AATCC Review

06 July 2025: Received
24 August 2025: Revised
02 September 2025: Accepted
01 October 2025: Available Online

https://aatcc.peerjournals.net/

Original Research Article

Open Access

Regression and machine learning-based estimation of wheat leaf area index using field spectral NDVI

Siddhant Gupta[1], Rajeev Ranjan*[1,2], Anurag Tripathi[1], Krishna Kumar Patel[3], Satya Prakash Gupta[4], Chinmaya Kumar Sahu[5] and Ravi Kiran[1]

[1]*G.B. Pant University of Agriculture & Technology, Pantnagar, Uttarakhand, India*
[2]*Senior Scientist, ICAR-Central Marine Fisheries Research Institute, Kochi, Kerala, India*
[3]*Department of Agriculture, Mangalayatan University, Beswan, Aligarh, Uttar Pradesh, India*
[4]*Department of Agronomy, ANDUAT, Kumarganj, Ayodhya, Uttar Pradesh, India*
[5]*Junior Agrometeorologist, GKMS, AMFU- Bhubaneswar, OUAT, Odisha, India*

ABSTRACT

This study tackles the important issue of plant conservation by highlighting the need to monitor vegetation health and diversity in agroecosystems. These areas face threats from environmental pressures like pollution, habitat loss, and climate change, which impact plant stability. To improve vegetation assessment, this research uses remote sensing tools, specifically the Normalized Difference Vegetation Index (NDVI) and Leaf Area Index (LAI), to create strong LAI estimation models for wheat. Field experiments conducted over two crop seasons at G.B. Pant University employed a Split-Split-Plot Design with different sowing dates, irrigation levels, and varieties to capture a wide range of canopy conditions. A variety of regression models, including linear, exponential, logarithmic, power, and sigmoid models, were created and assessed. Machine-learning methods, such as Support Vector Regression and Random Forest Regression, were also explored to improve predictive accuracy. The modeling faced challenges due to NDVI saturation at high canopy density, seasonal changes in microclimate, and complex interactions among treatments. These issues required careful calibration and validation of the models. Results showed that non-linear models, especially sigmoid regression, best represented the NDVI-LAI relationship, achieving high coefficients of determination (R^2 = 0.8625 for training and 0.9213 for validation). Meanwhile, machine-learning models also performed well with complex data structures. Overall, the study provides valuable insights into crop monitoring using remote sensing, offering better tools for precision agriculture, efficient water use, and long-term plant biodiversity conservation.

Keywords: *NDVI, LAI, agroecosystem, Regression models, Machine Learning methods, remote sensing, precision agriculture, biodiversity conservation, etc.*

Introduction

Plant conservation is a vital global concern that has garnered considerable attention recently [25]. The diversity and health of plant species are essential to the world's ecosystems, whether they are terrestrial, aquatic, or marine [35]. At the most basic level, primary producers fulfill the function of effectively starting all food webs and providing the basis for many human-used ecological services, facilitating all ecological processes, and functioning in life support for humans. With this in mind, the conservation status of the variety of plants in any ecosystem becomes critical. The world is facing a decline in the variety of plants and with many causes of extinction of some plants and the endangerment of many more [1]. The extinction of plants causes a loss, and this is a problem considering the many functions plants provide, and the many benefits plants provide to humans, like the stabilization of soil, the provision of oxygen, and the sequestration and purification of water.

*Corresponding Author: ***Rajeev Ranjan***

DOI: https://doi.org/10.21276/AATCCReview.2025.13.04.626

The observation of conserved crops provides the foundation for stress detection in the plants in order to initiate loss prevention action when stress is caused by disease or hostile environmental conditions. Crop monitoring also helps to enhance crop yields, maximize the usage of resources, and ultimately ensure food security. Crop monitoring using sensing has progressed more rapidly due to technological innovation [30].

Remote sensing technology is largely unopposed as a means for monitoring crops. It is a non-intrusive means of access, contemporary and extensive in coverage. The physical structure of plant canopies plays an important role in regulating ecosystems, as it modifies energy, moisture, and gas exchange between the soil, the atmosphere, and vegetation. The canopy's complicated structure includes the leaf area index (LAI), which is the most important of the several structural variables, and is defined as the total area of leaves, on one side of the leaves, which is for photosynthesis, and is per unit area of the ground within the canopy [2][8]. It is a dimensionless measure, but it very well approximates the density of the photosynthetic apparatus and functions as a measure of the canopy structure and function. The LAI value most probably determines the fraction of solar energy incident on the canopy in the photosynthetically active domain (PAR, ~ 400-700 nm) that is

captured for photosynthesis [16][29]. Higher LAI values enhance the Gross Primary Production (GPP) of the ecosystem, the total carbon captured through the photosynthesis process, as well as the chances of chlorophyll molecules to intercept photons to power their photosynthesis. This association, frequently defined in terms of light extinction rules similar to the Beer-Lambert Law, regulates the degree of light energy fixed in chemical form within the canopy. Increasing LAI does augment light capture, but the influence ultimately reaches an asymptote at high levels of LAI due to growing self-shading by lower leaves. LAI has a strong impact on Net Primary Production (NPP), the net gain of biomass after subtracting respiration by plants, which is the foundation of the food web of the ecosystem and the potential for carbon sequestration. LAI dictates the total surface area available for transpiration, the process by which water vapor is released through microscopic leaf pores called stomata [4][26]. Consequently, it significantly influences total canopy conductance to water vapor and is a major determinant of ecosystem-level evapotranspiration (ET). This profoundly impacts the local and regional water balance, affecting soil moisture depletion rates, runoff generation, and atmospheric humidity. This control over water vapor flux also means LAI is crucial in partitioning incoming net radiation (solar energy absorbed by the surface) into latent heat flux (energy consumed during transpiration, leading to evaporative cooling) and sensible heat flux (energy that directly heats the air). Canopies with greater LAI tend to have more evaporative cooling and lower surface temperatures than thinly vegetated sites under the same conditions, thereby affecting microclimate and even regional climate regimes. Importantly, LAI is not a fixed characteristic but a dynamic measure of the complex interaction between vegetation and its environment [2][33]. It has clear phenological patterns, following seasonal cycles of growth, like the emergence of leaves in spring and autumn senescence in deciduous species. Its size combines the impacts of resource availability with restrictions in light, water, or nutrients, usually limiting maximum leaf area development or causing decreased leaf longevity. LAI is sensitive to environmental stress (e.g., drought, heat waves, and air pollution) and disturbance (e.g., fire, insects, disease, logging, or grazing), typically through physiological mechanisms such as stomatal closure, early senescence of leaves, or defoliation. Tracking these changes in LAI therefore delivers an informative, integrated measure of vegetation health, stress, phenological timing, and ecosystem dynamics and productivity potential.

