn đề

Sức khỏe cây lúa có thể được đánh giá trên cơ sở các yếu tố đặc trưng của lá, bao gồm độ ẩm, hàm lượng diệp lục, cấu trúc bề mặt lá và khả năng chuyển hóa các dưỡng chất. Những yếu tố này không chỉ ảnh hưởng đến mức độ quang hợp mà còn phản ánh thực trạng dinh dưỡng và khả năng thích nghi của cây trồng trước các điều kiện môi trường. Diệp lục là loại sắc tố quan trọng đóng vai trò chủ yếu trong chu trình quang hợp của cây trồng nói chung và cây lúa nói riêng. Hàm lượng diệp lục cao thường phản ánh sức khỏe tốt và ngược lại. Khi lúa chịu ảnh hưởng của các yếu tố môi trường như ô nhiễm không khí hoặc thiếu nước, hàm lượng diệp lục có thể giảm, dẫn đến sự suy giảm hiệu quả quang hợp [1–2] và làm giảm năng suất cây lúa [2–3]. Việc theo dõi mức độ diệp lục trên lá lúa có thể giúp người dân điều chỉnh các giải pháp bón phân phù hợp nhằm tối ưu hóa tăng trưởng và năng suất cây lúa.

Có một số phương pháp được áp dụng để xác định hàm lượng diệp lục trên lá lúa, như phương pháp lấy mẫu và chiết xuất trong phòng thí nghiệm với máy quang phổ. Các phương pháp đạt độ chính xác cao nhưng đòi hỏi cần có cơ sở vật chất cần thiết trong phòng thí nghiệm [4]. Bên cạnh đó, nhiều nghiên cứu đã áp dụng phương pháp đo hàm lượng chất diệp lục qua giá trị SPAD bằng máy đo SPAD cầm tay. Giá trị SPAD có mối quan hệ thuận và cao với hàm lượng diệp lục [5–6]. Tuy nhiên, phương pháp đo quang phổ trong phòng thí

nghiệm và phương pháp đo giá trị SPAD bị hạn chế về mặt không gian, hàm lượng diệp lục chỉ được xác định ở phạm vi các điểm mẫu. Để giải quyết hạn chế này, ảnh viễn thám nói chung và ảnh UAV nói riêng kết hợp với giá trị đo SPAD để ước tính hàm lượng diệp lục trong lá. Ví dụ chỉ số thực vật tỉ số tính từ ảnh Sentinel-2 kết hợp với giá trị SPAD để ước tính hàm lượng diệp lục trên lá khoai tây [7]. Chỉ số thực vật chuẩn hóa (NDVI - *Normalized Difference Vegetation Index*), chỉ số thực vật chuẩn hóa xanh (GNDVI - *Green Normalized Difference Vegetation Index*) và chỉ số diệp lục tán cây (CCCI - *Canopy Chlorophyll Content Index*) tính từ ảnh UAV cũng được áp dụng trong ước tính hàm lượng diệp lục một cách hữu hiệu [8]. Cũng với dữ liệu ảnh UAV, chỉ số NDVI, Difference Vegetation Index (DVI), Ratio Vegetation Index (RVI), GNDVI cũng được mô hình hóa trong ước tính hàm lượng diệp lục trên lá lúa mì, ngô và lúa mạch [9–11].

Quan hệ giữa VIs với giá trị SPAD là cơ sở để áp dụng các mô hình học máy trong ước tính hàm lượng diệp lục của cây trồng. Một số mô hình học máy như mô hình hồi quy tuyến tính đa biến - *Multiple Linear Regression* (MLR), hồi quy hỗ trợ vector - *Support Vector Regression* (SVR), rừng ngẫu nhiên - *Random Forest* (RF), tăng cường độ dốc cực đại - *Extreme Gradient Boosting* (XGBoost) và mạng nơ ron lan truyền ngược - *Back Propagation Neural Network* (BP-NN). Những mô hình này đều được nhận định là có độ chính xác cao [9–15]. Bên cạnh đó, phương pháp KNN cũng được cho là mô hình hiệu quả trong ước tính hàm lượng diệp lục của cây trồng [16]. Việc lựa chọn mô hình phù hợp mà có khả năng cải thiện được độ chính xác là cần thiết khi nghiên cứu ước tính hàm lượng diệp lục trên lá của cây trồng. Mô hình SVR cũng được chứng minh là tốt hơn mô hình BP-NN khi ước tính hàm lượng diệp lục của cây táo [17]. Thêm nữa, mô hình RF được nhận định là tốt hơn các mô hình hồi quy tuyến tính đơn biến - *Univariate Linear Regression* (ULR), MLR và SVR trong ước tính hàm lượng diệp lục trên lá cây táo [18]. Tuy nhiên, hiện còn hạn chế về các nghiên cứu so sánh độ chính xác của các phương pháp học máy này trong ước tính hàm lượng diệp lục trên lá lúa [19].

Do đó, bài báo này có các mục tiêu chính sau: (1) Chứng minh hiệu quả của việc kết hợp VIs trong ước tính hàm lượng diệp lục trên lá lúa từ dữ liệu ảnh UAV đa phổ; (2) Đánh giá, lựa chọn mô hình học máy tốt nhất trong ước tính hàm lượng diệp lục trên lá lúa.

