INDIAN INSTITUTE OF REMOTE SENSING
Indian Space Research Organisation
Department of Space, Govt. of India

Project Report on LAI estimation of Wheat using empirical and ML Algorithms

M.Sc Agriculture Analytics

Submitted by:

Harshilkumar Ponkiya (TS90705)

Nidhi Gusai (TS90707)

Submitted to:

Dr. N.R.Patel

Agriculture and Soil Department, IIRS-ISRO

Mrs. Shweta Pokhariyal

Agriculture and Soil Department, IIRS-ISRO

Contents

List of Figures

Abstract

This study aimed to estimate the Leaf Area Index (LAI) of wheat crops in the Palampur Subdistrict of Kangra District, Himachal Pradesh, using various machine learning algorithms. The study's findings suggest that multiple linear regression (MLR) combined with key vegetation indices such as NDVI, CI, Osavi, EVI, and GNDVI, can estimate LAI accurately. Among different machine learning models, Linear Regression had the lowest RMSE of 0.3934, indicating its relatively accurate predictions. Additionally, it achieved a high R^2 Score of 0.6874, indicating a strong correlation between predicted and actual values. The high Index of Agreement (0.8782) for MLR underscores its robustness in agreement between predicted and observed LAI values, indicating its potential for practical implementation in crop management decision-making processes.

Furthermore, the study shows significant relationships between LAI and various vegetation indices, reaffirming their importance in remote sensing-based LAI estimation. Except for Red Edge NDVI, all investigated indices demonstrate relevance, emphasizing their collective utility in monitoring wheat crop dynamics. The data for all the vegetation indices were downloaded from Google Earth Engine Sentinel-2 satellite imagery.

In essence, this study contributes to the growing body of knowledge surrounding remote sensing applications in agriculture, highlighting the potential of machine learning algorithms and vegetation indices in improving crop management strategies for enhanced agricultural sustainability and productivity. Overall, the findings of the study provide insights into the use of machine learning algorithms for estimating LAI in wheat crops, and underscore the importance of vegetation indices in remote sensing-based LAI estimation.

Introduction

LAI typically stands for Leaf Area Index. Leaf Area Index (LAI) is a critical parameter in agriculture and crop management. It refers to the total area of leaves per unit ground area within a canopy. LAI is an essential metric because it directly relates to the amount of photosynthetically active radiation (PAR) intercepted by the crop canopy, which in turn influences plant growth, development, and ultimately yield.

Wheat crop is a significant crop in Himachal Pradesh, India. The state has a diverse topography and soil types, which make it suitable for growing various crops, including wheat. The crop is grown in the winter season, mainly in the districts of Kangra, Mandi, Una, and Bilaspur. The production of wheat in the state has been increasing over the years, and the state government has been taking several initiatives to improve the productivity and quality of the crop.

Remote sensing has become a useful tool for estimating Leaf Area Index (LAI) in crops in recent years. This technique enables fast and precise estimation of LAI across vast areas, making it an efficient tool for monitoring crop growth and productivity. Several studies have used various vegetation indices obtained from remote sensing data to accurately estimate LAI in crops.

Remote sensing-based LAI estimation provides valuable information for crop monitoring, yield prediction, precision agriculture, and management decision-making. It enables farmers and agronomists to assess crop health, identify areas of stress or potential yield limitations, and optimize inputs such as irrigation, fertilization, and pest control.

With the help of machine learning algorithms, the estimation of the Leaf Area Index (LAI) in agriculture has undergone a revolution. These algorithms analyze patterns in remote sensing data and ground measurements and learn complex relationships between spectral information and LAI values from training data. By predicting LAI across large agricultural areas, these algorithms provide farmers and agronomists with valuable insights for optimized crop management practices.

❖ Research Objectives:

1. To Investigate the relationship between different vegetation indices and leaf area index for wheat-crop in Palampur, Himachal Pradesh.
2. To develop machine learning models to estimate LAI.

Literature Review

Discovering the most accurate way to estimate LAI in wheat crops is crucial for achieving optimal yields. Fortunately, several studies have already been conducted that employ machine learning algorithms to achieve this goal. By utilizing these advanced techniques, we can ensure that our crops are healthy, productive, and profitable.

The Leaf Area Index (LAI) is an important metric used to measure the amount of green leaf area present per unit of ground area. This measurement is commonly used in the field of plant science and is a crucial element in the assessment of plant canopies and their performance (Jonckkheere,2003).

Formulating narrowband NDVIs using red edge (RE) bands is important for exploring field-collected winter wheat data over different growth stages and cultivars for quantity per unit surface area-based variables such as LAI.

In the empirical inquiry undertaken by Wang (2018), a sophisticated array of multivariate calibration techniques was employed to ascertain the Leaf Area Index (LAI) within the context of wheat crops. These techniques encompassed classical multilinear regression (MLR), partial least squares regression (PLS), and contemporary machine learning (ML) methodologies, inclusive of support vector regression (SVR), random forests (RF), and artificial neural networks (ANN). Notably, the study revealed that the RF model evinced the most promising efficacy in the estimation of LAI within paddy rice cultivation settings. Central to this investigation was the utilization of the Normalized Difference Vegetation Index (NDVI) as a pivotal vegetative metric.

The study conducted by Delegido in 2013 revealed that the Normalized Difference Vegetation Index (NDVI), derived from spectral reflectance at 674nm and 712nm, exhibited the strongest linear correlation with Leaf Area Index (LAI) across nine different crop types.

This finding underscores NDVI's significance as a reliable indicator for LAI estimation, warranting further investigation in agricultural research.

The research found that Support Vector Regression (SVR) surpassed Partial Least Squares (PLS) and Multilinear Regression (MLR) in predicting Leaf Area Index (LAI) for paddy rice using 15 selected bands from hyperspectral data spanning 350 nm to 2500 nm.