The Normalized Difference Vegetation Index (NDVI) is a radiometric indicator of photosynthetic activity based on the differential absorption of red (600-700 nm) and near-infrared (NIR, 700-1300 nm) radiation by vegetation canopies [32]. The NDVI formula, NDVI = (ρNIR - ρred)/(ρNIR + ρred), in which ρ is surface reflectance, takes advantage of chlorophyll's high absorption of the red end of the spectrum and high NIR reflectance from leaf mesophyll caused by internal scattering [11]. NDVI displays asymptotic behaviors with important biophysical variables: It is closely related to leaf area index (LAI) up to thresholds of saturation (LAI ≈ 3-4) [3], displays a near-linear relationship with a fraction of absorbed photosynthetically active radiation (fAPAR) for NDVI < ≈ 0.7 [27], and acts as a substitute for canopy chlorophyll concentration after correction from structural effects [14]. Although extensively applied in ecological surveys and productivity assessment [34], NDVI has some limitations, such as saturation in closed canopies (NDVI > 0.7), soil background reflectance sensitivity (especially in arid environments), and atmospheric interference effects [22]. These limitations have spurred the creation of amended indices such as the Soil-Adjusted Vegetation Index (SAVI) [19], Enhanced Vegetation Index (EVI) [20], and Wide Dynamic Range Vegetation Index (WDRVI) [13] that include more correction factors but retain NDVI's underlying biophysical foundation.

The relationship between LAI and NDVI is one of the most basic forms of remote sensing relationships of vegetation, where LAI is proxied by NDVI spectrally, because both rely on canopies' light interception [3]. The theory is based on adapted versions of the Beer-Lambert law, where LAI affects the light attenuation through the canopy, and NDVI represents the spectrally signed absorption [27]. It has been shown that there is an empirical asymptotic and nonlinear relationship, having: (1) low LAI (0-2) and linear NDVI rise sensitivity of approximately 0.1-0.15 units per LAI unit increase; (2) mid LAI (2-4) diminishing sensitivity because of mutual shading; and (3) saturation at high LAI (>4-6) where added leaf layers contribute little to the NDVI [6]. The LAI-NDVI correlation is highly predictable by biome, where coniferous stands saturate quickly on NDVI, while broad leaf canopies take longer to saturate, showing the contraries in leaf angle distribution and clumping [9]. The LAI-NDVI correlation is also modulated by canopy structure factors such as the leaf inclination angle distribution (planophile vs. erectophile canopies) and foliage clumping index that influence light penetration and scattering [23]. Current methods overcome saturation constraints using hybrid techniques that blend NDVI with alternative spectral bands (e.g., red-edge reflectance) or physical approaches that include canopy radiative transfer theory [21].

Materials and Methods

Experimental site:

Wheat is widely grown in the *rabi* season throughout the Udham Singh Nagar (US Nagar) district of Uttarakhand. A field experiment was conducted at C-6, N.E. Borlaug Crop Research Centre, G.B. Pant University of Agriculture and Technology, Pantnagar, Udham Singh Nagar, which falls in the Tarai region of the Kumaon division. Geographically, the region is located between 28°53′ to 29°23′ N latitude and 78°45′ to 80°08′ E longitude. The area has a climate varying from subtropical to humid subtropical. Summers, particularly in April and May, are generally warm, followed by the southwest monsoon that delivers most of the annual rains from mid-June to late September. Winter arrives by late October and lasts through to March, yielding favorable thermal conditions for winter crops such as wheat. The Tarai belt is characterized by deep, fertile alluvial soils and rich water resources, making it extremely suitable for wheat production. The good agro-climatic conditions, combined with well-established irrigation facilities, provide for maximum growth and yield of the wheat crop in this region. This makes the region an important contributor to wheat production in Uttarakhand.

Experimental details

The experiment was conducted using a Split-Split-Plot Design with two sowing dates, three irrigation levels, and two varieties, resulting in 12 unique treatment combinations. For each treatment combination, biophysical parameters, precisely the Normalized Difference Vegetation Index (NDVI) and Leaf Area Index (LAI), were measured at five distinct growth stages corresponding to 30, 45, 60, 75, and 90 Days After Sowing (DAS).

This sampling strategy yielded a total of 60 data points (12 treatments × 5 time points) per crop season. NDVI and LAI values were derived from (greenseeker and ceptometer) and subsequently extracted and organized into a numerical format for analysis. The data utilized in this study were collected over two consecutive wheat crop seasons: 2023-24 and 2024-25.

Leaf Area Index

LAI is a dimensionless biophysical parameter defined as the total one-sided green leaf area per unit ground surface area (m^2 leaf area / m^2 ground area) [39]. The Ceptometer LP-80 was used to measure the Leaf Area Index (LAI) and Photosynthetically Active Radiation (PAR). Temperatures ranging from -30°C to 50°C and relative humidity levels up to 100% are suitable operating conditions for the Ceptometer. LAI is computed using the Ceptometer LP-80 as follows:

$$LAI = \frac{\left(\left(1-\frac{1}{2k}\right)fb\right)ln\tau}{A(1-0.047fb)} (1)$$

Where K is the extinction coefficient, which indicates the amount of radiation absorbed by the canopy at a specific solar zenith angle and canopy leaf angle distribution, and f_b is the beam fraction, which is determined by dividing the ratio of beam radiation by diffuse radiation. The Ceptometer LP-80 automatically determines f_b by comparing observed incidence PAR values to the solar constant, which is a known value of light energy from the sun (assuming clear sky conditions) at any particular time and location on the surface of the globe. Leaf absorption is denoted by A. The LP-80 has A set to 0.9 by default. The ratio of incident and transmitted PAR is denoted by τ. Measurements of incident PAR above the canopy and transmitted PAR close to the ground surface are used to compute this ratio (τ).

Normalized Difference Vegetation Index:

A portable GreenSeeker® Model 505 (Trimble Navigation Limited, Sunnyvale, CA, USA) active optical sensor that incorporated red light (671 ± 6 nm) and near-infrared (780 ± 6 nm) was used to measure the wheat canopy spectra. The carrying sensor probe was positioned above the crop at a height of around 0.8 meters above the wheat canopy. The light beam was perpendicular to the seed row, while the sensor route ran parallel to the seed rows. Three rows made up each cell, and each row was measured using three replications. Each plot was represented by the average values. NDVI quantifies vegetation health using reflectance in the red (R) and near-infrared (NIR) bands [32]:

$$NDVI = \frac{\rho NIR-\rho Red}{\rho NIR+\rho Red} (2)$$

Where, ρNIR is Reflectance in the NIR band and ρRed is Reflectance in the red band.