## 2. Phương pháp nghiên cứu và dữ liệu

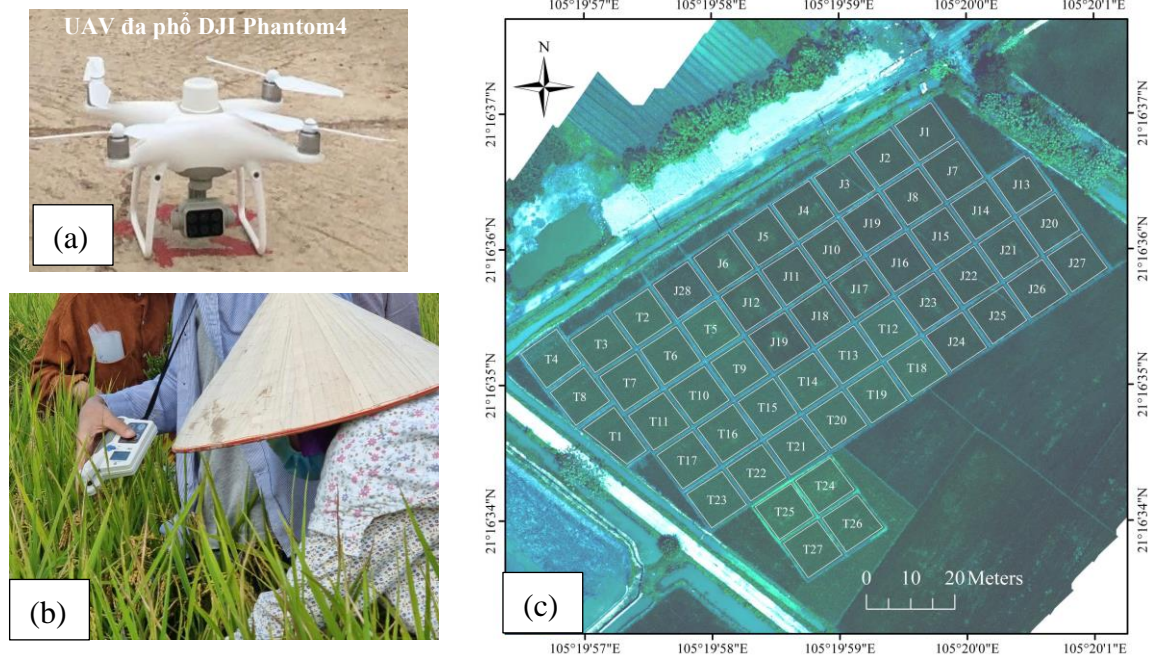
### 2.1. Khu vực nghiên cứu và dữ liệu

Khu vực thực nghiệm có diện tích 6500m<sup>2</sup>, thuộc khu vực xã Vĩnh Lại, huyện Lâm Thao, tỉnh Phú Thọ. Khu vực thực nghiệm là diện tích đất chuyên trồng lúa, có địa hình bằng phẳng, nằm cạnh ruộng nước nên thuận tiện cho việc tưới tiêu. Các ô mẫu được thiết kế dạng hình vuông 10m × 10m. Các ô mẫu liền kề được ngăn cách với nhau bởi các bờ có chiều rộng 1m. Trong tổng số 55 ô mẫu, 27 ô trồng giống lúa của địa phương TBR225 (T) và được ký hiệu lần lượt từ T1-T27. Bên cạnh đó, giống lúa Japonica 02 (J) của Nhật Bản được trồng ở 28 ô mẫu và được ký hiệu lần lượt từ J1-J28 (Hình 1c). Ảnh UAV được thu nhận bằng bộ cảm đa phổ, bao gồm các kênh: xanh lục (Blue: 450nm±16nm); xanh lá cây (Green: 560nm±16nm); đỏ (red: 650nm±16nm); rìa đỏ (Red Edge: 730nm±16nm) và cận hồng ngoại (Near - Infrared: 840nm±26nm). Ảnh UAV được thu nhận ngày 19/05/2022 tương ứng với giai đoạn chín sữa của cây lúa. Tuyến bay UAV được thiết kế song song với trục dọc ô ruộng mẫu (từ hướng Tây - Nam sang Đông - Bắc), độ phủ trùm ảnh dọc và ngang đều là 75%, chiều cao bay chụp ảnh 76m, độ phân giải ảnh mặt đất đạt 4cm. Ca bay chụp ảnh UAV được thực hiện trong 09 phút bay (từ 10 giờ 45 phút đến 10 giờ 54 phút) trong điều kiện thời tiết nắng, trời quang mây. Ảnh UAV được nắn chỉnh hình học và chuẩn hóa phản xạ phổ, xử lý và thành lập ảnh trực giao trên phần mềm Agisoft Metashape.

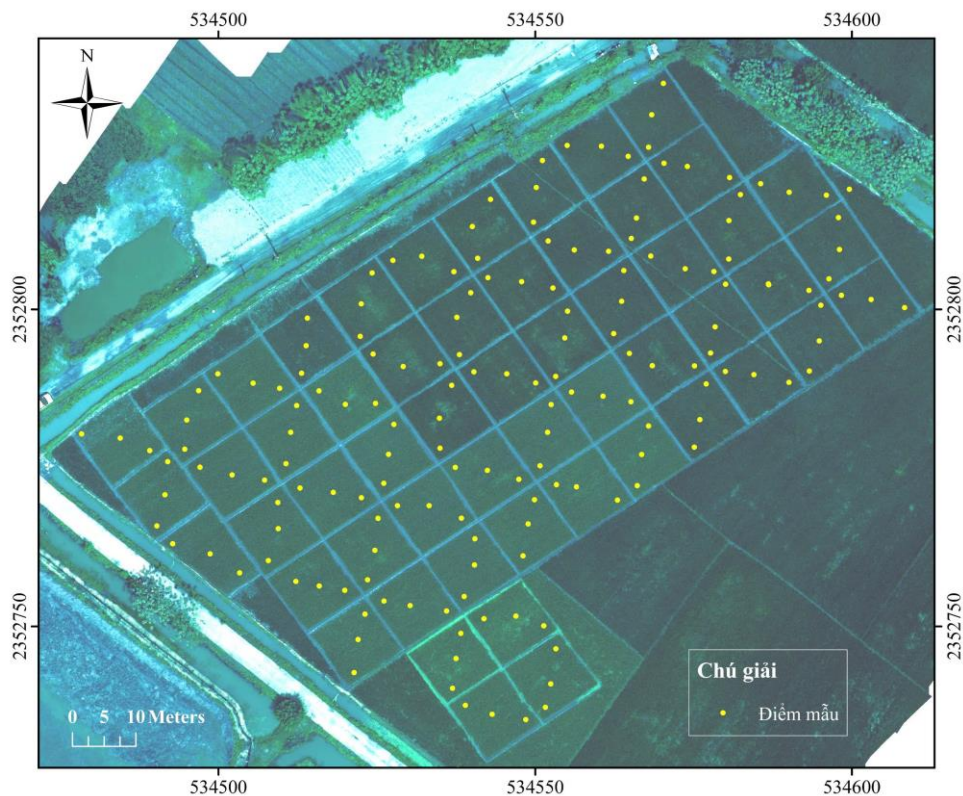
Bên cạnh đó, ở cùng thời điểm chụp ảnh UAV, dữ liệu hàm lượng diệp lục (giá trị SPAD) được đo trực tiếp bằng máy Chlorophyll Meter LCPM-A10 cầm tay (Hình 1b). Các điểm đo



SPAD trên lá lúa được phân bố đều ở các ô mẫu với trị đo ba điểm (Hình 2), giá trị SPAD tại mỗi điểm đo là giá trị trung bình đo được của 03 lá lúa. Máy Chlorophyll Meter LCPM-A10 hoạt động dựa trên nguyên lý đo mức độ hấp thụ ánh sáng của chất diệp lục trong lá tại vùng ánh sáng đỏ (650-660 nm) và gần vùng hồng ngoại (850-880 nm). Do vậy, giá trị SPAD được tính toán dựa trên tỷ lệ mức độ phản xạ giữa kênh đỏ và kênh hồng ngoại, biểu thị hàm lượng diệp lục trong lá cây. Để có cơ sở xác định điểm tương đồng trên ảnh UAV, vị trí lấy mẫu được xác định tọa độ bằng công nghệ GNSS động (Hình 2).



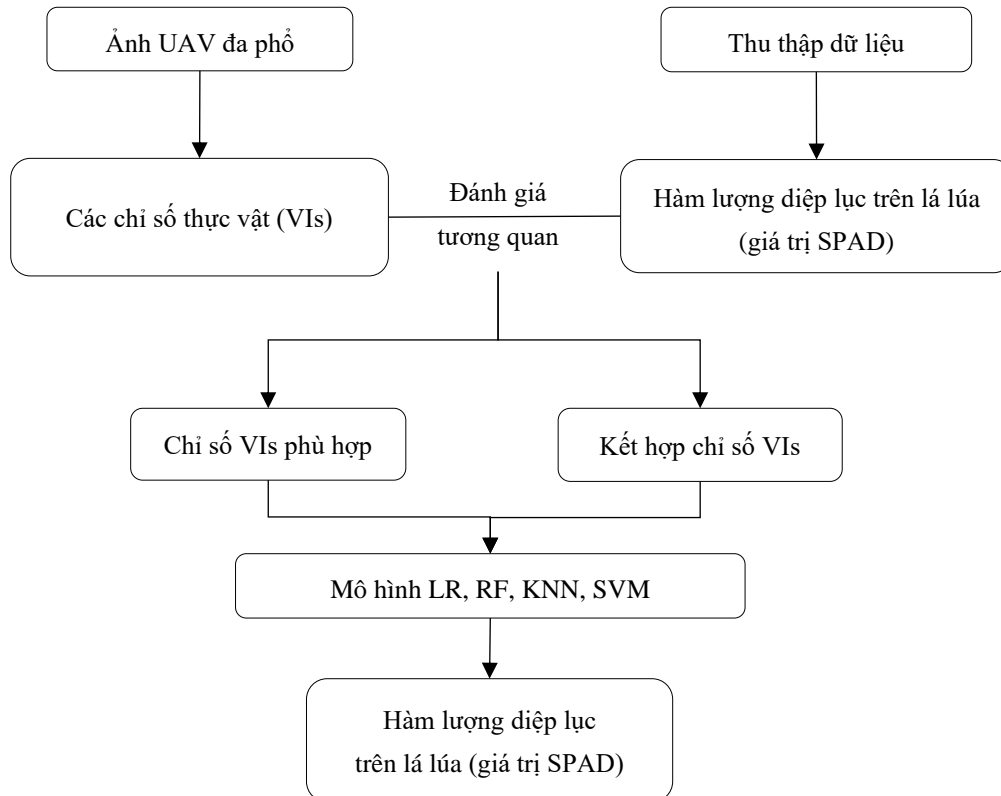
**Hình 1.** Dữ liệu và khu vực nghiên cứu: (a) UAV đa phổ DJI phantom4; (b) đo diệp lục trên lá bằng máy Chlorophyll Meter LCPM-A10; (c) Ảnh UAV chụp ngày 19/05/2022.



**Hình 2.** Phân bố các điểm mẫu.

## 2.2. Quy trình nghiên cứu

Hình 3 thể hiện chu trình ước tính hàm lượng diệp lục trên lá lúa sử dụng dữ liệu ảnh UAV đa phổ. Ảnh UAV đa phổ được sử dụng để tính VIs. Bên cạnh đó, việc khảo sát thực địa nhằm thu thập hàm lượng chất diệp lục (giá trị SPAD) trên lá bằng máy Chlorophyll Meter LCPM-A1. Tiếp theo đó, phương pháp phân tích tương quan được áp dụng để xác định mối quan hệ giữa VIs với giá trị SPAD nhằm lựa chọn những chỉ số phù hợp trong ước tính hàm lượng diệp lục cho toàn bộ khu vực thực nghiệm. Dựa trên chỉ số đã được lựa chọn, các mô hình học máy LR, KNN, SVR và RF được áp dụng để ước tính hàm lượng diệp lục trên lá lúa. Kết quả ước tính bởi các mô hình được đánh giá mức độ chính xác thông qua hệ số tương quan  $R^2$  của tập mẫu huấn luyện,  $R^2$  tập kiểm tra và sai số trung phương RMSE.



