

Crop Yield Estimation With Help of Unmanned Aerial Vehicle (UAV) & Advanced Machine Learning Techniques

Vedant Priyadarshi

vpriyadarshi6793@gmail.com

Centre for Advanced Studies (CAS), AKTU

Anuj Kumar Sharma

Centre for Advanced Studies (CAS), AKTU

Siddharth Yadav

Centre for Advanced Studies (CAS), AKTU

Yoggya Mehrotra

Centre for Advanced Studies (CAS), AKTU

Research Article

Keywords: UAV's, Precision Agriculture, Yield Prediction, Deep Learning, Convolutional Neural Networks, Random Forests, RGB Imagery, Vegetation Indices, Smart Farming

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9804916/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Crop Yield Estimation With Help of Unmanned Aerial Vehicle (UAV) & Advanced Machine Learning Techniques

Vedant Priyadarshi*¹, Anuj Kumar Sharma¹, Siddharth Yadav², Yoggya Mehrotra³

¹Department of Mechatronics, Centre for Advanced Studies (CAS), AKTU, Lucknow, Uttar Pradesh, 226031, India

²Department of Manufacturing Technology and Automation, Centre for Advanced Studies (CAS), AKTU, Lucknow, Uttar Pradesh, 226031, India

³Department of Electronics and Communication Engineering, Lucknow University, Lucknow, Uttar Pradesh, 226031, India

Email-id:

vpriyadarshi6793@gmail.com

Abstract

Determining Crop Yield Potential is increasingly becoming part of Intelligent Precision Agriculture Technology used to optimise management of farms, to enhance the use of resources & to achieve sustainable food production. Traditional Yield Assessment methods carried out through manual observation and statistical assessment (in most cases) require significant time, resources, and are often not effective because they do not account for the spatial variability present within agricultural fields. Therefore, this research proposes the development of an innovative intelligent crop yield prediction framework based on UAV technology, and advanced Remote Sensing & Artificial Intelligence Techniques, as an alternative to the traditional method of measurement. The current study reports on data collected from wheat fields using UAV collected high-resolution aerial photographs captured with RGB imaging cameras. The collected aerial imagery data were then processed using photogrammetric reconstruction techniques to create orthorectified maps of fields and 3D representations of wheat crops-from this data advanced image processing was performed for the purpose of Segmentation, Canopy Extraction and analysis of the Spatial Variability in Wheat Crops. UAV imagery was used to obtain various crop features (phenotypic and spectral), including: percentage of vegetation coverage, canopy texture, plant vigor, biomass characteristics, and growth variability. RGB-based vegetation indices (GLA, VARI, RGBVI, and ExG) were derived from UAV images for analysis of crop health and physiology. The extracted features were used to develop yield prediction models based on machine learning and deep learning algorithms. To predict yield, we used various predictive models (RF, SVM, DN, and LASSO). Performance of the models was compared using statistical measures such as root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). The results showed that the deep learning models outperformed the other prediction algorithms, due to their ability to learn complex spatial/non-linear relationships from UAV imagery. The suggested framework offers a quick, automated, and non-destructive method of estimating agricultural yield and for use within precision agriculture. The combination of aerial imagery from UAVs (drones), analysis through GIS, and data mining through Artificial Intelligence will permit a more efficient means for monitoring agriculture and making data-driven decisions at the farm level. The results of this research highlight the benefits that can be achieved from the development and application of UAV-enabled Smart Agriculture systems in terms of increasing productivity, decreasing resource wastage, and contributing to sustainable agricultural practices.

Keywords: UAV's, Precision Agriculture, Yield Prediction, Deep Learning, Convolutional Neural Networks, Random Forests, RGB Imagery, Vegetation Indices, Smart Farming.

1.0 Introduction

One of the oldest & most important sectors, agriculture has played an essential role in human civilization. Along with being responsible for economic development and food security, agriculture supports a large portion of the world population—directly or indirectly—for their food & livelihood. Agriculture is a significant part of the economy and is responsible for generating employment in developing nations like India. Agriculture plays a crucial role in rural development. Global food production systems face many challenges such as rapid population growth, climate change, diminishing arable land, water scarcity & increasing environmental stress that will threaten both the ability to grow food and the sustainability of

agricultural production systems. In order to meet the growing demand for food by the rapidly growing world population, there needs to be an increase in agricultural productivity through improving agricultural practices (more efficient) & implementing sustainable methods [1]. Accurate estimating of crop yields is one of the most important factors in agricultural management. Yield prediction supports food supply planning, policy development, irrigation scheduling, fertilizer application, transportation, storage & forecasting demands for markets. Emerging yield estimates enable farmers, researchers & policymakers to make educated decisions when determining how to manage crops & allocate resources. In addition, providing early crop yield estimates reduces economic risk & improves agricultural productivity. Obtaining accurate estimate of crop yields is a complex organization due to the interaction between the environmental, biological & management factors that will affect crop production [2]. Crop yield estimation has historically been accomplished using manual field surveys, visual inspections of the crops, empirical observations, and destructive sampling methods. The traditional approach to estimating yield has generally required considerable amounts of labor, high operational cost, and a significant amount of time to collect and analyze the data. Field measurements represent only a subset of the entire agricultural area and do not adequately represent the spatial variability found throughout the entire area of cropped land. There are many ways that human error, irregularity of the sampling procedure, and environmental variability can adversely affect estimation accuracy. Remote sensing technologies have become powerful tools to monitor agriculture and analyze crop yield within the last thirty years [3]. Remote sensing has made it possible to conduct large-scale, non-destructive observations of crop growth and vegetation, soil conditions, and environmental factors impacting plants and crops. Remote sensing systems have been successfully utilized for crop mapping, monitoring vegetation, and estimating crop yield using satellite-based imaging. However, many satellite images are restricted to use because of their low spatial resolution, cloud cover, atmospheric interference, and infrequent revisits; thus, remote sensing does not have the ability to provide high precision analysis of agriculture at the field level [4]. To solve the challenges faced in agriculture today, drones (also called Unmanned Aerial Vehicles (UAVs)) can offer flexible, low-cost, fast, and high-resolution aerial imaging capabilities for crop monitoring and agricultural assessment. Drones fly close to the ground and record precise images depending on how the user chooses to use them, allowing for immediate and site-specific evaluation of agriculture situations [5]. Today's UAVs can carry an assortment of sensing technologies like RGB cameras (digital) and sensors that capture spectral images. These sensing devices allow users to obtain a greater array of UAV-collected data regarding the health of plants, including information such as the vegetation's vigor, canopy structure, chlorophyll concentration, height, density, stress from water, dry matter, and nutrient levels [6]. Of all of the different types of sensing systems available, RGB cameras are the most commonly utilized in agriculture because they are affordable, lightweight, and very easy to work with on almost any