Data Structuring and Preprocessing:

For each crop season (2023-24 and 2024-25), the acquired NDVI and LAI data were structured into separate matrices. Each matrix consisted of 12 rows (treatment combinations) and 5 columns (DAS time points: 30, 45, 60, 75, 90). For regression analysis and model building that can be used for all growth stages measured, these two-dimensional matrices were converted into one-dimensional vectors (arrays). For each year 2023-24, the NDVI matrix was reduced to a single vector, NDVI_2023-21_flat, and the same was done for the LAI matrix as LAI_2023-24_flat.

Each resulting vector had 60 corresponding NDVI or LAI measurements, combining the values over all treatments and DAS stages. This methodology combines data from various phenological stages to determine empirical relationships between NDVI and LAI.

Regression Model Development and Comparison:

Using the dataset from the 2023-24 crop season as the training set (NDVI_2023-24_flat as input X, LAI_2023-24_flat as target Y), several regression models were developed and compared to identify the most suitable functional form for predicting LAI from NDVI. The models explored included linear, ML, and various non-linear forms. The goodness-of-fit for each model on the training data was evaluated using the Coefficient of Determination ®^{2}).

Regression Model:

A statistical method for determining a quantifiable relationship between a dependent variable and one or more independent variables is called a regression model.

Linear regression: Linear regression is a fundamental statistical method used to model the relationship between a dependent variable (Y) and one or more independent variables (X) by fitting a straight line to the observed data.

$$Y = \beta o + \beta 1X + \varepsilon (3)$$

Where Y is the dependent variable, X is the independent (predictor) variable, β_o is the intercept (value of Y when X = 0), β1 is the slope (rate of change of Y with respect to X), and ε is the random error (residual).

Non-linear regression: When a straight line cannot sufficiently depict the connection between variables, non-linear regression models are employed. Rather, the model incorporates parameters in non-linear combinations, and the data exhibits a curved trend.

$$Y = f(X, \beta) + \varepsilon (4)$$

Where f(X, β) is a non-linear function in terms of the parameters β, ε is a random error (residual).

Exponential Regression: when the rate of change in the dependent variable increases or decreases exponentially with the independent variable.

$$Y = \beta o.\, e^{\beta 1X} (5)$$

Where Y is the dependent variable, X is the independent (predictor) variable, β_o is the intercept (value of Y when X = 0), β1 is the slope and e is Euler's number (~2.718).

Power Regression: When the dependent variable scales as a power function of the independent variable, often used in allometric relationships.

$$Y = \beta o.\, X^{\beta 1} (6)$$

Where Y is the dependent variable, X is the independent (predictor) variable, β_o is the scale parameter, and β1 is the Exponent that determines curvature.

Logarithmic Regression: Models a rapid change in Y with small increases in X, followed by slower change, good for diminishing returns.

$$Y = \beta o + \beta 1.\ln(X) (7)$$

Where ln(X) is the Natural logarithm of X

Sigmoid (Logistic) Regression: An S-shaped curve often used to describe growth that begins slowly, increases rapidly, then plateaus - suitable for saturating biological processes.

$$Y = \frac{K}{1+e^{-(\beta o+\beta 1X)}} (8)$$

Where β_o, $\beta 1$ are scale parameters, K is the carrying capacity (maximum value), and e is Euler's number.

Machine learning regression:

Machine learning regression is a supervised learning method that uses input information to predict continuous outputs. It is widely used to model and forecast numerical data in a variety of fields, including environmental science, agriculture, and finance. While these models are powerful, they often face challenges such as overfitting, especially when dealing with high-dimensional data, and require meticulous hyperparameter tuning to achieve optimal performance. Recent advancements have introduced innovative approaches to enhance regression models.

Support Vector Regression (SVR): Support Vector Regression (SVR) is an extension of the Support Vector Machine (SVM) algorithm, designed specifically for regression problems. In contrast to conventional regression models that try to minimize the difference between actual and predicted values, SVR tries to fit the best function within a tolerance margin (ε), referred to as the epsilon-insensitive zone. The concept is that deviations within this range are not penalized, but those outside are minimized through a regularized loss function [36]. SVR operates by transforming input data into a high-dimensional feature space through a kernel function, usually the Radial Basis Function (RBF), and then applying a linear regression model in this transformed space. This allows SVR to handle non-linear relationships in the data effectively. The strength of SVR is its power to regulate the model's complexity using hyperparameters like the regularization parameter (C), the value of epsilon (ε), and the parameters of the kernel (e.g., gamma in RBF kernel). One of the prominent strengths of SVR is its strength against overfitting, particularly for small or medium datasets with high noisy cases. The model is sensitive to kernel selection and hyperparameter tuning, which usually calls for cross-validation and optimization methods.

Random Forest Regression (RF): Random Forest (RF) Regression is an ensemble method that creates a large number of decision trees when trained and produces the mean prediction of these trees for regression tasks. Random Forest is part of the family of bagging algorithms, whereby numerous models are separately trained on randomly selected subsets of the data, both along the rows (bootstrap sampling) and along the columns (random feature selection) [7]. This ensemble method lessens the variance observed in single decision trees, which are vulnerable to overfitting, particularly when dealing with noisy data. Random Forest can easily manage non-linear relations and interactions among features without the need for heavy preprocessing or feature engineering. It is parallelizable and scalable and works effectively with high-dimensional datasets, and thus it is very well-suited for real-world applications in agriculture, environmental modeling, and remote sensing [43]. A drawback of RF is that it will overfit if the model is too dense or not pruned or regularized. It is also less interpretable than models with fewer parameters, such as linear regression or decision trees.

Although variable importance scores give some idea of the importance of each feature, the actual decision path is difficult to follow when there are hundreds of trees [18]. Despite these constraints, RF is a robust and useful regression tool whenever predictive performance over model interpretability is more critical. Its resilience to outliers, missing data, and irrelevant attributes contributes to its popularity in applied research and decision-making based on data.

The predictive capability of the selected linear model was then rigorously evaluated using the independent dataset from the 2024-25 crop season. The NDVI_2024-25_flat array was input to the chosen linear equation to generate predicted LAI values (LAI_predicted_2024-25).