**Hình 3.** Quy trình ước tính hàm lượng diệp lục trên lá lúa từ dữ liệu ảnh UAV đa phổ.

## 2.3. Chỉ số thực vật (VIs)

Trong khả năng thu nhận các kênh phổ của ảnh UAV đa phổ, bài báo này sử dụng 17 chỉ số thực vật (Bảng 1).

**Bảng 1.** Chỉ số VIs.

TT	Chỉ số	Công thức	Tham khảo
1	NIR Shoulder Region Index	$NSRI = R_{NIR} / R_{RE}$	[21]
2	Difference Vegetation Index	$DVI = R_{NIR} - R_R$	[22]
3	Enhanced Vegetation Index	$EVI = 2,5(R_{NIR} - R_R) / (R_{NIR} + 6R_R - 7,5R_B + 1)$	[23]
4	False Color Vegetation Index	$FCVI = 1,5(2R_{NIR} + R_B - 2R_G) / (2R_G + 2R_B - 2R_{NIR} + 127,5)$	[24]
5	Green Normalized Difference Vegetation Index	$GNDVI = (R_{NIR} - R_G) / (R_{NIR} + R_G)$	[25]
6	Leaf Chlorophyll Index	$LCI = (R_{NIR} / R_{RE}) - 1$	[26]

7	Modified Chlorophyll Absorption in Reflec- tance Index	$MCARI = R_{RE} - R_R - 0,2(R_{RE} - R_G)R_{RE}/R_R$	[27]
8	Modified Soil Adjusted Vegetation Index	$MSAVI = 0,5(2R_{NIR} + 1 - \sqrt{(2R_{NIR} + 1)^3 - 8(R_{NIR} - R_R)})$	[28]
9	Modified Simple Ratio	$MSR = ((R_{NIR}/R_R) - 1)/((R_{NIR}/R_R) + 1)$	[29]
10	Normalized Difference Red Edge Index	$NDRE = (R_{NIR} - R_{RE})/(R_{NIR} + R_{RE})$	[30]
11	Normalized Difference Vegetable Index	$NDVI = (R_{NIR} - R_R)/(R_{NIR} + R_R)$	[31]
12	Optimized Soil Adjusted Vegetation Index	$OSAVI = (R_{NIR} - R_R)/(R_{NIR} + R_R + 0,16)$	[32]
13	Photochemical Reflec- tance Index	$PRI = (R_G - R_B)/(R_G + R_B)$	[33]
14	Red Edge Chlorophyll Index	$RECI = (R_{NIR} - R_{RE})/R_{RE}$	[34]
15	Ratio Vegetation Index	$RVI = R_{NIR}/R_R$	[22]
16	Optimized Soil Adjusted Vegetation Index	$TVI = (R_G - R_R)/(R_G + R_R)$	[35]
17	Structure Intensive Pig- ment Index	$SIPI = (R_{NIR} - R_B)/(R_{NIR} + R_B)$	[36]

Trong đó:  $R_R$ ,  $R_B$ ,  $R_G$ ,  $R_{RE}$  và  $R_{NIR}$  - Giá trị phổ phản xạ tương ứng trên ảnh UAV của các kênh màu đỏ (Red), xanh lục (Blue), xanh dương (Green), kênh rìa đỏ (Rededge) và cận hồng ngoại (Near infrared).

#### 2.4. Mô hình ước tính hàm lượng chất diệp lục

##### 1. Hồi quy tuyến tính (LR)

LR là một trong những mô hình học máy được ứng dụng để mô hình hóa mối quan hệ tuyến tính giữa hàm lượng diệp lục trong lá lúa với các chỉ số thực vật được xây dựng từ ảnh UAV đa phổ. Mô hình hồi quy tuyến tính nhằm mục đích tính toán các hệ số hồi quy mà đảm bảo sai số giữa giá trị mà mô hình ước tính được với giá trị khảo sát thực tế được tối thiểu hóa.

Công thức mô hình hồi quy tuyến tính được thể hiện bằng công thức (1):

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \epsilon \quad (1)$$

Trong đó  $y$  là hàm lượng chất diệp lục được ước tính - giá trị SPAD ước tính;  $x_i$  là chỉ số thực vật đã được lựa chọn;  $\beta_0$  là hằng số;  $\beta_i$  là hệ số hồi quy của chỉ số thực vật đã được lựa chọn;  $\epsilon$  là sai số ngẫu nhiên, phản ánh sự biến thiên không thể giải thích từ các chỉ số thực vật.

##### 2. Mô hình KNN Regression (KNN)

KNN là một thuật toán phi tham số dự đoán biến mục tiêu (hàm lượng diệp lục) dựa trên  $k$  lân cận gần nhất trong không gian các yếu tố độc lập (chỉ số thực vật) [37]. Trong việc ước tính hàm lượng diệp lục trong lá lúa, KNN tìm kiếm  $k$  điểm dữ liệu gần nhất với điểm cần dự đoán và lấy trung bình cộng của các giá trị hàm lượng chất diệp lục từ các láng giềng này nhằm đưa ra kết quả ước tính.

Công thức tổng quát của mô hình KNN được mô tả như sau:

$$y = \frac{1}{k} \sum_{i=1}^n y_i \quad (2)$$

Trong đó  $y$  là hàm lượng chất diệp lục được ước tính (giá trị SPAD ước tính);  $k$  là số lượng láng giềng gần nhất;  $y_i$  là giá trị hàm lượng chất diệp lục của láng giềng gần nhất.

##### 3. Support Vector Regressor (SVR)

Trong mô hình SVR, mục tiêu là tìm một hàm số  $f(x)$  sao cho sai số giữa giá trị ước tính hàm lượng diệp lục trong lá lúa và giá trị thực tế đã được đo bằng máy Chlorophyll Meter LCPM-A10 (giá trị SPAD) nằm trong một phạm vi chấp nhận được. Mô hình SVR tìm cách xây dựng một hàm hồi quy dưới dạng:

$$f(x) = (w, x) + b = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b \quad (3)$$

Trong đó  $f(x)$  là giá trị ước tính hàm lượng chất diệp lục (giá trị SPAD ước tính) dựa trên đặc trưng  $x$ ;  $x = (x_1, x_2, \dots, x_n)$  là vector đặc trưng ứng với các chỉ số thực vật đã lựa chọn;  $w = (w_1, w_2, \dots, w_n)$  là vector trọng số tương ứng với các chỉ số thực vật đã lựa chọn;  $b$  là độ lệch (bias), một tham số điều chỉnh.

#### 4. Random Forest Regressor (RF)

Để ước tính hàm lượng chất diệp lục trong lá lúa, mô hình RF có thể dựa vào một tập hợp các yếu tố như độ ẩm, nhiệt độ, hoặc các chỉ số quang phổ. Dữ liệu về chỉ số thực vật và giá trị SPAD được chia thành các tập con ngẫu nhiên. Dữ liệu được chia làm hai phần: một phần dùng để làm mẫu huấn luyện mô hình, phần còn lại được sử dụng để kiểm tra và tính toán độ tin cậy của mô hình.

$$y = \frac{1}{M} \sum_{i=1}^M f_i(x) \quad (4)$$

Trong đó  $y$  là hàm lượng chất diệp lục ước tính (giá trị SPAD ước tính);  $M$  là số lượng tập con;  $f_i(x)$  là dự đoán của cây  $i$  dựa vào dữ liệu đầu vào  $x$  (các chỉ số thực vật).

