ESTIMATION OF LAI WITH A COMPARATIVE ANALYSIS FOR YIELD PREDICTION IN RICE AND MAIZE

[1]K. Srinivasa Chakravarthy, [2]Dr. G.Shyama Chandra Prasad, [3]M. Shruthi, [4]D.V.S Anurag, [5]A. Prudhvi Raj Varma

0009-0007-1777-1900, 0000-0001-9223-3060, 0009-0000-2186-3209, 0009-0000-1581-8401, 0009-0001-6341-5022

Abstract

This study focuses on leveraging machine learning and remote sensing technologies for predicting crop type, Leaf Area Index (LAI), and crop yield estimation.Using satellite imagery and extracted vegetation indices such as NDVI, EVI, and GNDVI, various regression models were evaluated for LAI prediction. A pre-trained VGG deep learning model is employed to classify the crop type from dataset images, specifically targeting rice and maize.Among the models tested, Lasso Regression provided the best performance with the lowest RMSE. Vegetation indices derived from aerial images are utilized to estimate the LAI of these crops. Subsequently, yield prediction is achieved by correlating the estimated LAI with production data from an auxiliary dataset containing state-wise and season-wise crop information. This approach integrates advanced deep learning with practical agricultural data to provide a scalable solution for yield estimation, reducing reliance on traditional, resource-intensive methods. The platform enables users to upload satellite images for automated predictions, presenting results such as predicted crop type, LAI, estimated yield, and yield trends over time through a user-friendly interface. This research demonstrates the potential of combining vegetation indices with advanced machine learning techniques to support precision agriculture, enhance crop monitoring, and optimize decision-making in the agricultural sector. The results aim to enhance precision agriculture practices and contribute to efficient crop management and planning.

Keywords— *Leaf Area Index (LAI), Data-Driven Agriculture, Rice and Maize Yield Estimation, Lasso Regression.*

I. INTRODUCTION

A. Problem Statement

Accurately predicting crop yield is a critical challenge for sustainable agriculture. Traditional methods for estimating yield require extensive field measurements and are often time-consuming and resource-intensive. This research addresses this challenge by developing a deep learning and image-based approach to estimate the Leaf Area Index (LAI) and predict crop yield. Using RGB aerial imagery and vegetation indices, the study focuses on two key crops—rice and maize—to bridge the gap between LAI and yield prediction through a comparative analysis of LAI with production data.

B. Scope and Objectives

This research develops a system for crop yield prediction using RGB aerial imagery. It includes crop identification (rice and maize) using a deep learning model, LAI estimation through vegetation indices, and yield prediction by mapping LAI to production data. The framework is designed to be scalable and applicable to diverse agricultural settings.

- Estimate Leaf Area Index (LAI) from RGB imagery using vegetation indices.
- Correlate LAI with yield using production data from a secondary dataset.
- Predict crop yield for target images based on calculated LAI.
- Provide a scalable, cost-effective tool for precision agriculture.

II. LITERATURE SURVEY

In recent years, advancements in remote sensing technologies, machine learning algorithms, and agronomic models have significantly enhanced our capacity to estimate crop yields and monitor crop health. This section surveys the key methodologies and findings from three influential research studies focused on yield estimation and crop growth monitoring.

The study by de Magalhães et al. (2024) [1] explores the integration of remote sensing data with machine learning techniques for precise yield estimation. Key insights include the utilization of high-resolution satellite imagery to extract vegetation indices such as NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index). Machine learning algorithms, including Random Forest and Support Vector Machines, were applied to predict yield across multiple crop types. Results demonstrated improved accuracy when combining multispectral data with machine learning models compared to traditional statistical approaches. The study emphasized the importance of spatial and temporal resolution in predicting regional and field-level yield variability.