commercial drone platform. In comparison to multispectral sensors, RGB imagery does not carry as much spectral information; however, the use of advanced image processing methods in conjunction with machine-learning can still yield an abundance of information about crops, even when utilizing only visible-light imagery [7]. As such, RGB-based UAV systems have become a viable and cost-effective means for implementing precision agriculture, particularly in developing nations. Vegetation indices formulated from UAV images are frequently employed to assist in assessing the health of crops, estimating the levels of chlorophyll present in plants, determining whether or not vegetation is vigorous, and diagnosing stress within vegetative systems. Vegetation indices are mathematical representations that combine multiple spectral bands to increase the amount of vegetative information present and decrease any background noise. Several vegetation indices based on RGB data, such as the Green Leaf Algorithm (GLA), the Visible Atmospherically Resistant Index (VARI), the Red Green Blue Vegetation Index (RGBVI), the Excess Green Index (ExG), the Excess Red Index (ExR), and the Normalized Green-Red Difference Index (NGRDI) have been shown to be highly effective for performing agricultural analyses and monitoring crops [8]. In addition to using spectral analysis, UAV imagery provides information about the physical characteristics of crops, such as canopy area, height, density and volume, and texture features. The use of photogrammetry and image stitching techniques allows for the generation of orthomosaic (or orthophoto) images, digital surface models (DSM), digital terrain models (DTM), and crop surface models (CSM) for spatial analysis [9]. Pix4D Mapper and QGIS are examples of software packages that can process UAV imagery and perform geospatial analysis of the images by performing georeferencing, raster analysis and feature extraction. With the increase in high-resolution UAV datasets, there has also been an increase in the use of artificial intelligence (AI), machine learning (ML), and deep learning (DL) in agriculture. Agricultural systems are inherently dynamic and are influenced by nonlinear variables such as weather, soil properties, genotype of the plant, watering methods and fertilization methods. Because of the complexity of these relationships, conventional statistical models often perform poorly. Machine learning algorithms have the ability to analyze very large multidimensional datasets and identify hidden patterns that can be used to predict yield from a crop [10]. Various machine learning methods have demonstrated considerable results to forecast crop yields and classify agricultural data. These algorithms produce relationships amongst vegetation indices, crop structural characteristics, and respective/actual grow yield values which improves accuracy of predictions [11]. As well as conventional machine learning methods, deep learning algorithms have completely changed the way that image-based agricultural analysis is performed. Different types of deep learning architecture including Artificial Neural Networks (ANN), Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), and Long Short Term Memory (LSTM) can automatically discover multiple complex spacial (location) or temporal (time) attributes in large amounts of image file data [12]. The benefit of using CNN based

models for UAV (unmanned aerial vehicle) imagery data is that they provide the ability to extract spatial textures, edges, and canopy patterns from aerial images without the need to manually design features. The integration of UAV remote sensing with machine learning and deep learning techniques establishes an intelligent framework for crop monitoring and yield estimation. UAV-based precision agriculture systems allow for site-specific crop management through the identification of stressed areas, vegetation variability monitoring, and resource optimization. These systems also assist in reducing the excessive use of fertilizers, pesticides, and irrigation water, promoting sustainability and lessening environmental impact [13]. Additionally, precision agriculture technology supports the development of smart farming systems that leverage UAVs, IoT devices, cloud computing, robotics, artificial intelligence, and big data analytics to establish automated agricultural management. The adoption of intelligent agricultural monitoring systems is becoming increasingly necessary to enhance crop productivity, minimize operational expenses, and guarantee food security due to climate variability [14]. While UAVs can be used to monitor agricultural crops, there are still many issues that need to be resolved before getting an accurate estimate of crop yield. The data from agriculture is very heterogeneous and subject to many outside influences such as environmental variances. There are many complex workflows involved when processing the image data taken from UAVs and the different features in the images must be extracted in an efficient manner. Machine learning models must be optimized; otherwise overfitting will occur and generalizability will be compromised. Furthermore, little research has been done regarding the use of low-cost UAVs with RGB-type imagery for predicting crop yield using advanced machine learning methods, particularly in relation to regional ag. conditions. This research study developed a UAV-based crop yield estimation framework that incorporates RGB imagery, vegetation index analysis, GIS-based spatial analysis, and advanced machine learning techniques. The research study's primary focus is to examine how well different machine learning and deep learning algorithms can predict wheat crop yields using UAV aerial images. The framework was designed to be an automated, rapid, non-destructive, and low-cost solution targeted toward providing farmers with precision agriculture and sustainable farming solutions. In general, the contribution to the rapid advancement of smart agriculture is achieved through combining UAV remote sensing systems, remote-sensor imaging, GIS analysis, AI, and machine-learning to provide intelligent crop monitoring and facilitate data-supported decisions concerning agricultural production.

2.0 Methodology

The main purpose of the research methodology was to create an efficient and automated framework for predicting crop yields based on UAV remote sensing data and machine learning technologies. The entire workflow can be divided into a series of steps that include the collection of aerial images,

photogrammetric reconstruction, image preprocessing, vegetation analysis, feature extraction, and predictive modeling for precision agriculture purposes. The workflow began with a site survey and flight planning for the area of study to determine appropriate flight patterns to allow for proper aerial coverage of the experimental agricultural site. The area to be surveyed was broken down into several sampling areas in order to understand the variability of crops and improve the reliability of yield estimates. Additional parameters for the UAV flights (e.g., elevation, flight speed, percentage overlap of images, camera orientation) were optimized to ensure the collection of high resolution images for photogrammetric processing. Upon completion of the UAV imagery collection, photogrammetric processing was performed with the aid of advanced photogrammetric techniques in order to produce georectified orthomosaic maps and crop surface models. Image alignment, camera calibration, dense point cloud generation, and geometric correction were performed to increase the spatial accuracy of the images and eliminate any distortion seen within those images of the agricultural area. Once the rasterized images were orthorectified, they were then exported to GIS software for further analysis. Later on, orthomosaic images were imported into GIS software to do preprocessing and extract information about the spatial features of the image. The suitable operations used to enhance image quality so as to better depict the locations of crops included filtering noise and correcting radiometric distortions, adjusting for contrast, and segmenting the images. Within these raster images, vegetation was separated from soil by employing raster analysis methods and casualties or degree of fluctuation in crop production were identified based on the evaluation of vegetated portions of the field. Crop phenotypic and spectral characteristics were obtained from the UAV imagery and used to provide insights into the measurement of plant size and the health of the plant's leaf base. Some of the characteristics analysed include: Canopy Area, Plant Height, Crop Density, Tiller Coverage, Texture Characteristics, Canopy Structure, Biomass-related Characteristics. In addition to visual imagery analysis, vegetation indices obtained from RGB images were also calculated to determine the potential for growth and health of crops. Some of the vegetation indices used in the document included: GLA (Green Leaf Algorithm), VARI (Visible Ecosystem-based Atmospherically Resistant Index), RGBVI (Red Green Blue Vegetation Index), ExG (Excess Green Index), ExR (Excess Red Index), NGRDI (Normalized Green Red Difference Index) Data from the spectral and structural analyses were integrated into machine learning; deep learning approaches were used to estimate crop yields using predictive algorithms: KNN, DT, RF, SVM, and LASSO using a comparative analysis of performance. The data set was partitioned into training and test sets during model creation, validation hyperparameters were optimised and features were selected to maximize predictive accuracy and minimise complexity of the model. The models were evaluated with statistical metrics such as; Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2) and prediction accuracy. Spatial yield maps were generated based on the predicted outputs for visualising

crop production differences throughout the agricultural field. The methodologies outlined in this study represent a rapid, automated, non-destructive and scalable solution for precision agriculture.