In Wang's (2017) study, the application of Support Vector Machine (SVM) for estimating Leaf Area Index (LAI) in wheat crops proved superior to other machine learning algorithms in terms of accuracy. This highlights SVM's potential as an effective tool for precise LAI estimation in wheat cultivation, meriting further investigation in agricultural research.

Moreover, in addition to the aforementioned studies, other ML algorithms such as Artificial Neural Networks (ANN), K-Nearest Neighbours (KNN), and Linear Regression (LR) have been utilized for LAI estimation in wheat crops.

However, the selection of the most suitable ML algorithm depends on factors such as data availability, problem complexity, and desired accuracy levels.

Overall, previous studies have shown that ML algorithms can be used as effective tools for LAI estimation in wheat-crop. However, the choice of ML algorithm may depend on various factors, and further research is required to identify the most appropriate algorithm for specific crops and growing conditions**.**

Cumulatively, prior research highlights the effectiveness of ML algorithms in LAI estimation for wheat-crop. However, selecting the most suitable ML algorithm hinges on various factors. Further investigation is warranted to ascertain the optimal algorithmic choices for specific crop types and environmental conditions.

Materials and Methodology

❖ Study Area

The study area for this research is located in Kangra District, Himachal Pradesh, India. The area is well-known for its agricultural activities, particularly for the cultivation of wheat-crop. Ground LAI data were collected from various locations within the study area to establish the relationship between the vegetation indices and LAI.

To conduct the research, the study area chosen was Palampur-Kangra in Google Earth Engine, using a Shapefile as an outermost boundary. This helped to ensure that the data collected for the study was relevant and specific to the area of interest.

The AOI(Area of Interest) was created using the FAO GAUL: Global Administrative Unit Layers 2015 dataset. The Global Administrative Unit Layers (GAUL) compiles and disseminates the best available information on administrative units for all the countries in the world, contributing to the standardization of the spatial dataset representing administrative units. The GAUL always maintains global layers with a unified coding system at the country, first (e.g. departments), and second administrative levels (e.g. districts). Where data is available, it provides layers on a country-by-country basis down to third, fourth, and lower levels.

After AOI creation we like to obtain the Remote Sensing images for the study area from Sentinel-2. The use of the AOI (Area of Interest) helped to ensure that the study focused solely on the crop of interest and a specific area of interest. This approach led to successful LAI (Leaf Area Index) estimation using ML models.