#### 5. Đánh giá độ chính xác

Độ chính xác của mô hình được đánh giá bằng hệ số hồi quy bình phương ( $R^2$ ) của tập mẫu huấn luyện và của tập mẫu kiểm tra. Bên cạnh đó, sai số trung phương (RMSE) cũng được sử dụng để đánh giá độ chính xác của mô hình. RMSE được tính theo công thức sau đây:

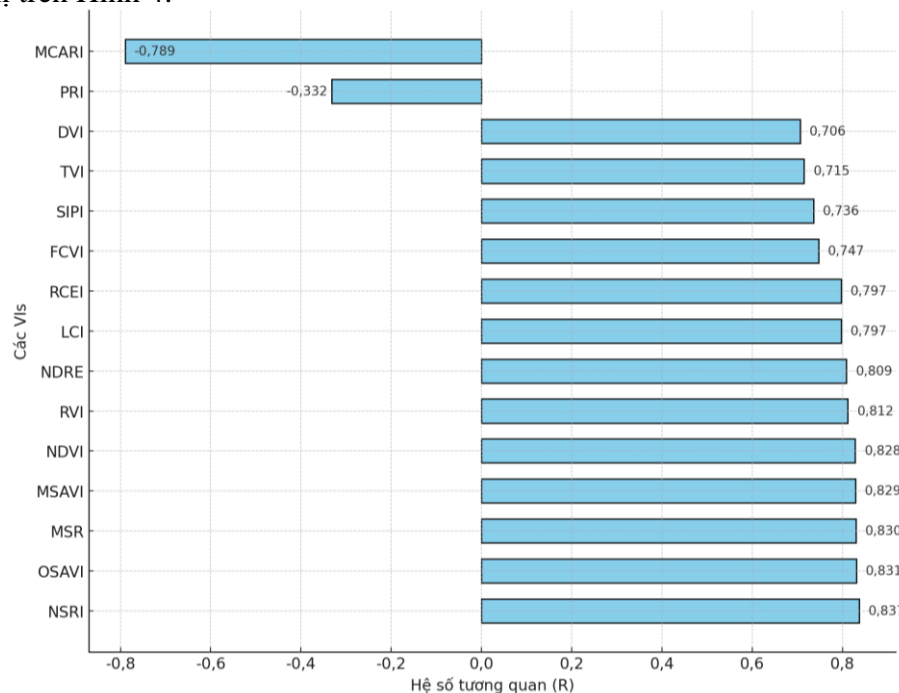
$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - X_i)^2}{n}} \quad (5)$$

Trong đó  $X_i$  là giá trị SPAD đo được trên mẫu thực địa;  $Y_i$  là giá trị SPAD ước tính bằng các mô hình LR, KNN, SVR và RF;  $n$  là số lượng mẫu.

### 3. Kết quả

#### 3.1. Lựa chọn chỉ số

Hệ số hồi quy là thông số được áp dụng để đánh giá, lựa chọn chỉ số VIs phù hợp nhất trong ước tính hàm lượng diệp lục trong lá lúa. Các hệ số này được xác định dựa trên mức độ tương quan giữa chỉ số VIs với giá trị SPAD thu thập được ở 165 mẫu thực địa, kết quả được biểu thị trên Hình 4.



**Hình 4.** Hệ số tương quan của các chỉ số VIs với hàm lượng diệp lục.



Hình 4 cho thấy mức độ tương quan giữa từng yếu tố VIs với giá trị SPAD có sự khác biệt. Các chỉ số như NRSI, OSAVI, MSR, MSAVI, NDVI, EVI, RVI, NDRE, LCI, và RCEI đều biểu thị mối tương quan mạnh với hàm lượng chất diệp lục, với các giá trị hệ số tương quan thay đổi trong khoảng từ 0,79 đến 0,84. Điều này chỉ ra rằng khi các chỉ số này gia tăng, hàm lượng diệp lục cũng tăng. Hệ số tương quan cao này cho thấy các chỉ số trên có vai trò đáng kể trong việc ước tính hàm lượng diệp lục. Bên cạnh đó, các chỉ số GNDVI, FCVI, SIPI, TVI, và DVI có mối tương quan trung bình với hàm lượng chất diệp lục, với các hệ số tương quan trong phạm vi từ 0,71 đến 0,79. Mặc dù sự ảnh hưởng của chúng đối với hàm lượng chất diệp lục không mạnh như các chỉ số trong nhóm trước, nhưng chúng vẫn có cần thiết trong việc giải thích sự thay đổi của hàm lượng diệp lục. Ngược lại, các chỉ số PRI, và MCARI lại có mối quan hệ nghịch với hàm lượng diệp lục, với các giá trị tương quan lần lượt là -0,332, và -0,789. Điều này chứng tỏ khi các chỉ số này gia tăng, giá trị hàm lượng chất diệp lục sẽ giảm. Từ những đánh giá kết quả trên, dựa vào hệ số tương quan của giá trị SPAD với các chỉ số VIs có thể xác định được hàm lượng chất diệp lục trên lá lúa bằng chỉ số VIs phù hợp.

Các chỉ số đã được chọn bao gồm NDVI, OSAVI và NSRI, sự lựa chọn này dựa trên cơ sở sau: (1) Chỉ số thực vật phổ biến đã được sử dụng trong các nghiên cứu về diệp lục cây trồng NDVI. Từ Hình 4 có thể thấy được hệ số tương quan của NDVI với hàm lượng diệp lục ở mức cao (0,828). Chỉ số NDVI được đặc biệt áp dụng rất nhiều trong các nghiên cứu đánh giá hàm lượng diệp lục và sự phát triển của thực vật do NDVI rất nhạy với sự thay đổi hàm lượng diệp lục đối với các sự biến đổi trạng thái thực vật; (2) Hai chỉ số thực vật có tương quan mạnh nhất, nhì với hàm lượng chất diệp lục, bao gồm NSRI và OSAVI, hệ số tương quan ở mức rất cao, lần lượt là 0,837 và 0,831. NSRI là một chỉ số phổ biến và hiệu quả trong việc đánh giá sức khỏe thực vật, đặc biệt là trong các môi trường có độ che phủ thực vật cao. Chỉ số này sử dụng các bức xạ trong dải phổ đỏ và đỏ cạnh (*Red-Edge*), nơi có sự thay đổi rõ rệt khi hàm lượng hàm lượng chất diệp lục thay đổi. Bên cạnh đó, OSAVI là một chỉ số tối ưu, điều chỉnh sự tác động của yếu tố khí quyển và giúp tăng độ chính xác khi đánh giá sức khỏe thực vật. OSAVI giảm ảnh hưởng của đất nền trong các vùng có mật độ thực vật thấp hoặc trung bình. Bên cạnh đó, OSAVI có thể được liên kết với các thông số sinh lý khác của cây trồng như hàm lượng diệp lục [38]; (3) Kết hợp hai chỉ số VIs có tương quan cao nhất và nhì với giá trị SPAD - NSRI và OSAVI. Sự kết hợp này nhằm cải thiện chính xác của kết quả ước tính hàm lượng chất diệp lục.

### 3.2. Mô hình ước tính hàm lượng chất diệp lục

Các mô hình học máy LR, KNN, SVR và RF được áp dụng để ước tính hàm lượng diệp lục trên lá lúa bằng các chỉ số NDVI, OSAVI và NSRI đã được lựa chọn. Các mô hình được chạy với bộ dữ liệu gồm 165 mẫu (Hình 2), được chia theo tỉ lệ mẫu huấn luyện: mẫu kiểm tra là 80:20. Các mẫu được trộn ngẫu nhiên để đảm bảo rằng các giá trị điểm mẫu được phân phối đều trong cả mẫu huấn luyện và mẫu kiểm tra. Việc ước tính hàm lượng chất diệp lục trên lá lúa được thực nghiệm với bốn kịch bản tương ứng với 3 nhóm tiêu chí chọn chỉ số đã phân tích trên mục 3.1. Bảng 2 là độ chính xác tương ứng với từng mô hình trong mỗi kịch bản.

**Bảng 2.** Độ chính xác của các mô hình.

TT	Chỉ số VIs	Tên mô hình	Hệ số tương quan ( $R^2$ )		RMSE
			Dữ liệu huấn luyện	Dữ liệu kiểm tra	
1	NDVI	LR	0,677	0,723	2,14
		KNN	0,759	0,682	2,29
		SVM	0,680	0,642	2,43
		RF	0,947	0,675	2,31
2	NSRI	LR	0,704	0,680	2,44
		KNN	0,812	0,668	1,94

TT	Chỉ số VIs	Tên mô hình	Hệ số tương quan ( $R^2$ )		RMSE
			Dữ liệu huấn luyện	Dữ liệu kiểm tra	
3	OSAVI	SVM	0,742	0,661	2,28
		RF	0,955	0,593	0,94
		LR	0,684	0,724	2,13
		KNN	0,756	0,713	2,18
		SVM	0,686	0,643	2,43
		RF	0,945	0,708	2,20
		LR	0,720	0,734	2,10
4	NSRI và OSAVI	KNN	0,791	0,671	2,33
		SVM	0,735	0,683	2,29
		RF	0,957	0,707	2,20

Trong kịch bản chỉ áp dụng NDVI để ước tính hàm lượng diệp lục, mô hình LR thể hiện có độ chính xác hơn cả, với giá trị  $R^2$  kiểm tra đạt 0,723, đồng thời có sai số RMSE thấp nhất (2,14). Do vậy, mô hình chỉ ra khả năng ước tính tốt và ổn định trên tập mẫu kiểm tra. Mặc dù mô hình KNN có  $R^2$  huấn luyện cao hơn (0,759), nhưng độ chính xác kiểm tra lại thấp hơn so với LR ( $R^2$  kiểm tra chỉ đạt 0,682). Hơn nữa, RMSE của KNN là 2,29, cao hơn so với LR, SVR và RF mặc dù có độ chính xác huấn luyện tốt, nhưng lại gặp hạn chế trong việc tổng quát hóa khi áp dụng trên dữ liệu kiểm tra, với  $R^2$  kiểm tra lần lượt là 0,642 và 0,675, và RMSE lần lượt là 2,43 và 2,31, cao hơn so với LR.