2.1 Image Acquisition

The Unmanned Aerial Vehicle image acquisition method used to capture aerial images of wheat crop fields consisted of UAVs equipped with high resolution RGB cameras. Proper flight planning and procedure was used to ensure good image acquisition quality and accuracy in regards to data measurement for crop analysis and yield estimation. To ensure quality aerial image data collection using a DJI Phantom 4 UAV, which has a very stable flight performance, is equipped with an integrated GPS System, and produces high resolution images, it was determined that it was best to operate the UAVs under good weather conditions (i.e., fully sunny; no to very little wind) to decrease image distortion and motion blur. The images were collected between the hours of 12:00 PM – 2:00 PM to create consistent sunlight conditions for all images collected and reduce the amount of shadows within the crop canopy area.

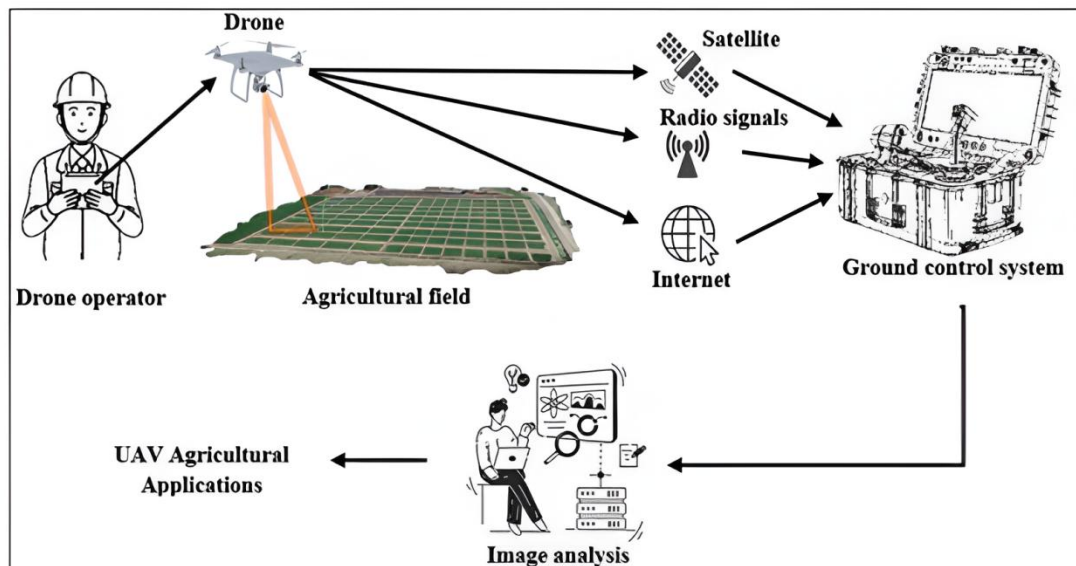


Fig 1: Methodological approach for UAV-based data acquisition



Fig 2: An image showing the DJI Phantom 4 drone used for aerial data acquisition

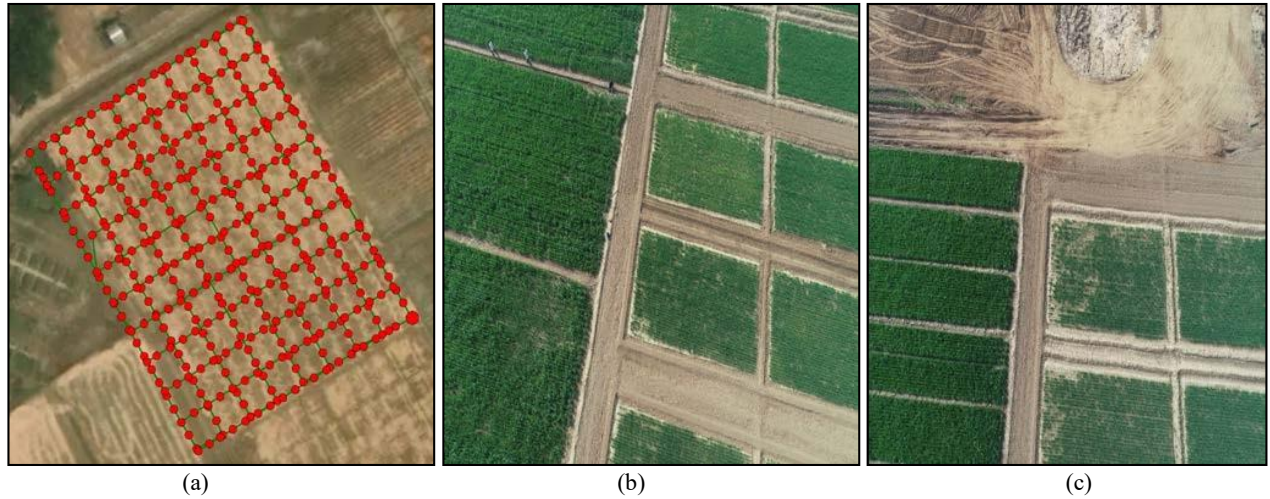


Fig 3: Aerial Data collection through UAVs: (a) Image Capture Locations, (b) and (c) Representative Captured Frames

2.2 Mosaicking of image followed by processing for analysis

1. Image Alignment & Orthomosaic Creation

To create a seamless georeferenced orthomosaic of the UAV images acquired, Structure-from-Motion (SF) techniques were used to align the images by identifying and matching common keypoint features across overlapping images. The matched features were then used to estimate the camera positions and orientations to create a sparse point cloud from which the images were orthorectified and subsequently stitched together.

2. Image Processing & Improvement

The orthomosaic was processed to add visual enhancement and increased reliability for analytical purposes. Radiometric correction was completed in order to produce images with consistent brightness and colour values. Noise reduction, contrast adjustment, and basic image segmentation were applied to the images to aid in the feature extraction process, which can be utilized for crop health monitoring and yield prediction.