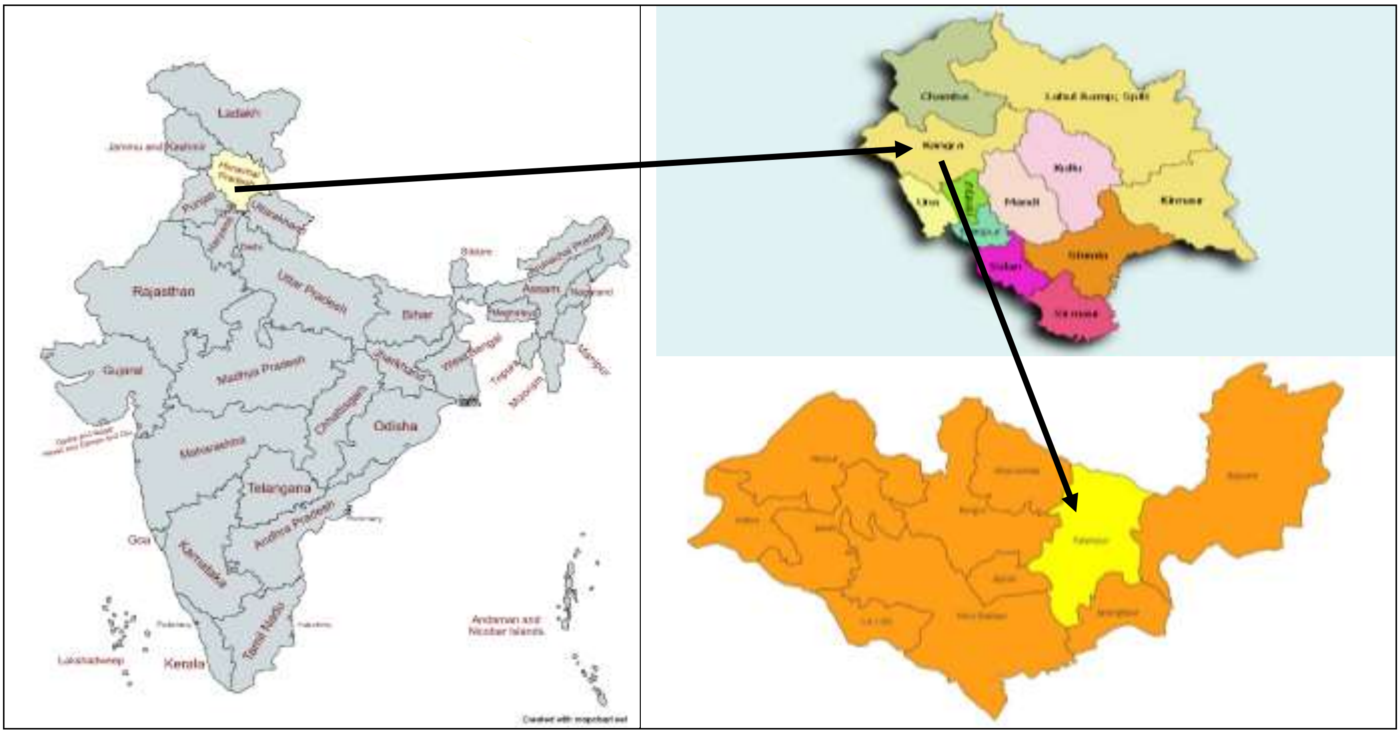

Figure-1: Study Area

❖ Flow chart

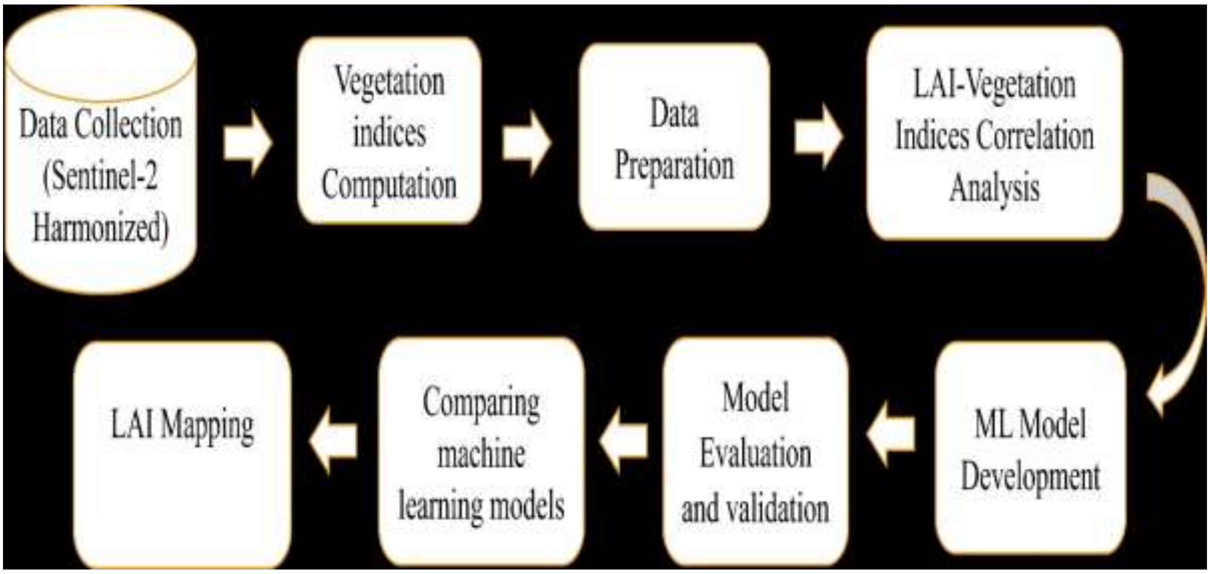

Figure 2: Project Flow-chart

❖ Data Collection

In this study, we have ground data for wheat in the Palampur subdistrict of Kangra District of Himachal Pradesh from 1st January 2023 to 31st January 2023.

We have the Ground Truth points of LAI with Geographic Locations, which helped us determine the LAI values for the same points with the help of various indices. We used this ground data to download several vegetation indices from the Sentinel-2 satellite imagery in the Google Earth Engine platform. The vegetation indices included NDVI, EVI, GNDRVI, WDRVI, CI, and NDRE, which are commonly used in LAI estimation studies.

The downloaded vegetation indices were then used to develop machine-learning models to predict LAI for wheat-crop. The models were validated using ground-based LAI measurements, and the results showed a high correlation between the predicted and actual LAI values.

Overall, this study highlights the potential of using machine learning algorithms and remote sensing data to predict LAI for wheat-crop. The use of ground-based measurements and satellite imagery from Sentinel-2 allowed for a more accurate estimation of LAI values, which can be used to improve crop management practices and increase crop productivity. The results of this study can be used as a baseline for future LAI prediction studies in the Palampur subdistrict of Kangra district of Himachal Pradesh and other similar agricultural regions.

❖ Vegetation Indices used to estimate LAI:

- **Normalized Difference Vegetation Index (NDVI):**

  - ✓ Normalized Difference Vegetation Index (NDVI) is a widely used vegetation index in remote sensing.
  - ✓ It is calculated using the ratio of the difference between the near-infrared (NIR) and red bands to the sum of the same two bands.

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

  - ✓ NDVI values range from -1 to 1, where values closer to 1 indicate high vegetation cover and values closer to -1 indicate low vegetation cover.
  - ✓ NDVI is used to estimate vegetation health, growth, and productivity, making it a useful tool in agriculture, forestry, and environmental studies.

- **Enhanced Vegetation Index (EVI):**

  - ✓ The Enhanced Vegetation Index (EVI) is a vegetation index obtained from remote sensing data, which provides information about the density, vigor, and overall health of vegetation.
  - ✓ EVI is a modified version of the Normalized Difference Vegetation Index (NDVI) that minimizes atmospheric influences and improves sensitivity in high biomass regions.

$$EVI = 2.5 * \frac{(NIR - RED)}{(NIR + C_1 * RED - C_2 * BLUE + L)}$$

  - ✓ It uses the blue band, red band, and near-infrared (NIR) band to calculate the vegetation index. EVI has been widely used in agriculture, forestry, and ecology to monitor vegetation dynamics, assess vegetation productivity, and predict crop yield.
  - ✓ EVI ranges from -1 to 1, with higher values indicating higher vegetation cover.

- **Green Normalized Difference Vegetation Index (GNDVI):**

  - ✓ GNDVI is a vegetation index that is commonly used to estimate vegetation density and health. It is calculated using the reflectance values obtained from remote sensing data and is a ratio of the difference between near-infrared (NIR) and green reflectance to the sum of NIR and green reflectance.

$$\mathrm{GNDVI} = \frac{(\mathrm{NIR} - \mathrm{green})}{(\mathrm{NIR} + \mathrm{green})}$$

- ✓ GNDVI is particularly useful in remote sensing because it is less sensitive to atmospheric interference and soil background effects than some other vegetation indices. It is also useful for distinguishing between green vegetation and non-vegetation areas.
- ✓ GNDVI values range from -1 to 1, with positive values indicating healthy vegetation and higher values indicating denser and more healthy vegetation.
- ✓ GNDVI can be used for a variety of applications, including crop monitoring, land-use mapping, and ecological studies.