The research by Yu et al. (2024) [2] investigates the application of multi-temporal remote sensing for tracking crop phenology and estimating Leaf Area Index (LAI). Key insights include the use of Sentinel-2 satellite data for multi-date observations to capture phenological stages. A hybrid model combining remote sensing indices and crop growth simulation was developed for estimating LAI. The research highlighted the correlation between LAI dynamics and yield potential, suggesting that monitoring LAI trends can serve as a proxy for early yield prediction. Key findings indicate that temporal resolution significantly influences LAI estimation accuracy, advocating for frequent data acquisition during critical growth phases.

The study by Castro Valdecantos et al. (2022) [3] delves into precision agriculture methods leveraging remote sensing and ground-based observations. Key insights include the development of yield maps using spatial interpolation techniques and field sensor data. The integration of proximal sensing devices for in-situ crop monitoring, complementing satellite observations, was also highlighted. The study presents a comparative analysis of kriging and inverse distance weighting for spatial yield prediction. Findings suggest that incorporating both satellite imagery and ground-truth data enhances predictive power, with implications for precision nutrient management and targeted irrigation.

These studies collectively illustrate the transformative role of remote sensing and machine learning in modern agriculture. The combination of machine learning models with high-resolution spectral indices yields more accurate predictions than traditional models. Temporal and spatial resolutions are critical factors in both LAI estimation and yield prediction. Integration of ground-based data with remote sensing enhances spatial accuracy, informing precision agricultural practices.

In conclusion, these advancements pave the way for more efficient and sustainable agricultural practices by enabling proactive decision-making through data-driven insights.

III. METHODOLOGY

This study aims to predict the crop type and estimate Leaf Area Index (LAI) values using vegetative indices derived from the rgb images. Additionally, it integrates machine learning models to forecast yield trends specific to rice cultivation in Kurnool district, Andhra Pradesh. The proposed methodology leverages remote sensing data, vegetative index calculations, and regression-based machine learning models for yield prediction.

A. System Architecture or Workflow

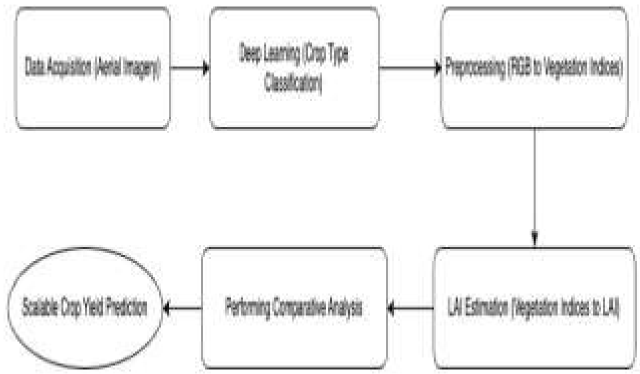

Fig. 1 Architecture for the Yield Prediction System

1. Data Acquisition (Aerial Imagery): Collect aerial images using UAVs (drones) equipped with high-resolution RGB cameras.

2. Deep Learning (Crop Type Classification): Use deep learning models (like CNNs) to classify the crop type in the acquired images.

3. Preprocessing (RGB to Vegetation Indices): Convert RGB images to vegetation indices (like NDVI) for assessing plant health and biomass.

4. LAI Estimation (Vegetation Indices to LAI): Estimate Leaf Area Index (LAI) from vegetation indices to quantify plant canopy density.

5. Performing Comparative Analysis: Using the LAI data, predict the potential crop yield.

6. Scalable Crop Yield Prediction: Implement scalable models for large-scale application across various agricultural settings.

   (Fig. 1 depicts the block diagram of the prediction pipeline, including preprocessing, feature extraction, and model deployment.)

B. Data Collection and Preparation

The datasets utilized for this research include the following:

1. **Image Dataset for Crop Classification**

High-resolution crop images were used to facilitate crop classification, focusing on distinguishing crops such as rice and maize. This dataset served as a critical input for model training and evaluation. [15]

2. **Crop Production Dataset**
   A structured CSV dataset containing detailed crop production statistics, including parameters like cultivated area and total production across various Indian states and districts. This dataset was instrumental in calculating the average yield for crops such as rice and maize. [16]

3. **Vegetation Indices Dataset**
   A dataset providing vegetation indices, including NDVI, EVI, GNDVI, RENDVI, and WDRVI, was employed to estimate the Leaf Area Index (LAI). These indices also contributed to the prediction of crop yield in the study. [17]

C. Feature Extraction and Selection

Vegetation indices used include NDVI, EVI, GNDVI, RENDVI, WDRVI, CI, and NDRE. Feature selection was automated using Lasso Regression, minimizing the RMSE by reducing multicollinearity among variables.