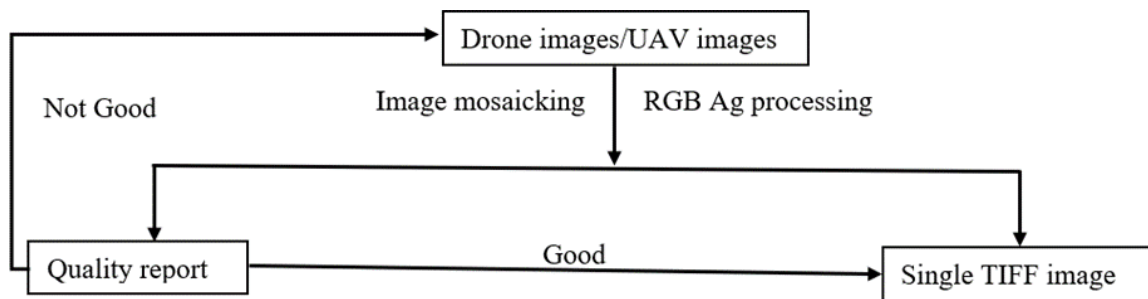


Fig. 4: Image processing was performed using Pix4Dmapper

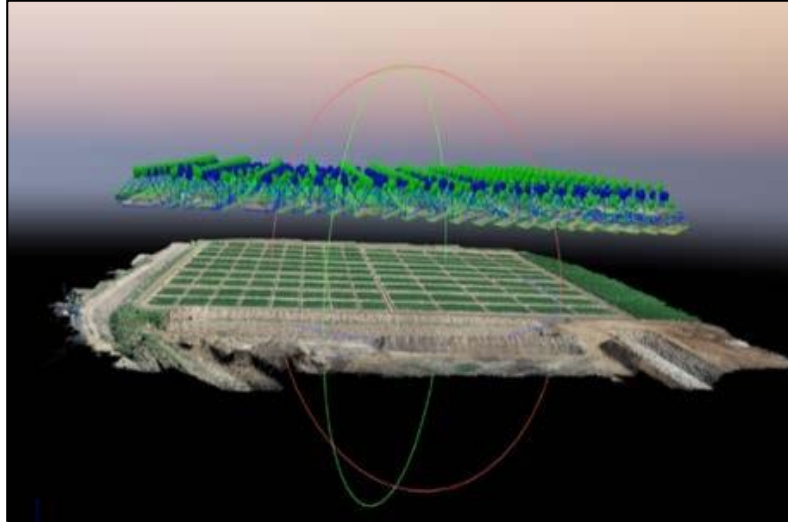


Fig. 5: Orthomosaic representation of the wheat field

The generated mosaic was utilized as a rasterized geographic layer in Quantum Geographic Information System software for the extraction of relevant crop parameters.

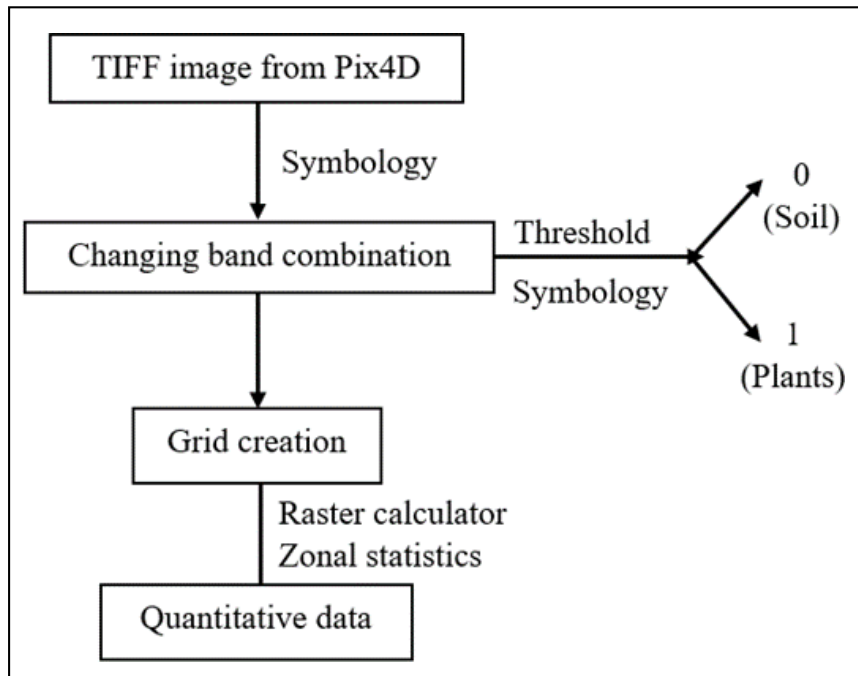


Fig. 6 : Image processing was performed using QGIS

2.3 Computation of Vegetation Index

The vegetation index (VI) is a mathematical combination of spectral bands that are designed to enhance the information associated with Vegetation and also provide a means for analyzing crop conditions based on remote sensing images. The use of vegetation indices in UAV derived precision agriculture will support the monitoring of crop health by determining the amount of vigour, chlorophyll concentration, biomass accumulation, canopy cover, and stress variability. As they can be derived through low-cost RGB cameras mounted on UAVs rather than expensive multispectral sensors, RGB vegetation indices have particular utility. In this study, several RGB vegetation indices were derived from orthomosaic

images using UAV image processing. These orthomosaic data were imported into the QGIS software environment and raster calculator tools or image processing techniques were used to derive the vegetation indices. The indices derived from these images were then analyzed for crop growth characteristics, covered area of vegetation and the physiological condition of the wheat crop in the field. Before carrying out the calculations of vegetation indices, UAV images were pre-processed in terms of radiometric correction, image normalisation and reduction of noise to improve image quality, and also to standardise the spectral information collected from the fields. The red (R), green (G) and blue (B) spectral bands from the RGB images served as the primary data for analyses of vegetation. The vegetation indices obtained included the Green Leaf Algorithm (GLA), the Visible Atmospherically Resistant Index (VARI), the Red Green Blue Vegetation Index (RGBVI), the Excess Green Index (ExG), Excess Red Index (ExR), Excess Green Minus Excess Red Index (ExGR), Water Index (WI), and Normalised Green-Red Difference Index (NGRDI). Therefore, these indices provide important information on crop vigour, chlorophyll content within vegetation, and the density of the vegetation, which gives us an idea of the condition of the crops under stress.