Kịch bản ước tính hàm lượng diệp lục bằng chỉ số NSRI cho thấy  $R^2$  huấn luyện khá cao, lên tới 0,945 với mô hình RF. Tuy nhiên,  $R^2$  lại không cao, giá trị giao động từ 0,593 tới 0,680. Bên cạnh đó, RMSE cao nhất đối với mô hình LR (2,44) và thấp nhất với mô hình RF (0,944). Mô hình RF trong kịch bản này có kết quả RMSE tốt nhất nhưng có  $R^2$  kiểm tra lại thấp nhất nên cũng không được xem là giải pháp tốt trong ước tính hàm lượng diệp lục.

Tương tự như trong kịch bản đầu tiên, kịch bản sử dụng chỉ số thực vật OSAVI, mô hình LR cũng cho thấy độ chính xác tốt nhất khi có  $R^2$  kiểm tra cao nhất (0,724) và RMSE thấp nhất (2,13). Mặc dù mô hình KNN cho thấy độ chính xác kiểm tra khá cao ( $R^2$  kiểm tra = 0,713), nhưng sai số RMSE của mô hình này (2,18) vẫn cao hơn so với LR. RF và SVR tiếp tục không đạt được kết quả cao về độ chính xác kiểm tra, với  $R^2$  kiểm tra lần lượt là 0,708 và 0,643, và RMSE cao hơn (2,20 và 2,43).

Trong kịch bản cuối cùng, hàm lượng chất diệp lục được ước tính khi kết hợp chỉ số NSRI và OSAVI. Trong kịch bản này, LR tiếp tục duy trì vị trí có độ chính xác tốt nhất với  $R^2$  kiểm tra cao nhất (0,734) và RMSE thấp nhất (2,10). RF có  $R^2$  huấn luyện cao (0,957), nhưng độ chính xác của mô hình này trên mẫu kiểm tra lại không tốt, khi  $R^2$  kiểm tra chỉ đạt 0,707 và RMSE là 2,20. KNN và SVR có hiệu suất kiểm tra thấp hơn nhiều so với LR, với  $R^2$  kiểm tra lần lượt là 0,671 và 0,683, cùng RMSE lần lượt là 2,33 và 2,29, cho thấy các mô hình này không phù hợp với mẫu kiểm tra.

### 3.3. Đánh giá hàm lượng chất diệp lục trong lá lúa

Từ phân tích ở mục 3.2, hàm lượng diệp lục tương đối trên lá lúa được xác định tốt nhất khi kết hợp OSAVI với NSRI, bằng phương pháp LR. Bảng 3 là kết quả phân tích LR đối với kịch bản kết hợp chỉ số OSAVI với chỉ số NSRI.

**Bảng 3.** Kết quả mô hình LR trong ước tính hàm lượng diệp lục.

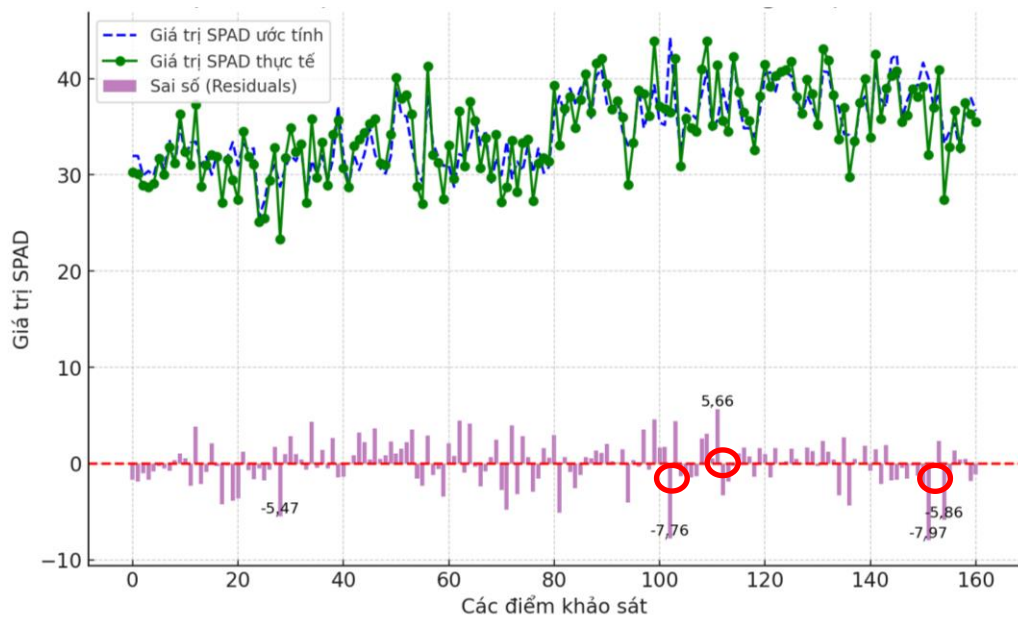
	Hệ số hồi quy ( $\beta$ )	Độ lệch chuẩn (Std.Err)	Thống kê kiểm định t	Giá trị xác suất: $P >  t $	Khoảng tin cậy 95% ([0,025, 0,975])	
					[0,025	0,975]
Hằng số ( $\beta_0$ )	-22,2815	4,4586	-4,9973	1,5269	-31,0878	-13,4752
OSAVI	40,3567	11,4467	3,5256	0,0006	17,7483	62,9650
NSRI	36,6640	8,5554	4,2854	0,0000	19,7663	53,5617



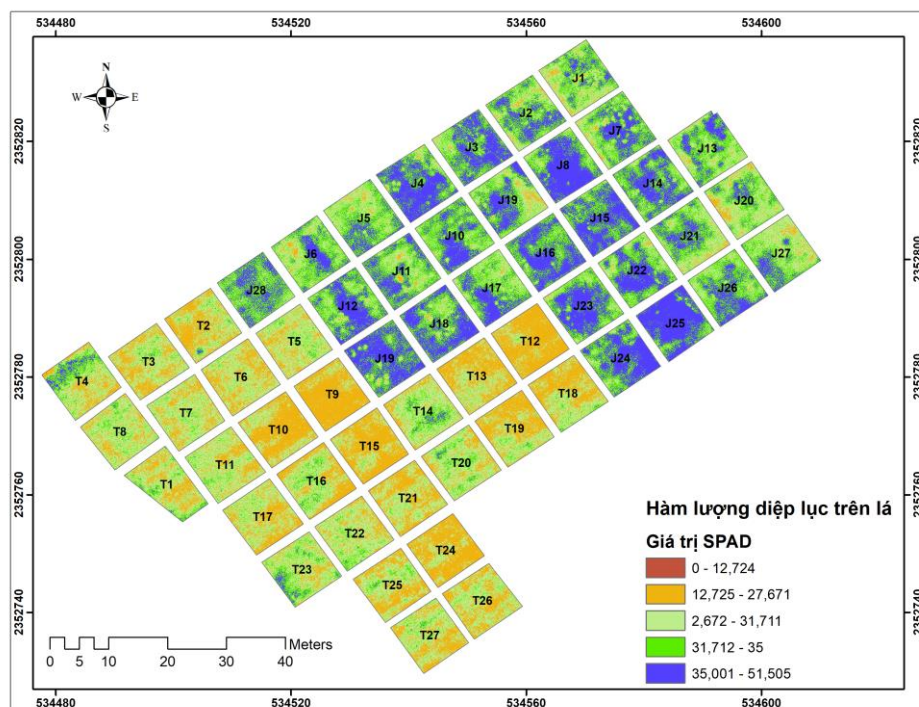
Ngoài giá trị  $R^2$  và RMSE của mô hình LR với kịch bản kết hợp OSAVI và NSRI đạt giá trị cao (Bảng 2), Bảng 3 còn thể hiện rằng, giá trị Std.Er, t, p và khoảng tin cậy trong mô hình LR ước tính hàm lượng diệp lục cũng đạt độ chính xác cao, đảm bảo có ý nghĩa về mặt thống kê. Do vậy, với kết quả có trong Bảng 3, hàm lượng diệp lục trên lá lúa được ước tính theo hàm hồi quy sau đây:

$$\text{Hàm lượng diệp lục} = -22,2815 + 36,6640 \times \text{NSRI} + 40,3567 \times \text{OSAVI} \quad (6)$$