These predicted LAI values were compared against the actual observed LAI values (LAI_2024- 25_flat) from the 2024-25 season. The model's validation performance was assessed using:

Coefficient of Determination (R^2): Measuring the proportion of variance in observed LAI explained by the model's predictions on the validation set.

Root Mean Square Error (RMSE): Quantifying the average prediction error in LAI units on the validation set.

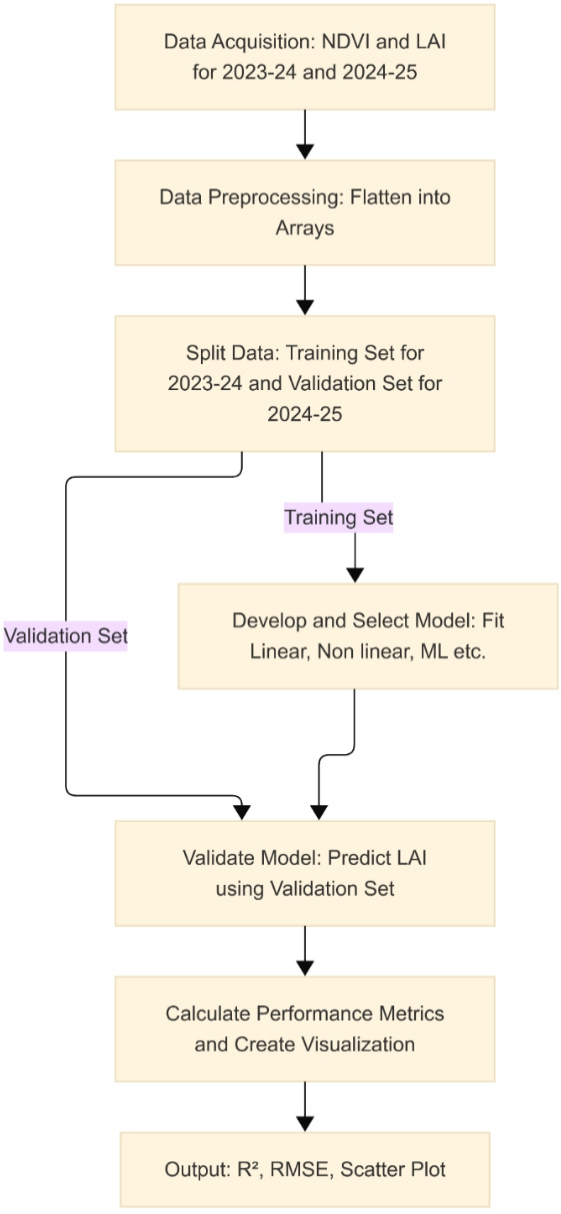

Fig.1: Flowchart of the methodology

Software

All data processing, statistical analysis, model development, comparison, validation, and visualization were performed using the Python programming language (Google Colab) with its scientific computing libraries, including NumPy for numerical operations, Pandas for data manipulation, scikit-learn for regression modeling and evaluation, and Matplotlib/Seaborn for data visualization.

Results and discussion

The temporal pattern of the Normalized Difference Vegetation Index (NDVI) and Leaf Area Index (LAI) between treatment combinations and over two seasons of growth (2023-24 and 2024-25) is depicted in the fig.2 NDVI, one of the best-known indicators of greenness and health of vegetation, is computed from differential reflectance of red and near-infrared (NIR) radiation, indicating chlorophyll concentration and photosynthetic activity [32].

The trends in NDVI in all treatments and years generally reflect a rise corresponding to vegetative development, peaking at around 75 Days After Sowing (DAS), reflecting maximum cover of the canopy during the reproductive stage, followed by a drop towards 90 DAS when plants senesce. Increasing NDVI typically reflects healthier, more vigorous cover.

Leaf Area Index (LAI), a measure of the total one-sided leaf area per unit ground area, is another important parameter affecting photosynthesis and overall plant productivity [39]. Like NDVI, the LAI patterns in the graphs show an increase during canopy growth, peaking at about 75 DAS, and then declining as plants mature or senesce. Comparison of NDVI and LAI patterns across different treatment combinations and between the two growing seasons provides a scientific basis for evaluating the efficiency of diverse farming practices and the influence of inter-annual variations in environmental conditions on crop growth and productivity.