The use of GLA allowed for the enhancement of the green vegetation areas when distinguishing between healthy crops versus background soil. GLA functions best when analysing green canopy cover, therefore, it is highly effective in determining differences in vegetation density across the agricultural fields.

$$GLA = (2G - R - B) / (2G + R + B)$$

The Visibility Atmospheric Resilient Index (VARI) was calculated to minimize the illumination effects of the atmosphere and enhance the ability to detect vegetation in varying levels of light. VARI can be used with visible spectrum image data to detect healthy vegetation.

$$VARI = (G - R) / (G + R - B)$$

Crop vigor and biomass properties were assessed using the Red Green Blue Vegetation Index (RGBVI), which was computed using RGB imagery to provide improved vegetation responses by combining all three spectral components of red, green, and blue.

$$RGBVI = (G^2 - R \times B) / (G^2 + R \times B)$$

To highlight pixels of green vegetation and increase the ability to segment crops, The Excess Green Index (ExG) was created from the RGB contributions from different areas of soil or non-vegetated backgrounds.

$$ExG = 2G - R - B$$

The Excess Red Index (ExR) was created to study crop drought stress and identify those crop areas that have lower concentrations of chlorophyll.

$$ExR = (1.4) R - G$$

Combining both of these indices together into one value, the Excess Green minus Excess Red Index (ExGR) that allows for a more accurate discrimination of vegetation and more precise performance in crop segmentations.

$$\text{ExGR} = \text{ExG} - \text{ExR}$$

The Water Index (WI) utilized red, green, and blue (RGB) bands to assess water availability related to crop health and condition.

$$\text{WI} = (\text{G} - \text{B}) / (\text{R} - \text{B})$$

To assess crop health and vegetation density based on visible (RGB) imaging, the Normalized Green-Red Difference Index (NGRDI) was used.

$$\text{NGRDI} = (\text{G} - \text{R}) / (\text{G} + \text{R})$$

In addition to analysing vegetation indices, digital values derived from the spectral images such as Green/Red, Green/Blue, and Red/Blue were analysed to determine crop variability and spectral response, respectively. Correlation analysis of the vegetation indices and spectral parameter data with yield and some structural characteristics (canopy area and plant height) was carried out. By utilizing QGIS software to produce vegetation index maps, a spatial representation of the areas of healthy and stressed plants within the agricultural field can be obtained. Vegetation Index (VI) values that were higher than the VI value of an individual plant indicated an area of dense canopies and significant biomass accumulation, while low values assign an area of sparse vegetation to nutrient deficient or stress conditions. Computed vegetation indices were then utilized as input features to design machine learning (ML) and deep learning (DL) models to predict crop yields. The integration of vegetation analyses with UAV imagery and artificial intelligence techniques substantially increased the degree of accuracy and reliability of the proposed precision agriculture framework.

2.4 Crop Height Measurement

The estimation of the height of plants was based on remote sensing aerial images obtained by UAVs (unmanned aerial vehicles) and through photogrammetry. The UAV used RGB (red, green, blue) high-definition photographs and software to create the DSM (Digital Surface Model), DSM (Digital Terrain Model) of the soil after all plants and objects were removed from the scene providing the extent of soil moisture and surface materials; thus providing a measure of plant height. Estimation of plant height occurred by producing a Crop Surface Model (CSM) generated by taking a DSM (Digital Surface Model) image minus a DTM (Digital Terrain Model) image of the same area i.e

$$\text{CSM} = \text{DSM} - \text{DTM}$$

The UAV imagery (i.e., aerial photography) was aligned and transformed into a dense three-dimensional

point cloud (XYZ) that contains elevation data of the crop canopy and ground elevation. Raster datasets derived from the point cloud data were generated by applying GIS and photogrammetric techniques to calculate the height of the vegetation. QGIS provided an additional method of calculating the average plant height for a number of experimental plots using both raster analysis and zonal statistics. In order to increase the accuracy of the vegetation height estimates and to reduce the vertical errors, several pre-processing steps were employed: (1) noise filtering, (2) filtering of the point cloud, (3) smoothing the raster, and (4) correcting geometric distortions (i.e., re-projecting the raster). To validate the UAV-derived plant heights, field data was collected from a subset of the study area sampling locations for a comparison between the two sources of information (i.e., field vs. UAV), which resulted in a high correlation between these datasets and supports the validity and applicability of the proposed methods for estimating plant height. UAV-based estimates of canopy height were then correlated with vegetation indices and yield data from crops, allowing analysis of crop growth conditions and variability in biomass. Canopy height is often a good indicator of crop health and biomass accumulation when tall, but low canopy heights may indicate crop stress, lack of nutrients, or uneven growth. The use of UAV-based plant height estimation along with GIS and machine learning technologies to enhance precision agriculture will greatly increase the accuracy of crop monitoring, biomass estimation, and accurate prediction of crop yield.

2.5 Field Data Acquisition

Field data, including crop height and yield parameters, were collected during the maturity stage of the crop. Plant height measurements for each plot were obtained by calculating the average of 15 individual observations.

2.6 Advanced ML Methods for Crop Yield Analysis

The analysis was carried out using advanced machine learning methods, discussed below

2.6.1 K-Nearest Neighbor (KNN)

K-Nearest Neighbor Regression (KNN) is a supervised learning approach used to estimate continuous output values by analyzing the similarity between nearby data samples [16]. In crop yield estimation, the algorithm predicts yield values based on neighboring observations with similar vegetation and canopy characteristics [21]. KNN is useful for agricultural datasets because it can model nonlinear relationships without requiring complex assumptions about the data distribution [22]. The algorithm is simple, flexible, and effective for analyzing UAV-derived crop features and vegetation indices [16][22].

2.6.2 Decision Tree (DT)

Decision Tree (DT) is a machine learning technique that uses a hierarchical tree-like structure to make predictions based on feature conditions [17]. The algorithm divides the dataset into smaller subsets using decision rules generated from input variables such as vegetation indices, canopy height, and crop density

[23]. Decision trees are widely used in precision agriculture because they are easy to understand and capable of handling nonlinear agricultural data efficiently [24]. DT models can support crop monitoring, yield prediction, and classification tasks with relatively low computational requirements [17][24].

2.6.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm designed for regression and classification problems involving complex datasets [18]. The method works by finding an optimal decision boundary that minimizes prediction error while maximizing the separation between data samples [25]. SVM performs effectively with high-dimensional agricultural data and is commonly applied in UAV-based crop analysis and vegetation monitoring [26]. Its strong generalization ability and high prediction accuracy make it suitable for crop yield estimation applications [18][26].

2.6.4 Random Forest (RF)

Random Forest (RF) is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting [19]. Each tree is generated using random subsets of the dataset, and the final prediction is obtained by combining the outputs of all trees [27]. RF is highly effective for agricultural applications because it can process large datasets and identify complex relationships among vegetation indices, canopy features, and crop productivity [28]. The algorithm is widely used for yield estimation, crop classification, and biomass analysis using UAV imagery [19][28].