D. Algorithm

Lasso Regression Model: Used for its regularization strength to prevent overfitting.

Equation:

$$y = \beta 0 + \sum_{i=1}^{n} \beta i * xi, \text{where} \sum |\beta i| \leq t$$

The constraint $\sum |\beta i|$ introduces a penalty proportional to the absolute value of coefficients.

Random Forest and XGBoost were compared but yielded higher error margins.

E. Training and Testing:

- Data Split: 70% training and 30% testing.
- Validation: Cross-validation with k=5 to generalize model performance.
- Library: Scikit-learn's train_test_split and cross_val_score functions were used.

The approach combines spectral indices with machine learning to accurately predict yields and LAI, enabling precision agriculture tailored to local climatic and soil conditions. Lasso Regression proved most effective, balancing simplicity and accuracy.

IV. RESULTS AND OBSERVATIONS

The model has been developed to predict crop types, Leaf Area Index (LAI), and crop yield. The process involves using both machine learning models for crop classification and vegetation indices for LAI prediction, combined with a yield estimation based on LAI values.

Crop Prediction: The model successfully classifies crop types from input images, predicting one of the five class labels: Maize, Rice. This is achieved by preprocessing satellite images and extracting relevant vegetation indices such as NDVI, EVI, GNDVI, RENDVI, WDRVI, CI, and NDRE, which are then input into a pre-trained deep learning model for classification.

LAI Prediction: The LAI (Leaf Area Index) prediction model, trained using synthetic data generated from vegetation indices, demonstrates a strong performance. The model uses RandomForestRegressor, indicating that it can reliably predict LAI based on the input indices. The predicted LAI values range from 0 to 6, which aligns with the expected range for agricultural crops.

Yield Estimation: The model estimates crop yield based on the predicted LAI, leveraging a reference dataset that includes historical crop yield data from various regions. By applying a scaling factor derived from the ratio of predicted to reference LAI values, the model provides a yield estimate with a ±10% range. This approach allows for the estimation of crop yield even in areas with limited data.

Yield Trend Analysis: Yield trends for individual crops over time were analyzed and plotted. The results reveal insights into the variations in crop yields year over year. For example, the yield trend for Maize shows consistent growth, while Sugarcane yields exhibit more volatility. These trends provide valuable information for farmers and agricultural stakeholders to anticipate future yield performances and adjust strategies accordingly.

Average Yield Calculation: The average yield of a crop for a given year and location was calculated by analyzing crop production data based on the state, district, and crop type. The model estimates the average yield for a specific crop, and the accuracy is validated by comparing the model's estimated yield with historical production data.

These findings show the potential of the developed model for aiding precision agriculture by enabling more accurate predictions of crop yields and facilitating better planning for resource allocation.

Predicted Result for a Rice Crop:
Input: An RGB image of a rice crop (Fig. 2) was provided as input to the model.

Fig. 2 RGB image of a rice crop

Results:

- **Predicted Crop**: Rice
- **LAI Prediction**: 2.708865854080525
- **Average Yield**: 3.7871914661638044
- **Predicted Yield**: 4.016762633532386

The system also generated a graphical representation of yield trends over the years, as shown in fig. 3, which demonstrates the steady increase in average yield for rice.