Mô hình LR đạt độ chính xác thống kê và sai số trung phương tổng thể của cả mô hình RMSE (2,1), sai số của các điểm khảo sát cũng được đánh giá như trên Hình 5. Trong Hình 5 chúng ta có thể thấy đa số các điểm khảo sát có sai số thấp, ngoại trừ một số điểm như điểm 102: Sai số lớn (-7,76), điểm 111: Sai số đáng kể (5,66) và điểm 151: Sai số lớn (-7,97). Ở vị trí những điểm này, lúa bị đập do công tác đi thực địa lấy mẫu lá và đo phổ mặt đất tại các lần đo trước.



**Hình 5.** Sai số các điểm khảo sát.



**Hình 6.** Hàm lượng diệp lục trên lá lúa ở giai đoạn chín sữa.

Hình 6 thể hiện kết quả ước tính hàm lượng diệp lục trên lá lúa ở giai đoạn chín sữa, qua đây có thể chỉ ra rằng giống lúa J có hàm lượng diệp lục cao hơn giống lúa T. Giống lúa J có hàm lượng diệp lục trên lá có giá trị nằm trong khoảng 31,7112 - 51,505. Trong khi đó, giống lúa T có giá trị diệp lục chỉ khoảng 12,725 - 31,711.

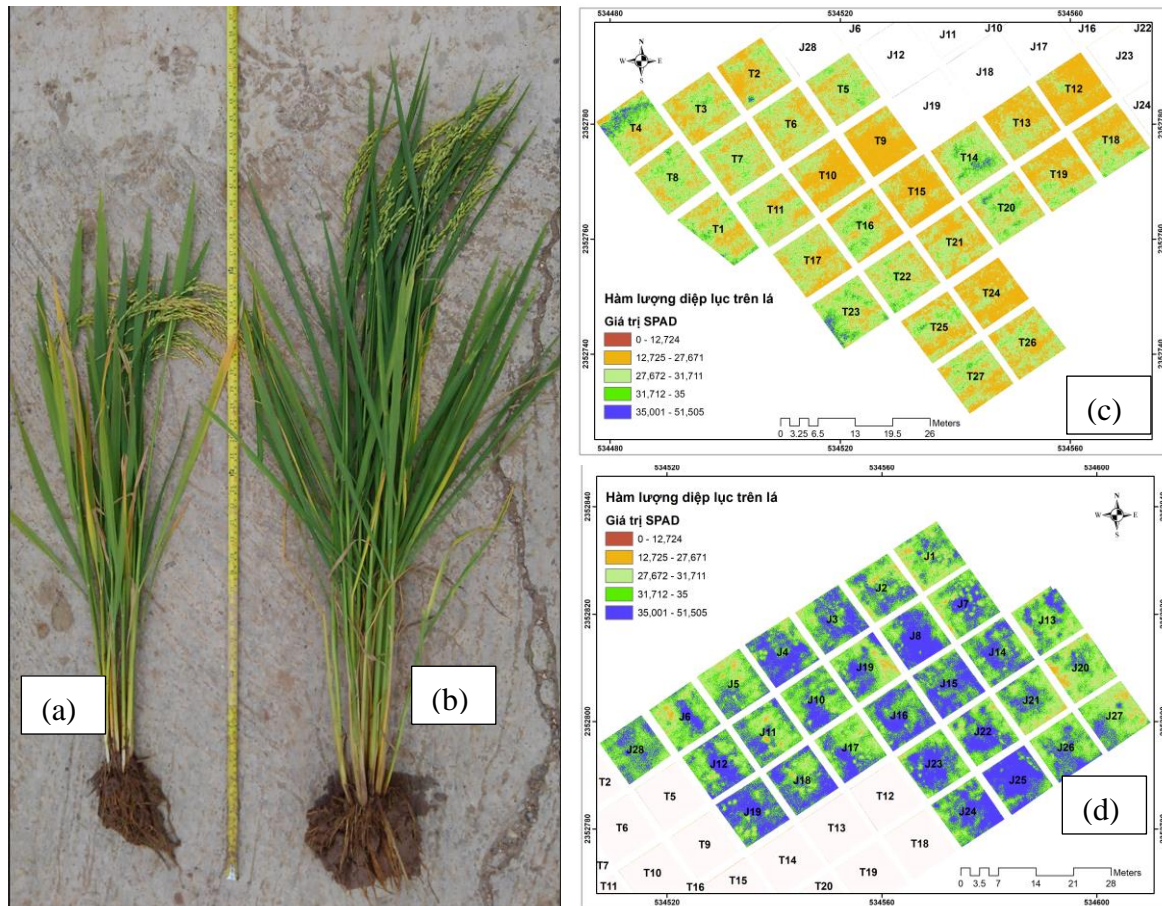
#### 4. Thảo luận

Việc kết hợp chỉ số VIs có vai trò thiết yếu trong việc nâng cao độ chính xác của các mô hình ước tính hàm lượng diệp lục trên lá lúa. Các chỉ số NDVI và OSAVI đã được nhiều nghiên cứu sử dụng để ước tính hàm lượng diệp lục [39–41]. Trong khi đó, chưa có nghiên cứu nào sử dụng chỉ số NSRI. Kịch bản tích hợp NSRI với OSAVI trong việc ước tính hàm lượng diệp lục bằng các mô hình LR, KNN, SVR và RF đã chỉ ra rằng chỉ số NSRI giúp tối ưu hóa hàm lượng diệp lục cần ước tính và nâng cao khả năng khái quát của mô hình. Kết hợp NSRI với chỉ số OSAVI trong kịch bản nghiên cứu giúp tăng độ chính xác của mô hình trên tập mẫu kiểm tra. Cụ thể, trong trường hợp của LR, mô hình sử dụng NSRI cho kết quả cao hơn trong việc ước tính các giá trị kiểm tra với  $R^2$  đạt 0,734 và RMSE là 2,10, thấp hơn so với các kịch bản chỉ sử dụng NDVI hoặc OSAVI. Trong các mô hình, mặc dù RF có thể đạt được độ chính xác huấn luyện rất cao, nhưng khi áp dụng vào mẫu kiểm tra, mô hình này đôi khi gặp phải vấn đề về khả năng tổng quát, thể hiện qua  $R^2$  kiểm tra và RMSE không đạt mức tối ưu. Tuy nhiên, khi kết hợp NSRI vào mô hình, có thể làm giảm hiện tượng overfitting (quá khớp với dữ liệu huấn luyện), từ đó nâng cao khả năng ước tính chính xác trên dữ liệu kiểm tra. Hơn nữa, việc kết hợp NSRI với OSAVI làm phong phú thêm các đặc trưng biểu thị đến hàm lượng diệp lục được sử dụng đối với các mô hình LR, KNN, SVR và RF. Điều này tạo ra một tập hợp đặc trưng đa dạng, giúp mô hình có thể được học tốt hơn, nhất là khi phải xử lý các dữ liệu phức tạp hoặc có tính chất không tuyến tính. Các mô hình như KNN và SVR có thể gặp hạn chế khi xử lý dữ liệu với tính đa dạng cao, sẽ trở lên hiệu quả hơn khi bổ sung thêm NSRI, giúp tăng độ chính xác của mô hình. Như vậy, kịch bản kết hợp chỉ số NSRI và OSAVI giúp nâng cao độ chính xác của việc ước tính hàm lượng diệp lục trên lá lúa.