2.6.5 Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO is a shrinkage-based regression technique that applies regularization to improve model performance. that performs feature selection while building predictive models [20]. The algorithm reduces the influence of less important variables by shrinking their coefficients toward zero, thereby improving model stability and reducing overfitting [29]. In crop yield estimation studies, LASSO helps identify the most significant vegetation and canopy parameters affecting crop productivity [30]. The method is particularly useful when working with high-dimensional agricultural datasets containing multiple correlated features [20][30].

2.7 Evaluation of Predictive Models

The five machine learning algorithms were evaluated using root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2). The formula for each metric is given below:

RMSE (Root Mean Square Error)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

MAE (Mean Absolute Error)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

R² (Coefficient of Determination)

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

n (i = 1, 2, . . . , n) represents the number of observations in building the machine learning model. The obtained winter wheat productivity values is denoted y_i; the estimated mean productivity value is represented by y_i; the predicted wheat yield (from our machine learning model) is denoted f_i; and the corresponding estimated yield (mean) is denoted f_i. The farther the value of the coefficient of determination R² from one, the less accurate the model has been in estimating yield. The amount of variability of actual versus estimated yield values is represented by reasonable RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) values.

3.0 Result

3.1 Comparative Analysis of Wheat Yield Prediction Models

Different machine learning models were evaluated against each other to determine how well they could be used to predict wheat yield based on UAV-derived vegetation/canopy features. A wide variety of supervised learning algorithms (KNN, DT, SVM, RF, LASSO) were used with crop parameters and vegetation indices as features. The results of the comparative analysis showed significant differences in prediction ability between the different models.

KNN had a stable learning behaviour and provided acceptable prediction accuracy for nonlinear crop data. However, the choice of neighbouring samples and feature distribution impacted its performance. Although Decision Tree models trained quickly and were easy to interpret, they were moderately consistent in terms of prediction due to sensitivity to changes within the data used for training.

The utilization of support vector machines provided high accuracy when predicting because of their ability to accurately find non-linear relationships and process large agricultural databases with high accuracy; and, by combining many decision trees into an ensemble learning approach with random forests.

Random forests were able to produce very accurate results and reduce the amount of overfitting and produced predictable generalization from the model and able to provide very good robustness against agricultural data that was non-consistent.

LASSO regression had the ability to decrease the complexity of the model by selecting features or by providing rules for coefficient regularization. The model produced the most significant contributors to

wheat yields and also reduced their contribution during training.

Overall, based on this study, using random forests and support vector machines provided superior training accuracy compared to the other algorithms, because they could process UAV derived complex data sets more effectively than any of the other algorithms. In addition, the analysis in this study showed that using vegetation indices, the height of the canopy and spatial features of the crops, was able to provide a significant improvement in the machine learning based prediction of yield.

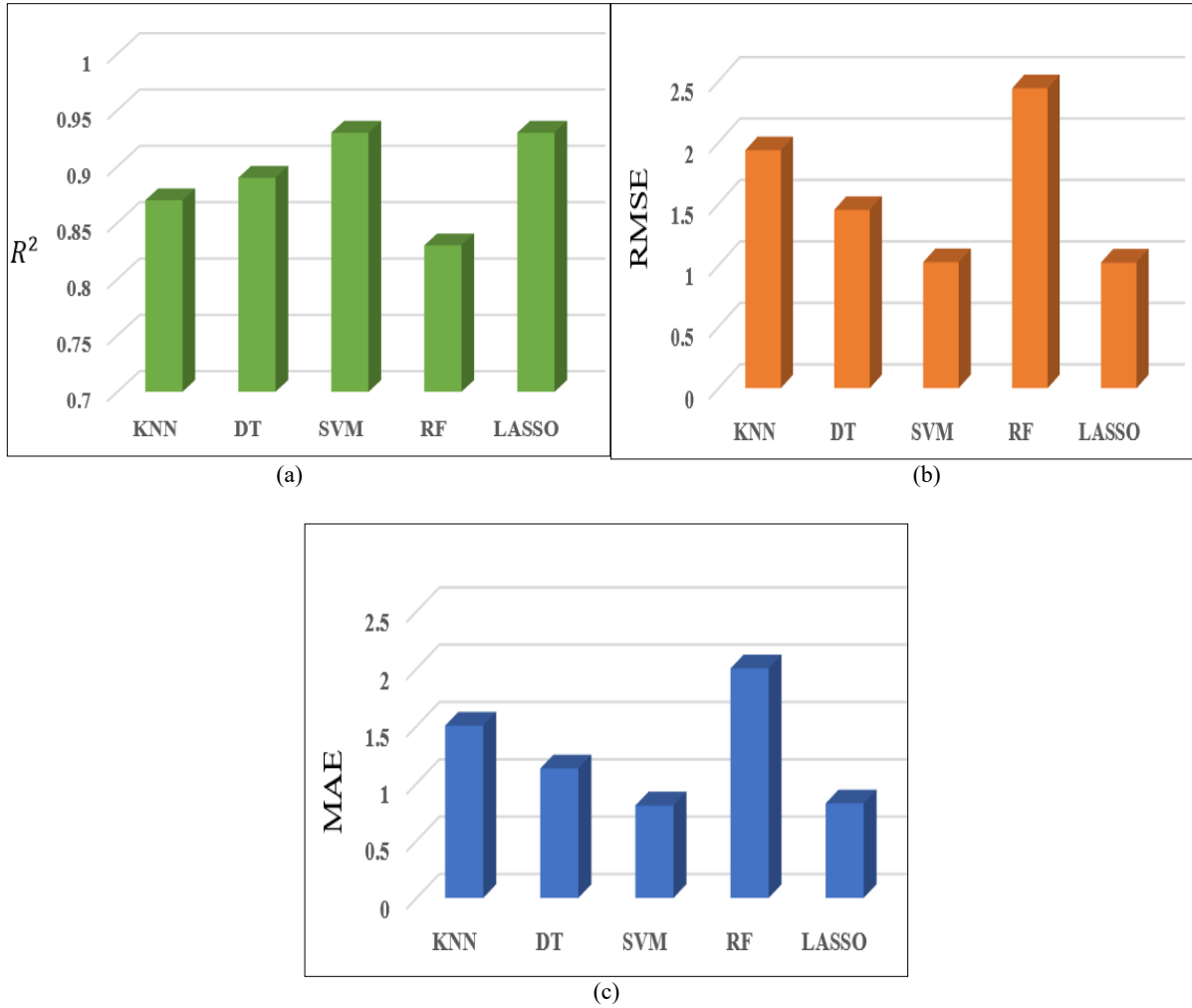


Fig. 7: (a), (b), (c) are Comparative Evaluation Metrics of ML Algorithms

3.2 Wheat Yield Estimation Using UAV Imagery

Five different predictive method (models) based upon UAV image data have now been applied; these models provide scatter diagrams (See Figure 8.) to illustrate predictions versus actual results. In the above study, we found that the two models – LASSO > SVM > DT > RF > KNN can accurately forecast wheat yield. Despite the fact that all expected yields were significantly nearer the line.