- **Chlorophyll Index (CI):**

  - ✓ Chlorophyll index (CI) is a vegetation index that has been widely used in remote sensing for the estimation of chlorophyll content in plants.
  - ✓ The chlorophyll index is calculated based on the reflectance of light at specific wavelengths, which are sensitive to the absorption of chlorophyll in leaves.

$$CI_{GREEN} = \frac{NIR}{GREEN} - 1$$

  - ✓ CI is an effective indicator of plant health and productivity, and it has been used in various studies to estimate crop yield and monitor plant stress.
  - ✓ The use of CI in remote sensing is particularly useful for large-scale monitoring of crop health and management, as it enables rapid and accurate estimation of chlorophyll content in crops over large areas.

- **Normalized Difference Red Edge (NDRE):**

  - ✓ Normalized Difference Red Edge (NDRE) is a vegetation index used in remote sensing to estimate crop biomass and chlorophyll content.
  - ✓ NDRE is calculated using the red edge band and the near-infrared band of the electromagnetic spectrum.

$$NDRE = \frac{NIR - Red\ Edge}{NIR + Red\ Edge}$$

  - ✓ NDRE is an improvement over NDVI because it is less sensitive to soil background and atmospheric influences. NDRE is particularly useful for crops with dense canopies, such as wheat and corn.
  - ✓ NDRE values range from -1 to 1, with higher values indicating higher biomass and chlorophyll content.

- **Optimized Soil-Adjusted Vegetation Index (OSAVI):**

  - ✓ OSAVI is a vegetation index used in remote sensing to estimate the vegetation cover and biomass of crops. It is an improvement over the traditional Soil-Adjusted Vegetation Index (SAVI), which was developed to minimize the effects of soil brightness on the vegetation index values.
  - ✓ The OSAVI formula further adjusts the SAVI by introducing an optimized soil adjustment factor.

$$\text{OSAVI} = (1 + 0.16)\frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R} + 0.16}$$

Where,

The optimized soil adjustment factor is 0.16, which is obtained using a numerical optimization technique.

- ✓ The OSAVI values range from -1 to 1, with higher values indicating higher vegetation cover and biomass.
- ✓ It is particularly useful in arid and semi-arid regions where the soil brightness can significantly affect the vegetation index values. Overall, the OSAVI is a valuable tool for monitoring vegetation cover and biomass in agricultural and ecological applications using remote sensing data.

In our study, we evaluated the performance of these seven vegetation indices in estimating LAI in wheat and rice crops using various ML algorithms.

❖ Correlation Analysis:

Correlation analysis is a statistical technique used to determine the strength and direction of the relationship between two variables. It is commonly used in research to identify the degree of association between variables and to test hypotheses about the nature of this association. In agriculture, correlation analysis can help identify relationships between crop parameters such as yield, growth, and nutrient levels, as well as environmental factors such as temperature, rainfall, and soil properties.

The correlation coefficient, usually denoted by r, is a numerical measure of the strength and direction of the relationship between two variables. It ranges between -1 and 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation. The closer the coefficient is to -1 or 1, the stronger the correlation, while coefficients close to 0 indicate weak or no correlation.

In summary, correlation analysis is a useful statistical technique for exploring the relationships between variables in agriculture research. It can help identify significant associations between crop parameters and environmental factors, providing valuable insights for crop management and decision-making.

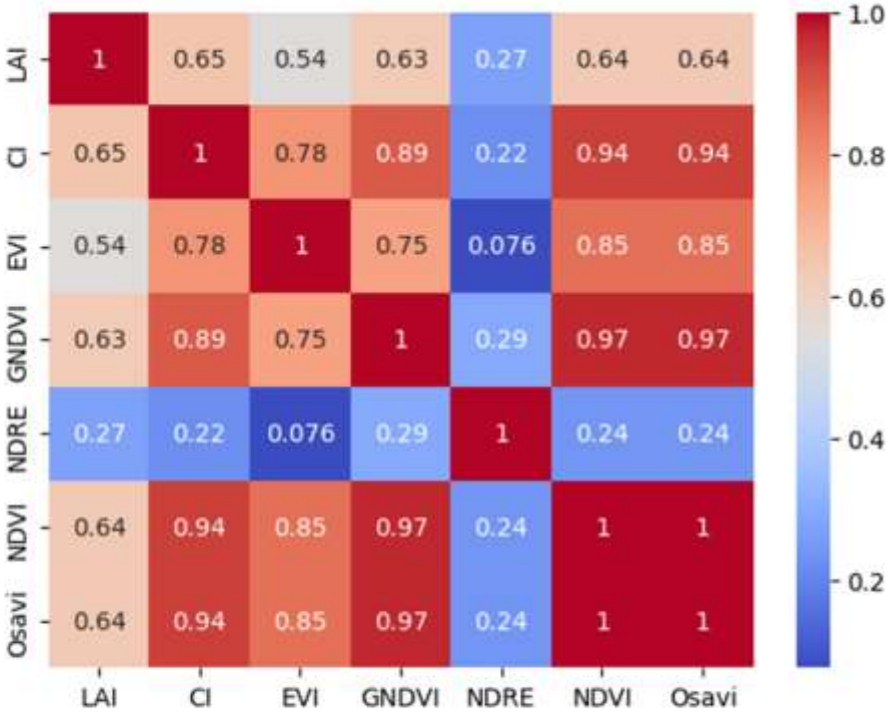

Figure 3: Vegetation Indices Correlation Analysis

This chart is a testament to the strength of the correlation between LAI and major vegetation indices derived from used vegetation indices. The data presented here clearly shows that these indices are highly indicative of the LAI.

❖ Pre-Processing:

1. **Data Cleaning:**

   Data cleaning is a critical step in the machine learning process. It ensures that the data is accurate, consistent, and reliable, which improves the performance and accuracy of the model.

   It involves identifying and correcting errors and inconsistencies in the dataset to improve the accuracy and reliability of the model. The process includes several steps such as identifying missing values, removing duplicates, handling outliers, correcting errors, and handling inconsistent data types.

2. **Data Normalization:**

   Data normalization is an essential pre-processing step in machine learning that involves transforming numerical data into a common scale to improve the performance and accuracy of machine learning models. Normalizing data helps in avoiding bias towards any particular feature and improves the convergence rate of optimization algorithms.