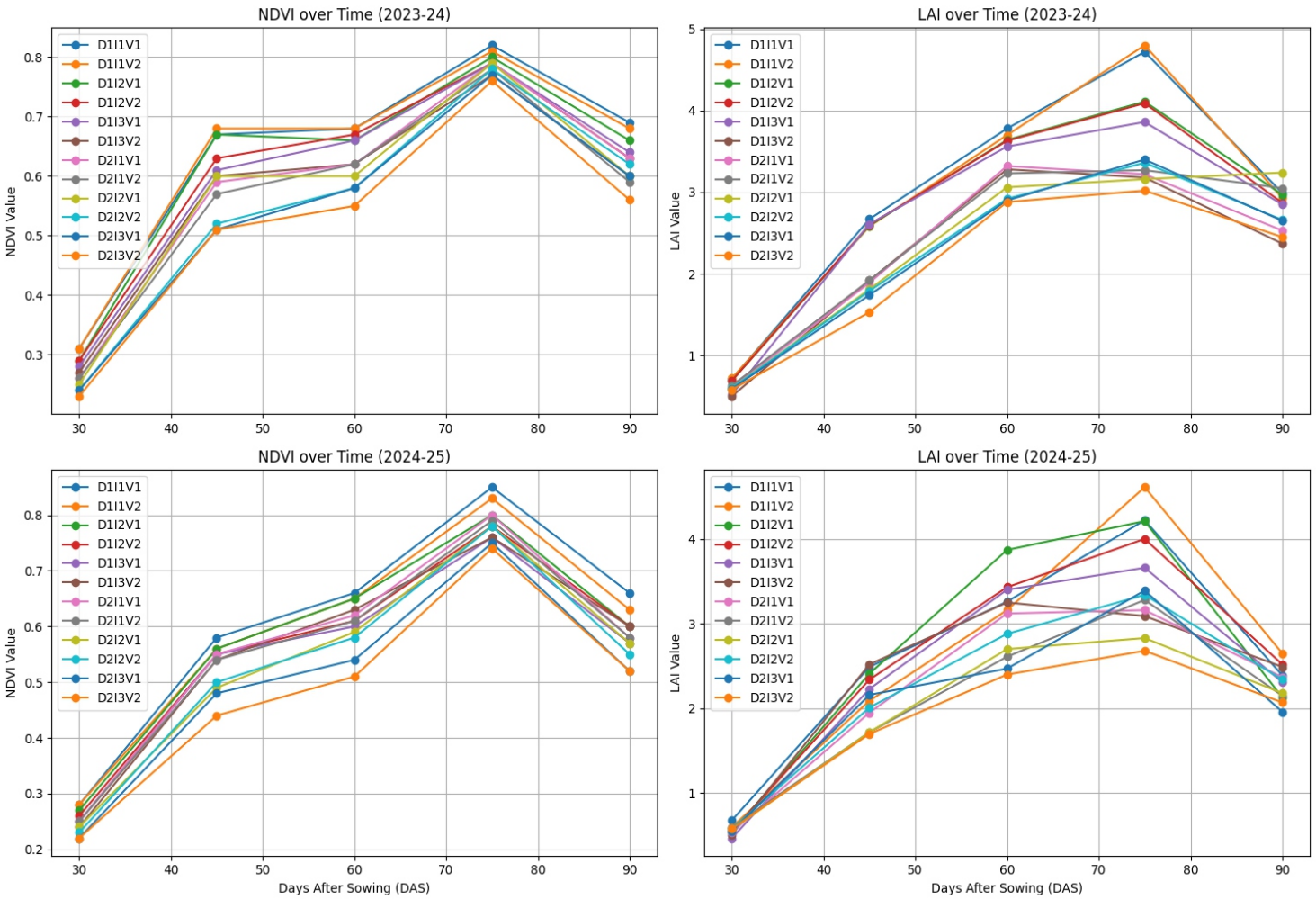

Fig. 2: NDVI and LAI Trends Over Time

This study developed regression models to predict the Leaf Area Index (LAI) from NDVI values recorded at five growth stages of wheat, using observations from one crop season for model training and a subsequent season for validation.

Linear regression model:
The linear regression model fitted to the 2023-24 data resulted in the following equation: $LAI = 6.08 * NDVI - 1.02$ (9)

The model was validated using data from the 2024-25 season, where predicted Leaf Area Index (LAI) values were compared with observed values. The performance statistics, Coefficient of Determination (R^2) of 0.92 and Root Mean Square Error (RMSE) of 0.30, reflect a good linear relationship between Normalized Difference Vegetation Index (NDVI) and LAI at varying growth stages and treatment levels. An R^2 value of 0.92 indicates that the variation in LAI can be explained by the variation in NDVI to the tune of 92%, and low RMSE ensures high predictive ability, validating the model's use for estimating LAI from NDVI data.
The linearity of the relationship between NDVI and LAI is consistent with findings from prior studies that highlight NDVI's robust performance for estimating canopy attributes under moderate to dense vegetation cover [5][14] The slope of 6.08 suggests a steep and positive increase in LAI with increasing NDVI, which is expected since NDVI values tend to increase with rising chlorophyll content and leaf density. However, the slightly negative intercept (-1.02) reflects potential limitations of linear models at very low NDVI values, where soil background or sparse canopies may influence reflectance more strongly [38]. Though the general performance of the linear model is satisfactory, it must be noted that NDVI does saturate with high LAI values, especially in close canopies, tending to decrease sensitivity over a certain point [42].

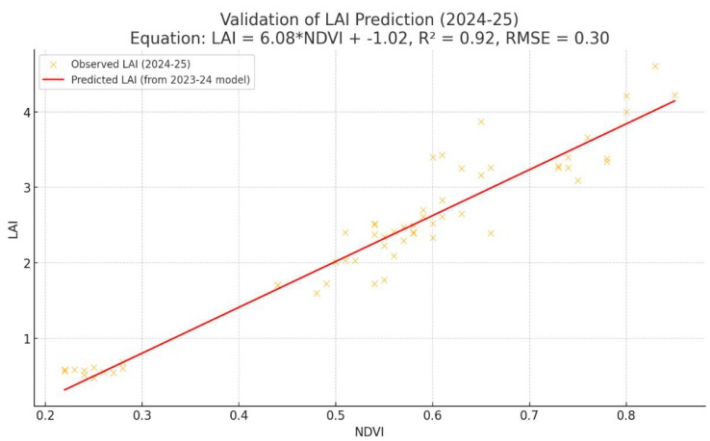

Fig. 3: Validation of predicted LAI using linear regression

Exponential regression model:
To estimate LAI from NDVI values taken at various stages of development and treatments, a regression model with an exponential form was established to fit the potentially non-linear relationship between vegetation indices and canopy biophysical variables. It is most applicable when LAI increases exponentially with NDVI, a sign of developing canopy and rapid biomass accumulation. The best-fit exponential regression equation obtained from the 2023-24 data is:

$$LAI = 0.5053.\, e^{2.6201.NDVI} \quad (10)$$

The model exhibited a coefficient of determination (R^2) of 0.8159. This indicates that approximately 81.6% of the variation in LAI could be explained by NDVI using the exponential relationship. Although this is slightly lower than the linear model performance (R^2 = 0.92), the exponential model is biophysically meaningful, especially in early growth stages when LAI tends to increase rapidly with increasing NDVI. Exponential functions have been widely used in remote sensing to model LAI-NDVI relationships due to their ability to capture non-linear dynamics in vegetation growth [6][17]. At low NDVI values, LAI tends to change slowly, but as canopy density increases, small increases in NDVI can lead to rapid increases in LAI, an effect well captured by exponential models. However, it is also important to note that NDVI saturation at high LAI levels may reduce sensitivity, which may necessitate using alternative indices or more flexible models [42].
The fitted exponential curve (blue line in Fig. 4) aligns well with both observed and validation data points, particularly in the middle NDVI range (0.4-0.7), supporting its applicability across key growth stages. Its visual fit reinforces the appropriateness of non-linear models for complex biological systems like crop canopies.

Logarithmic regression model:
To explore the non-linear relationship between Normalized Difference Vegetation Index (NDVI) and Leaf Area Index (LAI), a logarithmic regression model was developed using observational data from the 2023-24 wheat season and validated on an independent dataset from the 2024-25 season.

$$LAI = 4.1829 + 2.7671.\, ln(NDVI) \quad (11)$$

The Coefficient of Determination (R^2) for the logarithmic model was 0.8349, indicating that approximately 83.5% of the variation in LAI was explained by the natural logarithm of NDVI. This reflects a strong statistical fit across diverse treatments, growth stages, and seasonal conditions. The logarithmic relationship reveals that as NDVI increases, the rate of increase in LAI gradually diminishes, a trend that makes physiological sense. In the early stages of crop growth, NDVI rises quickly with increasing canopy cover, but as the canopy nears closure, NDVI begins to saturate. As a result, even small increases in NDVI during later stages correspond to much smaller gains in LAI. This pattern is typical of remotely sensed vegetation indices under moderate to high biomass conditions and aligns with earlier findings by [5] and [6], who also reported logarithmic and asymptotic trends in NDVI-LAI relationships due to spectral saturation at high leaf densities.