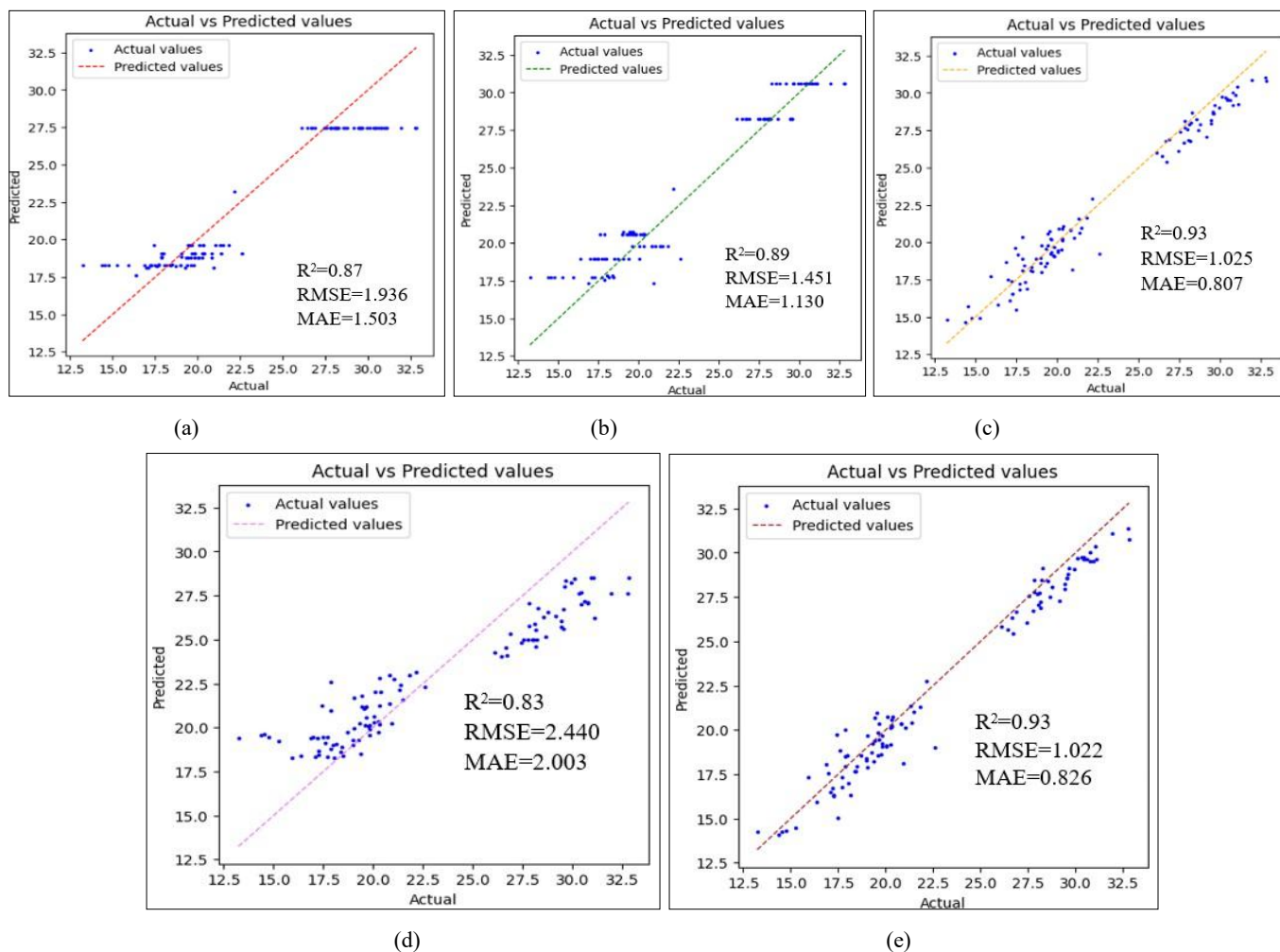


Fig. 8: Comparison of Actual and Predicted Yield Using Different ML Models i.e a) KNN, b) DT, c) SVM, d) RF, e) LASSO

Table 1: Relationship Matrix Between Crop Yield and Spectral Parameters Obtained from UAV Images

Variables	Yield (kg/plot)	Canopy area (sqm)	Mean canopy height (m)	GLA	WI	RGBVI	VARI	NGRDI	ExG	ExR	ExGR	G/R	G/B	R/B
Yield (kg/plot)	1													
Canopy area (sqm)	0.931***	1												
Mean canopy height (m)	0.388***	0.203***	1											
GLA	0.210***	0.360***	0.356***	1										
WI	-0.169***	-0.240**	0.206***	-0.248***	1									
RGBVI	0.364***	0.189***	0.798***	0.533***	0.033**	1								
VARI	0.416***	0.289***	0.832***	0.615***	0.118***	0.914***	1							
NGRDI	0.355***	0.069***	0.806***	0.078***	0.260***	0.858***	0.808***	1						
ExG	-0.007***	-0.302**	0.111***	-0.830***	0.318***	0.029***	-0.124***	0.475***	1					
ExR	-0.073***	0.244***	-0.255***	0.717***	-0.309**	-0.205***	-0.058***	-0.626***	-0.981***	1				
ExGR	0.119***	-0.205**	0.392***	-0.534***	0.241***	0.403***	0.255***	0.757***	0.896***	-0.957***	1			
G/R	0.378***	0.168***	0.878***	0.405***	0.181***	0.944***	0.945***	0.933***	0.148***	-0.318***	0.495***	1		
G/B	0.164**	0.249**	0.376***	0.909***	-0.311***	0.651***	0.562***	0.178***	-0.643***	0.533***	-0.347***	0.457***	1	
R/B	-0.401***	-0.294***	-0.722***	-0.594***	-0.316***	-0.596***	-0.854***	-0.794***	-0.412***	0.758***	-0.616***	-0.776***	-0.142***	1

Note: *, **, and *** indicate statistical significance at $p < 0.05$, $p < 0.01$, and $p < 0.001$ levels, respectively.