   It helps in improving the performance and accuracy of machine learning models. It involves scaling the data to a common range of values using various normalization techniques such as Min-Max Scaling, Z-Score Normalization, and Decimal Scaling.

3. **Feature Selection:**

   Feature selection is a critical step in machine learning that involves identifying and selecting a subset of relevant features from a larger set of available features. The goal of feature selection is to improve the performance of a machine learning model by reducing the dimensionality of the input data and removing irrelevant or redundant features.

   The benefits of feature selection include improved model performance, reduced training time, and improved interpretability and understanding of the underlying relationships between the input features and the target variable. However, it is important to note that feature selection should be done carefully and with consideration of the domain knowledge and context of the problem, as removing important features can lead to poor performance and inaccurate predictions.

4. **Data Splitting:**

   Data splitting is a crucial step in machine learning, which involves dividing a dataset into two or more subsets. The most common way to split data is into training and testing sets. The training set is used to train the machine learning model, while the testing set is used to evaluate the model's performance.

   It helps to ensure the model's performance on new, unseen data. It involves dividing the dataset into training, testing, and validation sets and can be done using different methods such as random splitting or stratified sampling.

   By following these pre-processing steps, we can prepare the data for ML-based LAI estimation of wheat crop.

❖ LAI-Vegetation Index Correlation Charts

To examine the relationship between Leaf Area Index (LAI) and Vegetation Indices extracted from Sentinel-2 imagery in wheat crops, we constructed a scatter plot.The LAI values were plotted on the x-axis, while the corresponding Vegetation Index values were plotted on the y-axis.

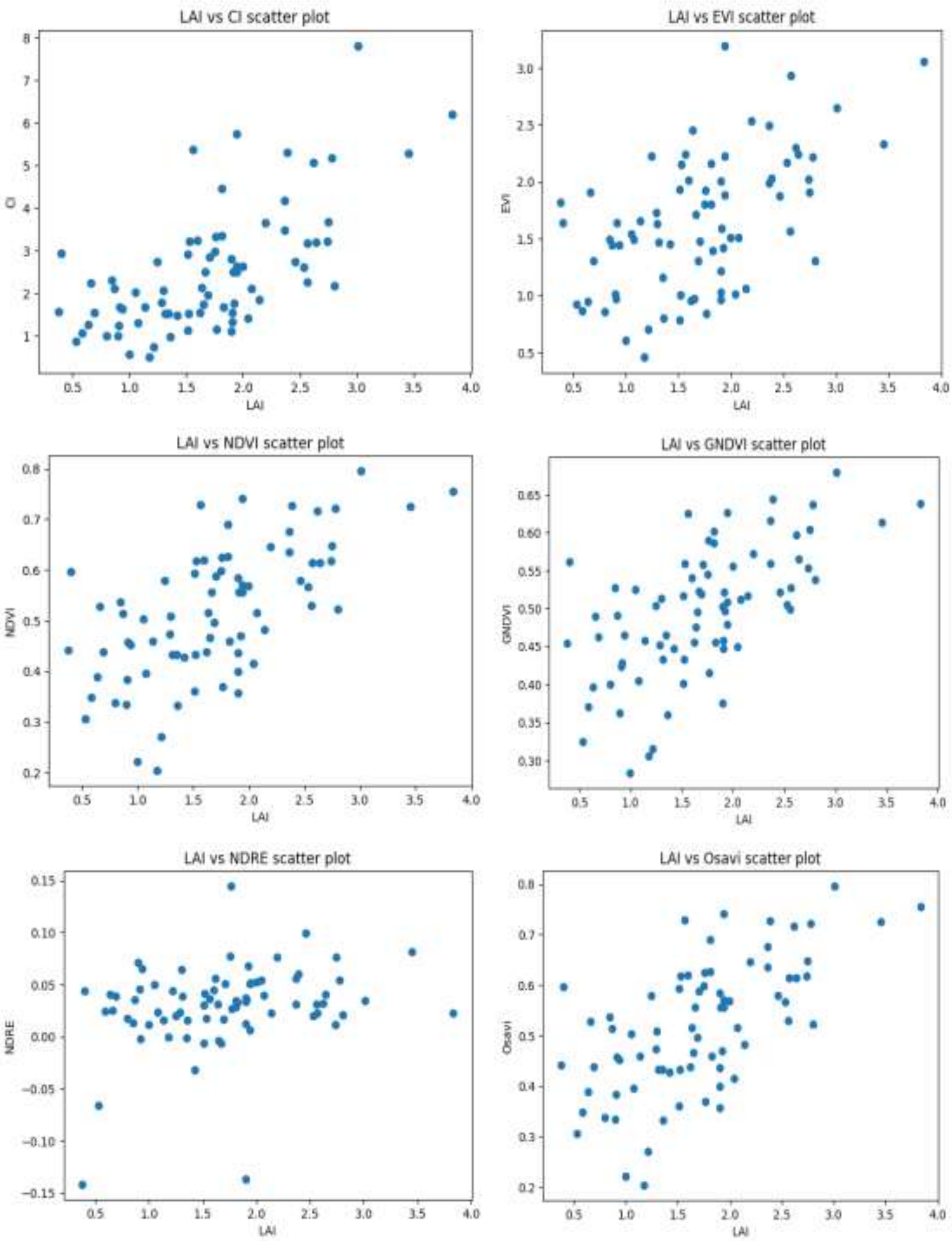

Figure 4: ***LAI-Vegetation Index Correlation Charts***

The graph revealed a discernible positive linear trend between the two variables, indicating that higher ground LAI values corresponded to elevated Vegetation Index values. This observation suggests that Vegetation Index values serve as reliable indicators of actual ground LAI values in wheat cultivation.