Power regression model:
To better describe the non-linear relationship between Normalized Difference Vegetation Index (NDVI) and Leaf Area Index (LAI) in wheat, a power regression model was formulated. This was calibrated with data from the 2023-24 season and tested with observations from the 2024-25 season from several treatments and stages of growth.

$$LAI = 5.5553.\, \mathrm{NDVI}^{1.5212} \quad (12)$$

Coefficient of Determination (R^2) 0.8557 suggests that about 85.6% of the observed LAI variation can be explained by power-transformed NDVI values, reflecting a very good fit of the model. The power model predicts a non-linear, accelerating relationship between LAI and NDVI. In particular, the fact that the exponent (1.5212) is larger than 1 indicates that increases in NDVI are linked to disproportionately larger increases in LAI, particularly at larger NDVI values. This is consistent with intense vegetative growth stages where NDVI is responsive to growing biomass and leaf area, especially under optimal irrigation and nutrient conditions. Power functions are often used in remote sensing processes where plant development is not linear but scale-dependent, as noted by [31]. In contrast to linear or logarithmic models, the power model fits better in intermediate-to-high NDVI ranges, where LAI grows very fast.

Sigmoidal (logistic) regression model:

To better describe the biological saturation behavior of the NDVI–LAI relationship in wheat, a sigmoidal (logistic) regression model was formulated. This model was trained on 2023-24 crop season field data and validated using 2024-25 data from several treatments and phenological stages.

$$LAI = \frac{4.541}{1+exp^{-(6.5915.(NDVI-0.552))}} \quad (13)$$

Coefficient of Determination (R^2) 0.8625 is the highest among all tested models, indicating that 86.3% of the variability in LAI is explained by the NDVI through this sigmoidal function. Sigmoid functions are commonly used in crop modeling and remote sensing to represent phenological progress and biomass accumulation. The logistic form is particularly suited for modeling vegetation indices and LAI due to its ability to handle early-stage sensitivity and late-stage saturation [28][37]. This model offers the best compromise between accuracy and biological realism, making it ideal for use in simulations, remote sensing-based estimations, and decision-support tools.

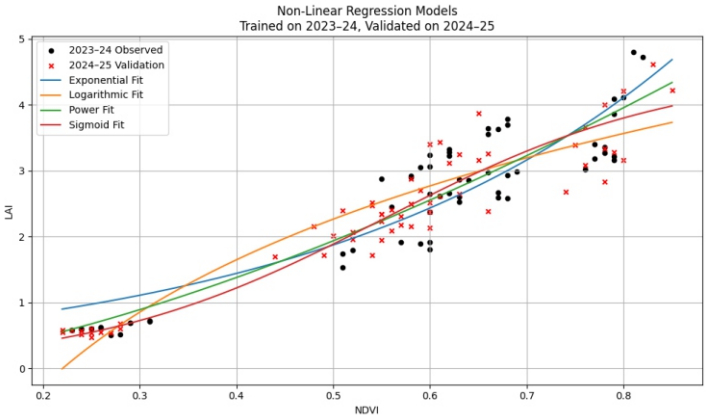

Fig.4: Training and validation of non-linear models

Comparison of Different Regression Models for Estimating LAI from NDVI:

Estimating LAI accurately and non-destructively through spectral vegetation indices, such as the Normalized Difference Vegetation Index (NDVI), has become a standard approach in precision agriculture and crop monitoring. Nevertheless, because of the non-linear NDVI-LAI relationship, especially NDVI saturation at high canopy densities, it is critical to determine suitable regression models that can account for this complexity [13]. To compare model performance in LAI estimation from NDVI, five linear, exponential, logarithmic, power, and sigmoid regression equations were tested using the 2023-24 wheat season data (training) and validated against the 2024-25 data. The fit of the models was compared using the coefficient of determination (R^2) and root mean square error (RMSE). The comparison is shown in Table 1

Table 1. Comparison of Different Regression Models for Estimating LAI from NDVI

Regression Model	Equation	R^2 (Train)	R^2 (Valid)	RMSE (Train)	RMSE (Valid)
Linear	$LAI = 6.08 * NDVI - 1.02$	0.8130	0.9200	0.3954	0.3000
Exponential	$LAI = 0.5053.e^{2.6201.NDVI}$	0.8159	0.9121	0.4714	0.3115
Logarithmic	$LAI = 4.1829 + 2.7671.ln(NDVI)$	0.8349	0.9168	0.4362	0.2946
Power	$LAI = 5.5553.NDVI^{1.5212}$	0.8557	0.9185	0.4063	0.2819
Sigmoid	$LAI = \frac{4.541}{1+exp^{-(6.5915.(NDVI-0.552))}}$	0.8625	0.9213	0.3945	0.2728

Among the tested models, the sigmoid equation exhibited the highest accuracy, with an R^2 of 0.8625 and RMSE of 0.3945 in the training dataset, and an even stronger performance in the validation dataset (R^2 = 0.9213, RMSE = 0.2728). This aligns with previous findings that emphasize the suitability of sigmoidal or logistic functions in capturing LAI dynamics, especially under high NDVI conditions where saturation effects are prominent [40].

The power and logarithmic models also performed well, with R^2 values exceeding 0.85 in training and 0.91 in validation. These models effectively capture the curvilinear relationship between NDVI and LAI, particularly at intermediate growth stages [24][41]. The exponential model, although slightly less accurate, still outperformed the linear model, indicating the importance of adopting non-linear approaches for better representation of the biophysical processes governing canopy development [10].

The linear model, despite its simplicity and interpretability, underestimates LAI at higher NDVI levels, supporting previous literature that cautions against its use in dense canopies due to NDVI saturation [13]. The results demonstrate that non-linear regression models, particularly the sigmoid function, offer a more robust and biologically consistent framework for estimating LAI from NDVI across different phenological stages in wheat. These findings are critical for improving the accuracy of remote sensing-based crop models and decision-support systems for irrigation management.