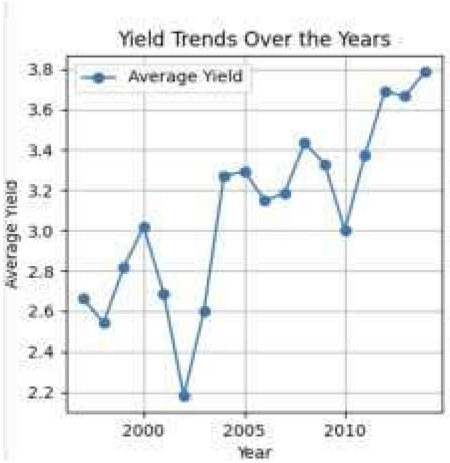

Fig.3:Average Yield vs. Year for the input crop shown in Fig.2

Predicted Result for a Maize Crop:
Input: An RGB image of a maize crop (Fig. 4) was provided as input to the model.

Fig. 4 RGB image of a maize crop

Results:

- **Predicted Crop:** Maize
- **LAI Prediction:** 2.9105587545176825
- **Average Yield:** 4.909019859521826
- **Predicted Yield:** 5.206593782006269

The system also generated a graphical representation of yield trends over the years, as shown in fig. 5, which demonstrates the steady increase in average yield for maize.

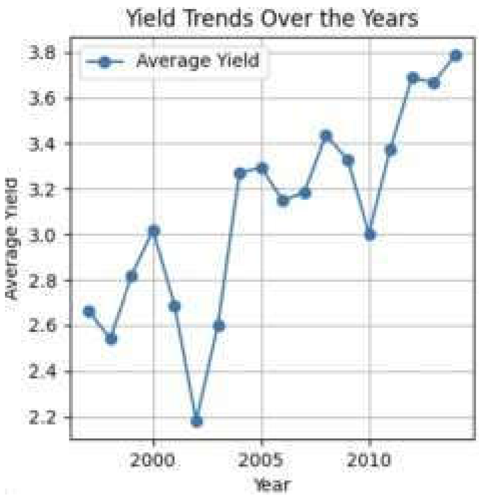

Fig.5: Average Yield vs. Year for the input crop shown in Fig.4

V. CONCLUSION AND FUTURE SCOPE

This research presents a robust methodology for predicting crop type, Leaf Area Index (LAI), and yield using vegetation indices and machine learning models. Among the evaluated models, Lasso Regression demonstrated the best performance, achieving the lowest RMSE, highlighting its suitability for the dataset. The proposed approach leverages vegetation indices derived from remote sensing data to deliver accurate predictions while minimizing overfitting. Though tested specifically on rice crops in Kurnool, Andhra Pradesh, the framework provides a scalable solution for precision agriculture, with potential adaptability to other regions and crop types through tailored datasets. Future work will focus on enhancing model generalizability and integrating additional data sources to further improve accuracy and applicability.

Future Scope:

Model Generalization: Expanding the dataset to include multiple crops and regions to enhance the model's adaptability and scalability.

Integration of Advanced Data Sources: Incorporating additional data like weather conditions, soil properties, and temporal satellite imagery to improve prediction accuracy.

Real-Time Predictions: Developing a cloud-based system for real-time crop monitoring and yield prediction accessible to farmers and stakeholders.

AI Optimization Techniques: Exploring advanced deep learning models or hybrid approaches for more precise predictions.

VI. REFERENCES

[1] de Magalhães, L. P., & Rossi, F. (2024). Use of Indices in RGB and Random Forest Regression to Measure the Leaf Area Index in Maize. Agronomy, 14(4), 750. https://doi.org/10.3390/agronomy14040750

[2] Yu, M., He, J., Li, W., Zheng, H., Wang, X., Yao, X., Cheng, T., Zhang, X., Zhu, Y., Cao, W., & Tian, Y. (2024). Estimation of Rice Leaf Area Index Utilizing a Kalman Filter Fusion Methodology Based on Multi-Spectral Data Obtained from Unmanned Aerial Vehicles (UAVs). Remote. Sens., 16, 2073.

[3] Castro Valdecantos, Pedro & Apolo, O. Enrique & Pérez-Ruiz, Manuel & Egea, Gregorio. (2022). Leaf area index estimations by deep learning models using RGB images and data fusion in maize. Precision Agriculture. 23. 1-18. 10.1007/s11119-022-09940-0.