The scatter graphs served as essential tools for unraveling the intricate relationship between Leaf Area Index (LAI) and Vegetation Indices values within the context of wheat crops. By visually representing the data, these graphs elucidated the nuanced patterns and trends underlying the correlation between LAI and Vegetation Indices. Such insights are pivotal for discerning which Vegetation Index offers the most robust indicator of LAI levels in wheat fields.

In this study, we estimated LAI in wheat crop using a variety of ML algorithms. Among these algorithms are:

1. **<u>Multiple Linear Regressor</u>**:

Multiple linear regression is a statistical method used to identify the relationship between two or more independent variables and a dependent variable. It is commonly used in data analysis and can help predict future outcomes based on the relationships between the variables. The multiple linear regression model assumes that the relationship between the independent variables and the dependent variable is linear and additive. The goal of the model is to find the best fit line that represents the relationship between the variables.

2. **<u>Random Forest Regression</u>**:

Random forest regression is a machine learning algorithm used for regression problems. It works by creating multiple decision trees and combining their predictions to obtain a final prediction. Each decision tree is trained on a random subset of the data and a random subset of the features. This helps to reduce overfitting and improve the accuracy of the model. Random forest regression can be used for both continuous and categorical variables and is a popular choice for many real-world applications due to its high accuracy and flexibility.

3. **<u>Support Vector Machines Regression</u>:**

Support Vector Machines Regression (SVMR) is a machine learning algorithm that is commonly used for regression analysis. It is a type of supervised learning algorithm that can be used to predict continuous numeric values, such as the amount of rainfall in a particular region. SVMR works by finding the optimal hyperplane that separates the data into two classes, with the largest margin between the classes. In the case of regression, SVMR finds the optimal hyperplane that best fits the data points and predicts the output value based on the distance from the hyperplane. SVMR is known for its ability to handle large datasets and its ability to handle non-linear relationships between the input and output variables.

4. **Gradient Boosting Regression**:

Gradient Boosting Regression is a popular machine learning algorithm that is used for regression problems. It's a type of ensemble method where multiple weak models are combined to form a strong model. The algorithm works by fitting a model to the data and then adjusting the model to minimize the error. It then fits the next model to the residual errors of the first model and continues to do so for a specified number of iterations. This process continues until the desired level of accuracy is achieved. Gradient Boosting Regression is known for its ability to handle large datasets, noisy data, and outliers.

5. **XGBoost**:

XGBoost is a leading machine learning algorithm known for its exceptional performance in predictive modeling. It belongs to the ensemble learning family, specifically the gradient boosting framework, and is highly scalable, efficient, and flexible. With features like parallel computing and built-in regularization, XGBoost excels in diverse datasets and predictive tasks, making it a popular choice for structured data analysis and machine learning competitions.

❖ Evaluation of the Model:

Once a machine learning model is trained, it needs to be evaluated using new and unseen data. This is done by testing the model on a separate dataset that was not used during the training phase. If the same data used for training is used for testing, the model may perform better than expected as it is already familiar with the data and can recognize patterns it has already learned. Therefore, to accurately assess the model's performance and efficiency, it is essential to evaluate it on testing data. Three metrics used for this evaluation are Mean Absolute Error, R2 Score, and Cross Validation Score.

Mean Absolute Error(MAE):

It is the average distance between Predicted and original values. It gives how we have predicted from the actual output. However, there is one limitation i.e. it doesn’t give any idea about the direction of the error which is whether we are under-predicting or over-predicting our data. It can be represented mathematically in this way:

$$\mathrm{MAE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{j}=1}^{\mathrm{N}} |\mathrm{y_j} - \hat{\mathrm{y}}_\mathrm{j}|$$

R2 Score:

The coefficient of determination also called the R2 score is used to evaluate the performance of a linear regression model. It is the amount of variation in the output-dependent attribute that is predictable from the input independent variable(s). It is used to check how well-observed results are reproduced by the model, depending on the ratio of total deviation of results described by the model.

R2 score, or the coefficient of determination, is a statistical metric used to assess the performance of a machine learning model. It measures the amount of variation in the model's predictions and how well the independent variable predicts the dependent variable.

The R2 score is calculated by dividing the sum of the squared differences between the predicted values and the actual values by the sum of the squared differences between the actual values and their mean. The equation for the R2 score is:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

While a perfect score of 100% indicates a perfect fit, there is no set standard for an ideal R2 score, and models with low scores can still be useful.

RMSE (Root Mean Squared Error):

We can say that RMSE is a metric that can be obtained by just taking the square root of the MSE value. As we know the MSE metrics are not robust to outliers and so are the RMSE values. This gives higher weightage to the large errors in predictions.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{N}\left(y_j - \hat{y}_j\right)^2}{N}}$$

RMSE is useful because it penalizes large errors more than small errors. A smaller RMSE indicates a better fit between the predicted and actual values.

Index of Agreement (IA)

Index of agreement (IA), also known as Willmott's index or Willmott's similarity index, is a statistical measure used to evaluate the similarity between two sets of data. It is commonly used in the field of hydrology to assess the accuracy of model predictions compared to observed data.

The formula for the index of agreement is:

IA = 1 - (Σ|i=1,n|(|Oi - Pi|)/(Σ|i=1,n|(|Oi - O_mean|) + Σ|i=1,n|(|Pi - O_mean|))))

where:

- Oi is the observed value for the i^{th} data point
- Pi is the predicted value for the i^{th} data point
- O_mean is the mean of the observed values
- n is the number of data points

The IA value ranges from 0 to 1, with a value of 1 indicating perfect agreement between the observed and predicted data, and a value of 0 indicating no agreement.

To calculate the IA, you need to have a set of observed values and a set of predicted values for the same set of data points. Then, you can use the formula to calculate the IA value.

Result and Conclusion

❖ Result

The table presents the performance evaluation results of various machine learning algorithms applied to regression tasks. Five models were assessed, including Linear Regression, Random Forest Regressor, Support Vector Machine (SVM) Regressor, Gradient Boosting Regressor (GBR), and XGBoost Regressor. Three key evaluation metrics were considered: Root Mean Squared Error (RMSE), R-squared Score (R^2 Score), and Index of Agreement.