Khi đánh giá các mô hình dựa trên các thông số như  $R^2$  và RMSE, mô hình LR có giá trị cao nhất trong bốn kịch bản thử nghiệm. Cụ thể, đối với NDVI, OSAVI, NSRI và NSRI kết hợp OSAVI, LR đều có độ chính xác cao trên mẫu kiểm tra ( $R^2$  kiểm tra) và sai số RMSE thấp nhất, phản ánh khả năng tổng quát và ước tính tốt. Trong khi đó, mặc dù RF và KNN đạt độ chính xác huấn luyện cao hơn, nhưng  $R^2$  trên dữ liệu kiểm tra và RMSE của chúng lại không tốt bằng LR. Đặc biệt, mô hình RF mặc dù có  $R^2$  huấn luyện rất cao, nhưng lại không đạt được kết quả tốt đối với dữ liệu kiểm tra với  $R^2$  thấp hơn và RMSE cao hơn. Do đó, mô hình LR là lựa chọn phù hợp nhất vì có khả năng ước tính chính xác trên cả mẫu huấn luyện và kiểm tra.

Hàm lượng diệp lục có thể thay đổi đáng kể giữa các giống lúa khác nhau với các điều kiện môi trường khác nhau [42]. Hình 7c chỉ ra rằng giống lúa T có hàm lượng diệp trên lá thấp hơn ngưỡng phù hợp trong giai đoạn chín sữa, giá trị SPAD đã xác định rằng giá trị SPAD dưới 35 đơn vị thường được cho là là giá trị quan trọng cho thấy nhu cầu bón đạm cho lúa để đạt sự phát triển tối ưu của hạt [43]. Giá trị SPAD của giống lúa T thấp có thể do giống lúa T đang ở giai đoạn chín sữa muộn, gần bước sang giai đoạn thu hoạch (Hình 7a). Giá trị SPAD của giống lúa T cũng có thể biểu hiện sức khỏe của lúa đang trong tình trạng thiếu phân bón cần thiết và phù hợp trong giai đoạn này. Ngược với giống lúa T, Hình 7d lại cho ta thấy giống lúa J đạt ngưỡng tốt, với giá trị SPAD của giống Japonica vùng ôn đới lớn hơn 37,2 [42]. Bên cạnh đó, nghiên cứu [44] cho rằng giá trị SPAD là 45,57 đối với sữa sớm và 35,91 đối với giai đoạn sữa muộn. Hình 5b cho thấy lúa J đang ở giai đoạn chín sữa sớm. Giá trị SPAD ở ngưỡng này là điều kiện quan trọng để đảm bảo mức nitơ đầy đủ và tăng cường sự phát triển của hạt khỏe mạnh [44]. Với sự khác biệt về hàm lượng diệp lục của hai giống lúa T và J, người nông dân cần bón đạm đối với giai đoạn chín sữa sớm để duy trì hàm lượng

diệp lục cần thiết trước khi bước sang giai đoạn chuyển hóa và tích lũy tinh bột của hạt. Bên cạnh đó, lúa cũng cần được bón kali với giai đoạn chín sữa muộn để tối ưu hóa sự tích lũy tinh bột của hạt hướng tới năng suất lúa cao nhất. Giống lúa J có tích lũy diệp lục cao hơn cũng là cơ sở quan trọng để người nông dân lựa chọn giống J được trồng rộng rãi cho những khu vực tương tự về yếu tố tự nhiên, môi trường của khu vực xã Vĩnh Lại, huyện Lâm Thao, tỉnh Phú Thọ.



**Hình 7.** Lúa ở giai đoạn chín sữa: (a) giống lúa TBR225 (T), (b) giống lúa Japonica 02 (J).

## 5. Kết luận

Nhờ có kênh ảnh cận hồng ngoại, ảnh UAV đa phổ đã trở thành dữ liệu hiệu quả trong ước tính hàm lượng diệp lục trên lá thông qua các ảnh chỉ số VIs. Trong số các chỉ số VIs mà bài báo thực nghiệm, chỉ số NSRI được xác định là chỉ số mới hiệu quả trong ước tính hàm lượng diệp lục trên lá của giống J và T tại khu vực thực nghiệm tại xã Vĩnh Lại, huyện Lâm Thao, tỉnh Phú Thọ. Bên cạnh đó, sử dụng kết hợp chỉ số VIs, cụ thể là chỉ số OSAVI kết hợp với NSRI giúp cải thiện độ chính xác khi ước tính hàm lượng diệp lục trên lá lúa từ ảnh UAV đa phổ.

Mô hình LR là mô hình tối ưu nhất trong số các mô hình học máy LR, KNN, SVR, và RF được sử dụng để ước tính hàm lượng diệp lục tại khu vực thực nghiệm. Kết quả của bài báo này chỉ ra rằng giống lúa J có hàm lượng diệp lục trên lá cao hơn giống lúa T, giá trị SPAD giao động từ 31,7112 - 51,505. Điều này cho thấy các giống lúa J có hiệu quả sản xuất diệp lục cao hơn giống lúa T.

Kết quả của bài báo là cơ sở tin cậy để ứng dụng các chỉ số VIs từ ảnh UAV trong xác định hàm lượng diệp lục cho các khu vực trồng lúa với diện tích lớn. Hàm lượng diệp lục trên lá khác nhau ở từng giống lúa cũng giúp việc lựa chọn giống lúa và quản lý việc bón phân hiệu quả nhằm tối ưu hóa năng suất lúa.



Bên cạnh những kết quả mà bài báo đạt được, bài báo còn tồn tại những hạn chế khi chỉ thực nghiệm ước tính được hàm lượng diệp lục trên lá thời kỳ chín sữa của vụ mùa trên hai giống lúa T và J. Để khắc phục được tồn tại này ở các nghiên cứu trong tương lai, dữ liệu mẫu diệp lục, ảnh bay chụp và phương pháp cần được thu thập và thực nghiệm ở nhiều giai đoạn sinh trưởng và phát triển của những giống lúa khác nhau.

**Đóng góp của tác giả:** Xây dựng ý tưởng nghiên cứu; Thu thập, xử lý số liệu: P.T.L., L.V.C.; Phân tích kết quả; Viết bản thảo bài báo, chỉnh sửa bài báo: P.T.L.

**Lời cảm ơn:** Nghiên cứu này được hỗ trợ bởi đề tài mã số T24–30 của Trường đại học Mở - Địa chất do NCS. Lê Văn Cảnh chủ trì.

**Lời cam đoan:** Tác giả cam đoan bài báo này là công trình nghiên cứu của các tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây.

### Tài liệu tham khảo

1. Sikuku, P.; Onyango, G.N.C.; Musyimi, P.D.M. Chlorophyll fluorescence, protein and chlorophyll content of three nerica rainfed rice varieties under varying irrigation regimes. *ARPJ. Agric. Biol. Sci.* **2010**, *2*, 19–25.
2. Shimono, H.; Hasegawa, T.; Fujimura, S.; Iwama, K. Responses of leaf photosynthesis and plant water status in rice to low water temperature at different growth stages. *Field Crops Res.* **2004**, *89*(1), 71–83.
3. Mina, U.; Chandrashekara, T.K.; Kumar, S.N.; Meena, M.C.; Yadav, S.; Tiwari, S.; Singh, D.; Kumar, P.; Kumar, R. Impact of particulate matter on basmati rice varieties grown in Indo-Gangetic Plains of India: Growth, biochemical, physiological and yield attributes. *Atmos. Environ.* **2018**, *188*, 174–184.
4. Feng, H.; Chen, G.; Xiong, L.; Liu, Q.; Yang, W. Accurate Digitization of the Chlorophyll Distribution of Individual Rice Leaves Using Hyperspectral Imaging and an Integrated Image Analysis Pipeline. *Front Plant Sci.* **2017**, *8*, 1238.
5. Duan, B.; Fang, S.; Zhu, R.; Wu, X.; Wang, S.; Gong, Y.; Peng, Y. Remote Estimation of Rice Yield With Unmanned Aerial Vehicle (UAV) Data and Spectral Mixture Analysis. *Front Plant Sci.* **2019**, *10*, 204.
6. Jia, Y.; Zhang, H.; Zhang, X.; Su, Z. Quantitative analysis and hyperspectral remote sensing inversion of rice canopy SPAD in a cold region. *Eng. Agric.* **2022**, *42*(4), e20220030.
7. Kooistra, L.; Clevers, J.G.P.W. Estimating potato leaf chlorophyll content using ratio vegetation indices. *Remote Sens. Lett.* **2016**, *7*(6), 611–620.
8. Ban, S.; Liu, W.; Tian, M.; Wang, Q.; Yuan, T.; Chang, Q.; Li, L. Rice leaf chlorophyll content estimation using UAV-based spectral images in different regions. *Agronomy* **2022**, *12*(11), 2832.
9. Deng, L.; Mao, Z.; Li, X.; Hu, Z.; Duan, F.; Yan, Y. UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 124–136.
10. Zhang, S.; Zhao, G.; Lang, K.; Su, B.; Chen, X.; Xi, X.; Zhang, H. Integrated satellite, unmanned aerial vehicle (UAV) and ground inversion of the SPAD of winter wheat in the reviving stage. *Sensors* **2019**, *19*(7), 1485.
11. Wang, J.; Zhou, Q.; Shang, J.; Liu, C.; Zhuang, T.; Ding, J.; Xian, Y.; Zhao, L. Wang, W.; Zhou, G.; Tan, C.; Huo, Z. UAV- and machine learning-based retrieval of wheat SPAD values at the overwintering stage for variety screening. *Remote Sens.* **2021**, *13*(24), 5166.
12. Nio, S.A.; Pirade, M.; Ludong, D.P.M. Leaf chlorophyll content in North Sulawesi (Indonesia) local rice cultivars subjected to polyethylene glycol (PEG) 8000-induced water deficit at the vegetative phase. *Biodiversitas J. Biological Diversity.* **2019**, *20*(9), 2462–2467.