[4] Du, L., Yang, H., Song, X. et al. Estimating leaf area index of maize using UAV-based digital imagery and machine learning methods. Sci Rep 12, 15937 (2022). https://doi.org/10.1038/s41598-022-20299-0

[5] Zhang, J., Cheng, T., Guo, W. et al. Leaf area index estimation model for UAV image hyperspectral data based on wavelength variable selection and machine learning methods. Plant Methods 17, 49 (2021). https://doi.org/10.1186/s13007-021-00750-5

[6] Yu, M., He, J., Li, W., Zheng, H., Wang, X., Yao, X., Cheng, T., Zhang, X., Zhu, Y., Cao, W., & Tian, Y. (2024). Estimation of Rice Leaf Area Index Utilizing a Kalman Filter Fusion Methodology Based on Multi-Spectral Data Obtained from Unmanned Aerial Vehicles (UAVs). Remote Sensing, 16(12), 2073. https://doi.org/10.3390/rs16122073

[7] Castro-Valdecantos, P., Apolo-Apolo, O.E., Pérez-Ruiz, M. et al. Leaf area index estimations by deep learning models using RGB images and data fusion in maize. Precision Agric 23, 1949–1966 (2022). https://doi.org/10.1007/s11119-022-09940-0

[8] Han, K.; Liu, B.; Liu, P.; Wang, Z. The Optimal Plant Density of Maize for Dairy Cow Forage Production. Agron. J. 2020, 112, 1849–1861.

[9] Panigrahi, N.; Das, B.S. Evaluation of Regression Algorithms for Estimating Leaf Area Index and Canopy Water Content from Water Stressed Rice Canopy Reflectance. Inf. Process. Agric. 2021, 8, 284–298.

[10] Rasmussen, J.J.; Ntakos, G.; Nielsen, J.F.; Svensgaard, J.; Poulsen, R.; Christensen, S. Are Vegetation Indices Derived from Consumer-Grade Cameras Mounted on UAVs Sufficiently Reliable for Assessing Experimental Plots? Eur. J. Agron. 2016, 74, 75–92.

[11] Luo, P.; Liao, J.; Shen, G. Combining Spectral and Texture Features for Estimating Leaf Area Index and Biomass of Maize Using Sentinel-1/2, and Landsat-8 Data. IEEE Access 2020, 8, 53614–53626.

[12] Chen, Y.; Ma, L.; Yu, D.; Feng, K.; Wang, X.; Song, J. Improving Leaf Area Index Retrieval Using Multi-Sensor Images and Stacking Learning in Subtropical Forests of China. Remote Sens. 2022, 14, 148.

[13] Han, K.; Liu, B.; Liu, P.; Wang, Z. The Optimal Plant Density of Maize for Dairy Cow Forage Production. Agron. J. 2020, 112, 1849–1861.

[14] Wang, Y.; Zhou, H.; Ma, X.; Liu, H. Combining Data Assimilation with Machine Learning to Predict the Regional Daily Leaf Area Index of Summer Maize (Zea mays L.). Agronomy 2023, 13, 2688.

[15] K. Srinivasa Chakravarthy, G. Shyama Chandra Prasad LAI calculation for Landsat, Hyperion, Sentinel using Time Series Analysis, IJISAE, 2024, 2808–2813

[16] K. Srinivasa Chakravarthy, G. Shyama Chandra Prasad, G. Monica, D. Lohitha Plant Species Identification Using a Custom CNN Model and Geometric Morphometrics – A Comparative Analysis, IJISAE, 2024, 2888–2895

[17] G. Shyama Chandra Prasad A Novel Method for Object Shape Description with Discrete Segments, IJCIHI, 2011, 0973-7413

[18] https://www.kaggle.com/datasets/aman2000jaiswal/agriculture-crop-images

[19]https://www.kaggle.com/datasets/iamtapendu/crop-production-data-india

[20] https://www.mdpi.com/2072-4292/16/6/954