Sr. No.	Model	RMSE	R2_Score	Index of Agreement
1	**Multiple Linear regression**	0.3934	0.6874	0.8782
2	**Random forest regressor**	0.3934	0.6047	0.8500
3	**SVM regressor**	0.4629	0.5672	0.7963
4	**GBR**	0.5431	0.4041	0.8118
5	**XG boost regressor**	0.4417	0.6058	0.8524

- Among these models, Linear Regression exhibited the lowest RMSE of 0.3934, indicating its relatively accurate predictions. Additionally, it achieved a high R^2 Score of 0.6874, signifying a strong correlation between predicted and actual values.

- The Index of Agreement for Linear Regression was also notably high at 0.8782, demonstrating a high level of agreement between predicted and observed values.

- Random Forest Regressor and XGBoost Regressor also performed competitively across these metrics, while SVM Regressor and Gradient Boosting Regressor showed slightly lower performance.

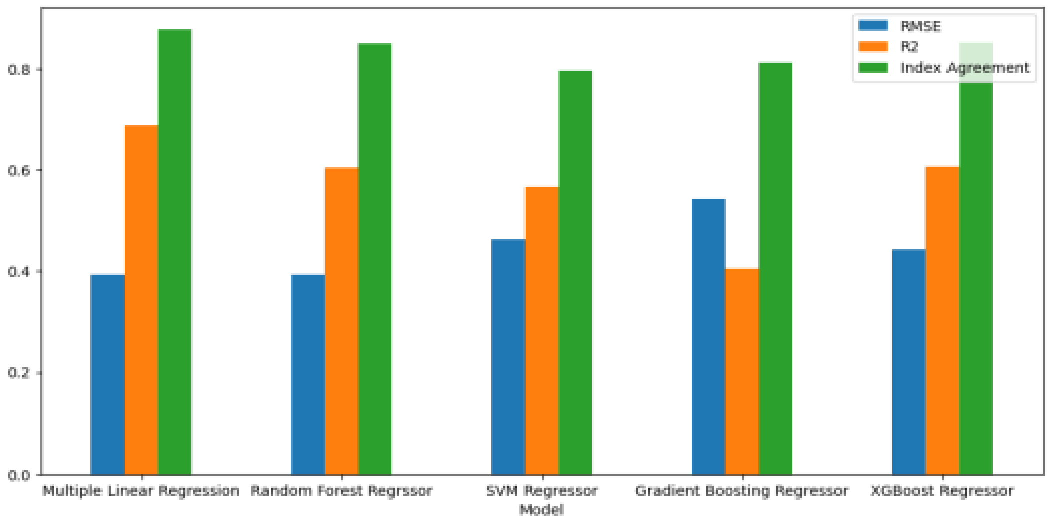

Figure 5: Performance Comparison of ML Algorithms

❖ **Palampur subdistrict's LAI map for wheat.**

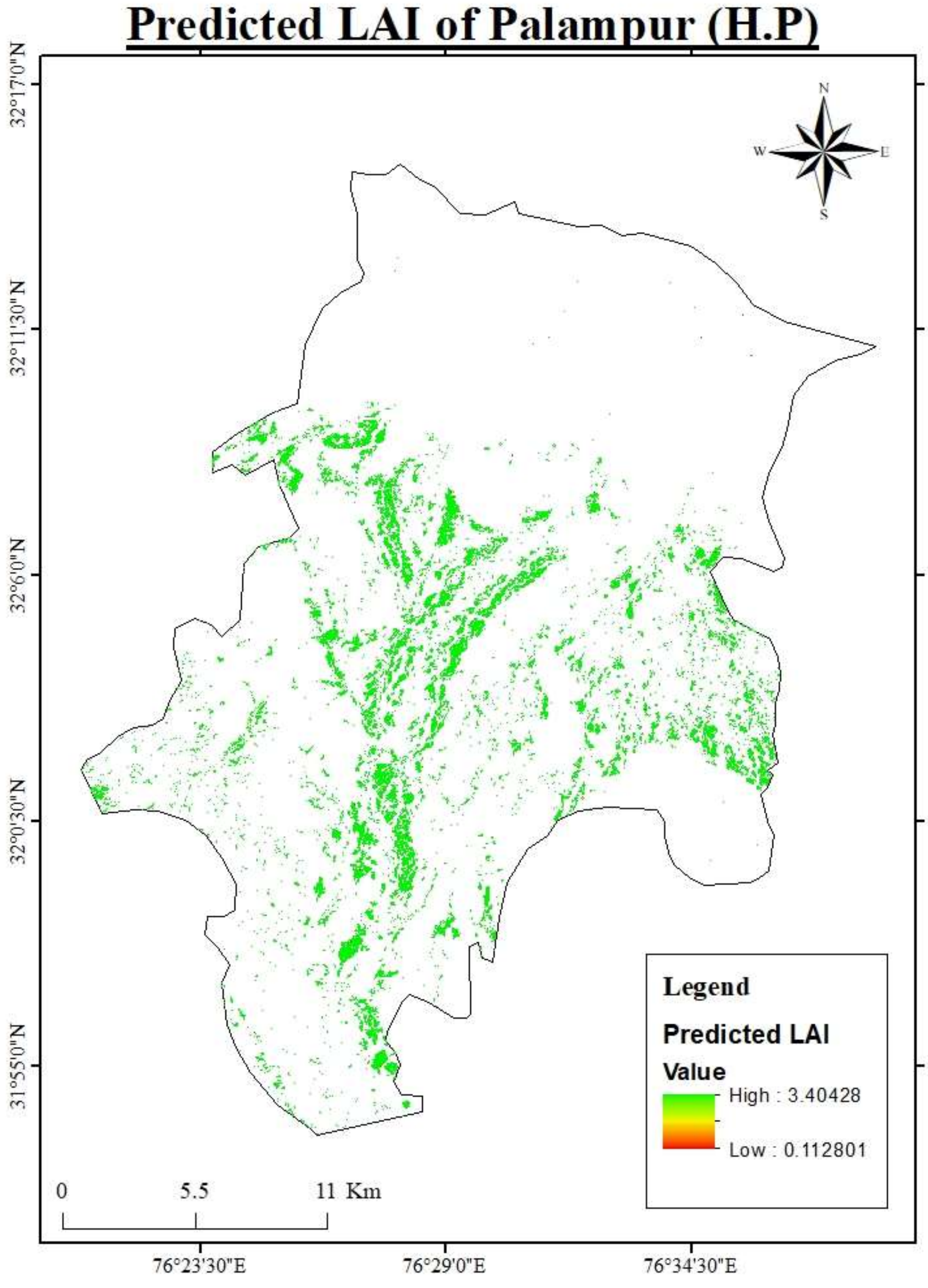

Figure 6: LAI Map for Wheat in Palampur Subdistrict

Conclusion

The study's findings highlight the effectiveness of multiple linear regression (MLR) in conjunction with key vegetation indices - namely NDVI, CI, Osavi, EVI, and GNDVI, in accurately estimating LAI.

The high Index of Agreement (0.8782) for MLR underscores its robustness in agreement between predicted and observed LAI values, indicating its potential for practical implementation in crop management decision-making processes.

Furthermore, the analysis reveals significant relationships between LAI and various vegetation indices, reaffirming their importance in remote sensing-based LAI estimation. Notably, except for Red Edge NDVI, all investigated indices demonstrate relevance, emphasizing their collective utility in monitoring wheat crop dynamics.

In essence, this study contributes to the growing body of knowledge surrounding remote sensing applications in agriculture, highlighting the potential of machine learning algorithms and vegetation indices in improving crop management strategies for enhanced agricultural sustainability and productivity.

These findings provide valuable insights into the effectiveness of different machine learning algorithms for regression tasks, aiding in informed model selection for future predictive modeling endeavors.

References

1. *Srinet, Ritika & Nandy, Subrata & N R Patel. (2019). Estimating leaf area index and light extinction coefficient using Random Forest regression algorithm in a tropical moist deciduous forest, India. Ecological Informatics. 52. 10.1016/j.ecoinf.2019.05.008.*

2. *Zhang, Juanjuan & Cheng, Tao & Guo, Wei & Xu, Xin & Qiao, Hongbo & Xie, Yimin & Ma, Xinming. (2021). Leaf area index estimation model for UAV image hyperspectral data based on wavelength variable selection and machine learning methods. Plant Methods. 17. 10.1186/s13007-021-00750-5.*