13. Meher; Shivakrishna, P.; Ashok Reddy, K.; Manohar Rao, D. Effect of PEG-6000 imposed drought stress on RNA content, relative water content (RWC), and chlorophyll content in peanut leaves and roots. *Saudi J Biol Sci.* **2018**, 25(2), 285–289.
14. Ziyang, Y.; Zhang, X.; Liu, H.; Zhang, Z.; Meng, L.; Han, Y.; Lu, L. Improving SPAD spectral estimation accuracy of rice leaves by considering the effect of leaf water content. *Crop Sci.* **2022**, 62(6), 2382–2395.
15. Yin, Q.; Zhang, Y.; Li, W.; Wang, J.; Wang, W.; Ahmad, I.; Zhou, G.; Huo, Z. Estimation of winter wheat SPAD values based on UAV multispectral remote sensing. *Remote Sens.* **2023**, 15(14), 3595.
16. Kumar, C.; Mubvumba, P.; Huang, Y.; Dhillon, J.; Reddy, K. Multi-stage corn yield prediction using high-resolution UAV multispectral data and machine learning models. *Agronomy* **2023**, 13(5), 1277.
17. Li, C.; Zhu, X.; Wei, Y.; Cao, S.; Guo, X.; Yu, X.; Chang, C. Estimating apple tree canopy chlorophyll content based on Sentinel-2A remote sensing imaging. *Sci Rep.* **2018**, 8(1), 3756.
18. Ta, N.; Chang, Q.; Zhang, Y. Estimation of apple tree leaf Chlorophyll content based on machine learning methods. *Remote Sens.* **2021**, 13(19), 3902.
19. An, G.; Xing, M.; He, B.; Liao, C.; Huang, X.; Shang, J.; Kang, H. Using machine learning for estimating rice Chlorophyll content from in situ hyperspectral data. *Remote Sens.* **2020**, 12(18), 3104.
20. Gao, B.C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, 58(3), 257–266.
21. Liu, L.Y.; Huang, W.J.; Pu, R.L.; Wang, J.H. Detection of internal leaf structure deterioration using a new spectral ratio index in the near-infrared shoulder region. *J. Integr. Agric.* **2014**, 13(4), 760–769.
22. Jordan, C.F. Derivation of leaf-area index from quality of light on the forest floor. *Ecol.* **1969**, 50(4), 663–666.
23. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, 83(1), 195–213.
24. Jiang, J.; Cai, W.; Zheng, H.; Cheng, T.; Tian, Y.; Zhu, Y.; Ehsani, R.; Hu, Y.; Niu, Q.; Gui, L.; Yao, X. Using Digital Cameras on an Unmanned Aerial Vehicle to Derive Optimum Color Vegetation Indices for Leaf Nitrogen Concentration Monitoring in Winter Wheat. *Remote Sens.* **2019**, 11(22), 2667.
25. Buschmann, C.; Nagel, E. In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *Int. J. Remote Sens.* **1993**, 14(4), 711–722.
26. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **2003**, 160(3), 271–282.
27. Daughtry, C.S.T.; Walthall, C.L.; Kim, M.S.; de Colstoun, E.B.; McMurtrey, J.E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* **2000**, 74(2), 229–239.
28. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, 48(2), 119–126.
29. Chen, J.M. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* **1996**, 22(3), 229–242.
30. Gamon, J.A.; Surfus, J.S. Assessing leaf pigment content and activity with a reflectometer. *New Phytol.* **1999**, 143(1), 105–117.
31. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the great plains with ERTS (Earth Resources Technology Satellite). Proceedings

- of 3<sup>rd</sup> Earth Resources Technology Satellite Symposium. NASA: Greenbelt, 1973, pp. 309–317.
32. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, 55(2), 95–107.
33. Gamon, J.A.; Serrano, L.; Surfus, J.S. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia*. **1997**, 112(4), 492–501.
34. Gitelson, A.; Merzlyak, M.N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol., B: Biology*. **1994**, 22(3), 247–252.
35. Broge, N.H.; Leblanc, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* **2001**, 76(2), 156–172.
36. Peñuelas, J.; Gamon, J.A.; Fredeen, A.L.; Merino, J.; Field, C.B. Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. *Remote Sens. Environ.* **1994**, 48(2), 135–146.
37. Kutylowska, M. K-nearest neighbours method as a tool for failure rate prediction. *Periodica Polytechnica Civil Eng.* **2018**, 62(2), 318–322.
38. Vincini, M.; Frazzi, E. Comparing narrow and broad-band vegetation indices to estimate leaf chlorophyll content in planophile crop canopies. *Precis. Agric.* **2010**, 12(3), 334–344.
39. Xu, X.; Gu, X.; Song, X.; Li, C.; Huang, W. Assessing rice chlorophyll content with vegetation indices from hyperspectral data. Computer and Computing Technologies in Agriculture IV (CCTA 2010). 2010.
40. Wu, C.; Niu, Z.; Tang, Q.; Huang, W. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agric. For. Meteorol.* **2008**, 148(8–9), 230–241.
41. Zarco-Tejada, P.J.; Miller, J.R.; Morales, A.; Berjón, A.; Agüera, J. Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops. *Remote Sens. Environ.* **2004**, 90(4), 463–476.
42. Karn, E.; De Leon, T.; Espino, L.; Al-Khatib, K.; Brim-DeForest, W. Phenotypic diversity of weedy rice (*Oryza sativa*f.spontanea) biotypes found in California and implications for management. *Weed Sci.* **2020**, 68(5), 485–495.
43. Tao, Y.; Zhang, J.; Song, L.; Cai, C.; Wang, D.; Wei, W.; Gu, X.; Yang, X.; Zhu, C. Projected elevated [CO<sub>2</sub>] and warming result in overestimation of SPAD-based rice leaf nitrogen status for nitrogen management. *Atmosphere* **2021**, 12(12), 1571.
44. Wan, W.; Zhao, Y.; Xu, J.; Liu, K.; Guan, S.; Chai, Y.; Cui, H.; Wu, P.; Diao, M. Reducing and delaying nitrogen recommended by leaf critical SPAD value was more suitable for nitrogen utilization of spring wheat under a new type of drip-irrigated system. *Atmosphere* **2022**, 12(10), 2331.

## **Estimating chlorophyll content in rice leaves using vegetation indices derived from multispectral UAV imagery**

**Lan Pham Thi<sup>1\*</sup>, Canh Le Van<sup>1</sup>**

<sup>1</sup> Hanoi University of Mining and Geology; phamthilan@humg.edu.vn;  
levancanh@humg.edu.vn

**Abstract:** Estimating chlorophyll content in rice leaves is crucial for monitoring crop growth and optimizing fertilizer management to enhance rice yield. In this study, chlorophyll content was estimated using vegetation indices (VIs) derived from UAV multispectral imagery. NDVI (Normalized Difference Vegetation Index), OSAVI (Optimized Soil Adjusted Vegetation Index), and NSRI (NIR Shoulder Region Index) were identified as optimal indices based on their correlation with chlorophyll content (SPAD values) measured in the field. Chlorophyll content was estimated using VIs through various machine learning models, including Linear Regression (LR), Random Forest (RF), K-Nearest Neighbor Regression (KNN), and Support Vector Regression (SVR). The combination of NSRI and OSAVI yielded the best results, with the LR model demonstrating higher accuracy compared to other models. The findings indicate that combining VIs from UAV multispectral imagery with the LR model improves the accuracy of chlorophyll content estimation in rice leaves. These results provide a reliable basis for applying this method to large-scale areas, guiding fertilizer management to optimize rice productivity.

**Keywords:** Chlorophyll content (SPAD value); Vegetation indices (VIs); UAV; Rice.