Comparison of Support Vector Regression and Random Forest Regression:

In recent years, machine learning models like Support Vector Regression (SVR) and Random Forest Regression (RF) have become prominent due to their capacity to capture complex relations in regression problems. SVR uses the kernel trick, specifically the Radial Basis Function (RBF) kernel, to deal with non-linear data by projecting it into higher-dimensional spaces. Performance of SVR on both the training (RMSE = 0.4519, R^2 = 0.8435) and validation sets (RMSE = 0.3896, R^2 = 0.8686) is excellent in generalization, though with slight improvement in

the validation set. SVR is very capable of avoiding overfitting, and so is apt to be used for smaller datasets or when high generalization is a necessity but runs high in computational cost, particularly with large datasets. Conversely, Random Forest Regression (RF), which is an ensemble learning algorithm that aggregates several decision trees, works well on both training (RMSE = 0.3069, R^2 = 0.9278) and validation sets (RMSE = 0.4129, R^2 = 0.8523), though there is a slight decrease in performance on the validation set. This indicates a possible overfitting. RF is especially effective at handling high-dimensional, complex data and is less likely to overfit compared to single decision trees. Nevertheless, it may be computationally demanding and non-interpretable, particularly when there are many trees.

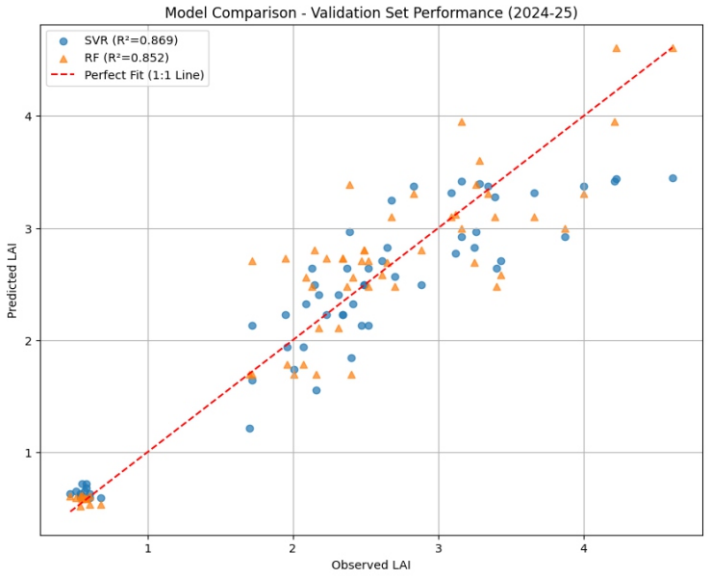

Fig.5: Validation of Machine Learning regression models (Support Vector regressor and Random Forest)

Both models excel in different contexts. SVR offers robust performance with smaller datasets and maintains generalization, making it suitable for applications requiring stability, such as financial modeling or time-series predictions. RF, on the other hand, performs better with large, complex datasets and is capable of capturing intricate patterns, making it ideal for environmental and agricultural modeling. Recent research highlights SVR's advantages in avoiding overfitting, while RF is preferred in high-dimensional applications where capturing complex feature interactions is critical [43]. The choice between SVR and RF depends on the dataset size, complexity, and the need for interpretability versus predictive power.

Conclusion

This study highlights the importance of modeling and validating regression techniques to estimate wheat Leaf Area Index (LAI) using field-based Normalized Difference Vegetation Index (NDVI) measurements. As plant conservation and biodiversity become increasingly critical in the face of environmental challenges, effective monitoring of crop health is essential for ensuring food security. The study utilized a comprehensive dataset collected over two crop seasons, employing various regression models, including linear, exponential, logarithmic, power, and sigmoid functions. The findings revealed that non-linear models, particularly the sigmoid regression, provided the most accurate estimates of LAI, demonstrating the importance of capturing the complex relationships between NDVI and LAI, especially in dense canopies. Additionally, the comparison of machine learning approaches, such as Support Vector Regression (SVR) and Random Forest Regression (RF), further emphasized the potential of these advanced techniques in handling complex datasets. Overall, this research provides valuable insights into remote sensing-based crop monitoring, enhances decision-making in agricultural practices, and supports sustainable farming efforts.

Future Scope

Future studies should evaluate the proposed NDVI-LAI models across wider agroecological regions, wheat cultivars, and management conditions to improve their robustness and transferability. Integrating additional vegetation indices, thermal data, and UAV- or satellite-based multispectral imagery may help reduce NDVI saturation and enhance LAI prediction accuracy. Advanced machine-learning and deep-learning methods also offer potential for developing more automated, real-time LAI estimation systems. Further integration of remote sensing outputs with crop simulation models could support climate-smart decision-making, particularly under moisture and heat stress. Such advancements will strengthen precision agriculture practices and contribute to sustainable crop monitoring and resource-use efficiency.

Acknowledgement: The authors are grateful to the Department of Agrometeorology, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India, for providing laboratory facilities in carrying out this research.

Funding: The research work was conducted without external funding; however, it benefited from the support of a university fellowship and resources.

Conflict of Interests: The authors declare that there is no conflict of interest in this article.

Author Contributions:

Siddhant Gupta: Conceptualization, Field investigation, data collection, analysis, modelling, and writing the original draft.; Rajeev Ranjan: Methodology, analysis support, validation, Supervision and project administration; Anurag Tripathi: Field investigation and data collection; Krishna Kumar Patel & Satya Prakash Gupta: Methodology and manuscript review; Chinmaya Kumar Sahu: Data curation and validation support. Ravi Kiran: Supervision, project administration.

References

1. Adebayo, O. (2019). Loss of biodiversity: The burgeoning threat to human health. *Ann Ibadan Postgrad Med*, 17,1-3

2. Asner, G. P., Scurlock, J. M., & A. Hicke, J. (2003). Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global ecology and biogeography*, *12*(3), 191-205. https://doi.org/10.1046/j.1466-822X.2003.00026.x

3. Asrar, G. Q., Fuchs, M., Kanemasu, E. T., & Hatfield, J. L. (1984). Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat 1. *Agronomy journal*, *76*(2), 300-306. https://doi.org/10.2134/agronj1984.00021962007600020024x

4. Baldocchi, D. D., Hincks, B. B., & Meyers, T. P. (1988). Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology*, *69*(5), 1331-1340. https://doi.org/10.2307/1941811

5. Bannari, A., Morin, D., Bonn, F., & Huete, A. (1995). A review of vegetation indices. *Remote sensing reviews*, *13*(1-2), 95-120. https://doi.org/10.1080/02757259509532298

6. Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote sensing of environment*, *35*(2-3), 161-173. https://doi.org/10.1016/0034-4257(91)90009-U

7. Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

8. Chen, J. M., & Black, T. A. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell & Environment*, *15*(4), 421-429. https://doi.org/10.1016/0168-1923(91)90074-z

9. Chen, J. M., & Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote sensing of Environment*, *55*(2), 153-162. https://doi.org/10.1016/0034-4257(95)00195-3

10. Verma, B., Prasad, R., Srivastava, P. K., Yadav, S. A., Singh, P., & Singh, R. K. (2022). Investigation of optimal vegetation indices for retrieval of leaf chlorophyll and leaf area index using enhanced learning algorithms. *Computers and electronics in agriculture*, *192*, 106581. https://doi.org/10.3390/rs14030615

11. Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. (1965). Spectral properties of plants. *Applied optics*, *4*(1), 11-20. https://doi.org/10.1364/AO.4.000011

12. Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology*, *161*(2), 165-173. https://doi.org/10.1078/0176-1617-01176

13. Gitelson, A. A., Peng, Y., Arkebauer, T. J., & Schepers, J. (2014). Relationships between gross primary production, green LAI, and canopy chlorophyll content in maize: Implications for remote sensing of primary production. *Remote Sensing of Environment*, *144*, 65-72.

14. Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical research letters*, *30*(5). https://doi.org/10.1029/2002GL016450

15. Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment*, *58*(3), 289-298. https://doi.org/10.1016/S0034-4257(96)00072-7

16. Gower, S. T., Kucharik, C. J., & Norman, J. M. (1999). Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote sensing of environment*, *70*(1), 29-51. https://doi.org/10.1016/S0034-4257(99)00056-5

17. Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote sensing of environment*, *90*(3), 337-352. https://doi.org/10.1016/j.rse.2003.12.013

18. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning (2nd ed.)*. Springer.

19. Huete, A.R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*. 25(3), 295-309. https://doi.org/10.1016/0034-4257(88)90106-X

20. Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment*, *83*(1-2), 195-213. https://doi.org/10.1016/S0034-4257(02)00096-2

21. Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote sensing of Environment*, *112*(10), 3833-3845. https://doi.org/10.1016/j.rse.2006.02.016

22. Kaufman, Y. J., & Tanre, D. (1996). Strategy for direct and indirect methods for correcting the aerosol effect on remote sensing: from AVHRR to EOS-MODIS. *Remote sensing of Environment*, *55*(1), 65-79. https://doi.org/10.1016/0034-4257(95)00193-X

23. Kucharik, C. J., Norman, J. M., & Gower, S. T. (1999). Characterization of radiation regimes in nonrandom forest canopies: theory, measurements, and a simplified modeling approach. *Tree physiology*, *19*(11), 695-706. https://doi.org/10.1093/treephys/19.11.695

24. Ma, J., Wang, L., & Chen, P. (2022). Comparing different methods for wheat LAI inversion based on hyperspectral data. *Agriculture*, *12*(9), 1353. https://doi.org/10.3390/agriculture12091353

25. Maunder, M. (2013). *Plant conservation. In: Levin SA (ed) Encyclopedia of biodiversity*. Academic Press, 76-89

26. Monteith, J., & Unsworth, M. (2013). *Principles of environmental physics: plants, animals, and the atmosphere*. Academic press.

27. Myneni, R. B., & Williams, D. L. (1994). On the relationship between FAPAR and NDVI. *Remote Sensing of Environment*, *49*(3), 200-211. https://doi.org/10.1016/0034-4257(94)90016-7

28. Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., ... & Running, S. W. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote sensing of environment*, *83*(1-2), 214-231. https://doi.org/10.1016/S0034-4257(02)00074-3

29. Norman, J.M., Campbell, G.S. (1989). Canopy structure. In: Pearcy, R. W., Ehleringer, J. R., Mooney, H., & Rundel, P. W. (Eds.). *Plant physiological ecology: Field methods and instrumentation.* Springer Science & Business Media, 301-325.

30. Omia, E., Bae, H., Park, E., Kim, M. S., Baek, I., Kabenge, I., & Cho, B. K. (2023). Remote sensing in field crop monitoring: A comprehensive review of sensor systems, data analyses and recent advances. *Remote Sensing*, *15*(2), 354. DOI: 10.3390/rs15020354

31. Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote sensing of environment*, *48*(2), 119-126. https://doi.org/10.1016/S0034-4257(03)00007-7

32. Rouse Jr, J. W., Haas, R. H., Deering, D. W., Schell, J. A., & Harlan, J. C. (1974). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation* (No. E75-10354).

33. Running, S. W., & Coughlan, J. C. (1988). A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production processes. *Ecological modelling*, *42*(2), 125-154. https://doi.org/10.1016/0304-3800(88)90112-3

34. Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*, *54*(6), 547-560. https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2

35. Sandifer, P. A., Sutton-Grier, A. E., & Ward, B. P. (2015). Exploring connections among nature, biodiversity, ecosystem services, and human health and well-being: Opportunities to enhance health and biodiversity conservation. *Ecosystem services*, *12*, 1-15.

36. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

37. Tucker, C. J., & Sellers, P. J. (1986). Satellite remote sensing of primary production. *International journal of remote sensing*, *7*(11), 1395-1416. https://doi.org/10.1080/01431168608948944

38. Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote sensing of environment*, *115*(12), 3468-3478. https://doi.org/10.1016/j.rse.2011.08.010

39. Watson, D. J. (1947). Comparative physiological studies on the growth of field crops: I. Variation in net assimilation rate and leaf area between species and varieties, and within and between years. *Annals of botany*, *11*(41), 41-76.

40. Müller, J., Behrens, T., & Diepenbrock, W. (2006). Use of a new sigmoid growth equation to estimate organ area indices from canopy area index in winter oilseed rape (Brassica napus L.). *Field crops research*, *96*(2-3), 279-295. https://doi.org/10.1016/j.fcr.2005.07.009

41. Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors*, *2017*(1), 1353691. https://doi.org/10.1080/01431161.2017.1346410

42. Zou, X., & Mõttus, M. (2017). Sensitivity of common vegetation indices to the canopy structure of field crops. *Remote Sensing*, *9*(10), 994. https://doi.org/10.3390/rs9100994

43. Zhang, L., Huettmann, F., Zhang, X., Liu, S., Sun, P., Yu, Z., & Mi, C. (2019). The use of classification and regression algorithms using the random forests method with presence-only data to model species' distribution. *MethodsX*, *6*, 2281-2292. 10.1016/j.mex.2019.09.035 https://doi.org/