3. *Durbha, Surya & King, Roger & Younan, N.H.. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. Remote Sensing of Environment. 107. 348-361. 10.1016/j.rse.2006.09.031.*

4. *Dong, Taifeng & Liu, Jiangui & Shang, Jiali & Qian, Budong & Ma, Baoluo & Kovacs, John & Walters, Dan & Jiao, Xianfeng & Geng, Xiaoyuan & Shi, Yichao. (2018). Assessment of red-edge vegetation indices for crop leaf area index estimation. Remote Sensing of Environment. 222. 133-143. 10.1016/j.rse.2018.12.032.*

5. *Asner, Gregory & Scurlock, Jonathan & Hicke, Jeffrey. (2003). Global synthesis of leaf area index observations: implications for ecological and remote sensing studies: Global leaf area index. Global Ecology and Biogeography. 12. 191-205. 10.1046/j.1466-822X.2003.00026.x.*

6. *Casa, Raffaele & Upreti, Deepak & Pelosi, Federico. (2019). Measurement and estimation of leaf area index (LAI) using commercial instruments and smartphone-based systems. IOP Conference Series: Earth and Environmental Science. 275. 012006. 10.1088/1755-1315/275/1/012006.*

7. *Clevers, J.G.P.W. & Kooistra, Lammert & Brande, Marnix. (2017). Using Sentinel-2 Data for Retrieving LAI and Leaf and Canopy Chlorophyll Content of a Potato Crop. Remote Sensing. 9. 405. 10.3390/rs9050405.*

8. *Jonckheere, Inge & Fleck, Stefan & Nackaerts, Kris & Muys, Bart & Coppin, Pol & Weiss, Marie & Frederic, Baret. (2003). Methods for leaf area index determination. Part I: Theories, techniques and instruments. Agricultural and Forest Meteorology. 121.*

9. *Pokovai, Klara & Fodor, Nándor. (2019). Adjusting Ceptometer Data to Improve Leaf Area Index Measurements. Agronomy. 9. 866. 10.3390/agronomy9120866.*

10. *Verrelst, Jochem & Rivera Caicedo, Juan & Veroustraete, Frank & Muñoz, Jordi & Clevers, J.G.P.W. & Camps-Valls, Gustau & Moreno, Jose. (2015). Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison. ISPRS Journal of Photogrammetry and Remote Sensing. -. 10.1016/j.isprsjprs.2015.04.013.*

11. *Turner, David & Cohen, Warren & Kennedy, Robert & Fassnacht, Karin & Briggs, John. (1999). Relationships between Leaf Area Index and Landsat TM Spectral Vegetation Indices across Three Temperate Zone Sites. Remote Sensing of Environment. 70. 52-68. 10.1016/S0034-4257(99)00057-7.*

12. *Boegh, E. & Soegaard, H. & Broge, Niels & Hasager, C. & Jensen, N.O. & Schelde, Kirsten & Thomsen, A.. (2002). Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. Remote Sensing of Environment. 81. 179-193. 10.1016/S0034-4257(01)00342-X.*

13. *Kross, Angela & McNairn, Heather & Lapen, David & Sunohara, Mark & Champagne, Catherine. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. International Journal of Applied Earth Observation and Geoinformation. 34. 235–248. 10.1016/j.jag.2014.08.002.*

Thank You!!