3.3 Correlation Between Vegetation Indices and Wheat Yield

The vegetation indices (VIs) from UAV-RGB (aerial) images indicated that there was a good correlation between wheat yield and their use as measures of plant growth, canopy vigour and biomass accumulation. The calculated VIs provided important spectral measurements regarding chlorophyll concentration, vegetation density and the physiological state of the wheat plants. These indices were then used to correlate with the actual observed yield of the wheat plants and to determine how well they functioned as measures of crop productivity. Among the VIs evaluated, the Visible Atmospherically Resistant Index (VARI) showed a high positive correlation with wheat yield and demonstrated its effectiveness as a measure of identifying healthy vegetation and identifying denser populations of wheat plants. Similarly, the Red Green Blue Vegetation Index (RGBVI) had a strong positive relationship with wheat yield because of its ability to enhance the spectral response of vegetation viewed in RGB image. The Green Leaf Algorithm (GLA) also had a positive correlation with wheat yield because it could distinguish between the green vegetation and soil background as well as highlight the differences in vigour of the wheat within the field. Additionally, the Normalized Green-Red Difference Index (NGRDI) had a moderate correlation with wheat yield and provided useful information for vegetation density assessment and monitoring crop growth. Excess Red Index (ExR) and Red/Blue (R/B) ratio had negative correlations

to wheat yields indicating that greater dominance of red spectral energy was typically associated with stressed vegetation or lower chlorophyll content or suboptimal crop growth. Excess Green Index (ExG) had weaker relationships with yield, however, it was still useful in segmenting vegetation and extracting crop areas. The correlation analysis demonstrated that vegetation indices related to a healthy green canopy produced generally positive correlations with wheat productivity. VARI demonstrated the strongest association with wheat yield among all the vegetation indices analyzed, achieving a correlation coefficient (r) of 0.391.

3.4 Assessment of Variable Significance in Prediction Models

We computed the reduced accuracy (mean square error) from the RF model in order to examine the significance of various factors on the wheat yield forecast. The predictor variable relevance according to the RF model is shown in Figure 9.

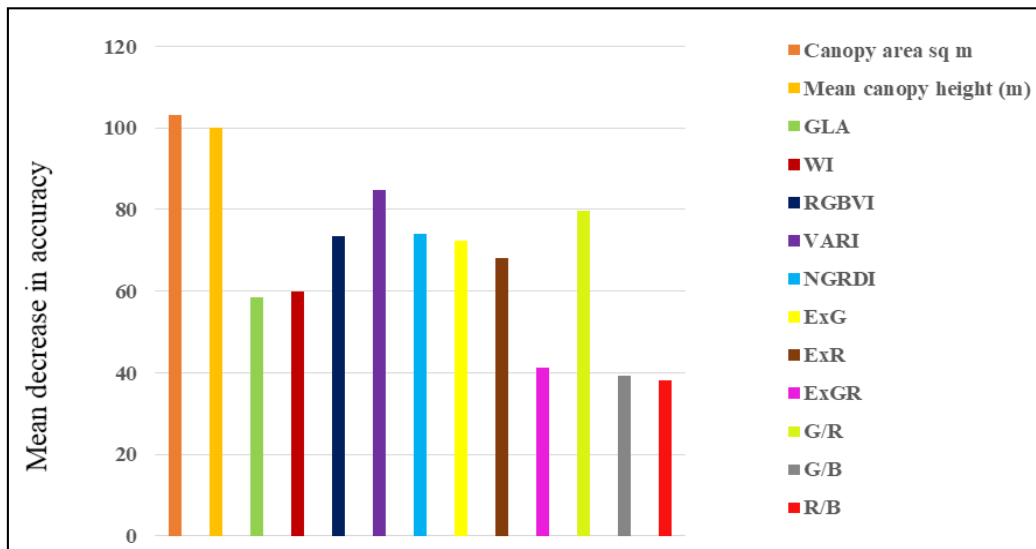


Fig. 9: Variable Importance Analysis Using the Random Forest Model

4.0 Discussion of Findings

4.1 Performance Evaluation of Yield Estimation Models

A number of different types of models were used to evaluate yield estimates for wheat using vegetation indices and canopy parameters obtained from UAVs. The machine learning models used to analyze this data included K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF) and LASSO Regression with their performance evaluated based on. Statistical evaluations for the various models were performed using the following indicators: RMSE, MAE and coefficient of determination (R^2). The SVM and Random Forest models produced superior predictions for yield due to their capability to model complex nonlinear agricultural datasets. The SVM and Random Forest yielded greater accuracy and stability with respect to their yield predictions than the other types of models. KNN yielded satisfactory results for the relationships between local data; DT produced simple

yield interpretation with moderate predictive capability. The LASSO Regression model was useful in selecting yield-related features as well as reducing extraneous variables from the analyses. Overall, the overall analysis demonstrated that the combination of UAV-derived vegetation indices and machine learning produced a significant improvement in accurate estimations of yields for wheat. Overall, this study indicates that combined methods of remote sensing and machine learning through UAVs provide a fast, effective, non-destructive means to accurately assess effective precision agriculture and estimating crop productivity.

4.2 Relationship Evaluation of Drone Imagery and Wheat Productivity

Wheat yield has an excellent relationship to UAV images, giving a clear and detailed picture of crop growth, canopy structure, vegetation health and variability across the field [21]. UAV airborne imagery provides the opportunity for assessing crop conditions through spectral and structural assessment methodologies [22]. UAV imagery has improved spatial resolution compared to standard field survey techniques, and UAV provides a means of rapidly collecting data from a large area and efficient monitoring of crops. UAV will provide an essential tool for precision agriculture [23]. From UAV-RGB imagery-derived vegetation indices (VIs), including VARI, RGBVI, GLA and NGRDI, showed notable positive correlation with wheat yield via chlorophyll activity, canopy health and biomass accumulation [24]. Vegetation index (VI) values had a general response in terms of healthier crop growth being representative of potential for more productive crops [25]. Spatial differences in the response of crops to different zones of productivity within the agricultural field could also be evaluated utilizing UAV imagery spectral responses. Spectral parameters also demonstrated meaningful relationships with structural crop attributes collected and quantified from UAV images such as canopies, plant height and vegetation cover, all of which were also relatively correlated with wheat productivity. Large canopies, greater heights and an overall estimated biomass were all factors that would typically be associated with an increased yield of grain produced [28]. As such, canopy-related traits derived from UAV imagery are reliable indicators of both assessing crop performance and identifying variability within agricultural fields [29]. UAV (unmanned aerial vehicle) imagery and machine learning model integration led to the increased precision of estimates of wheat yield [30]. The machine learning algorithms utilized vegetation indices and parameters of the canopy effectively to establish a correlation between crop condition and yield value. The findings suggest that the combination of UAV-based remote sensing with artificial intelligence enables an effective framework for non-destructive precision agriculture and intelligent crop monitoring systems [32]. Overall, the research supports that UAV-derived image data will provide a reliable source for assessing wheat yield, evaluating crop health, and implementing sustainable agricultural management practices.

5.0 Conclusions

In this study we demonstrated the feasibility of using UAV-based images combined with advanced machine learning techniques to estimate wheat yields and for precision agriculture. UAV-acquired imagery at high spatial resolution combined with crop vegetation indices and canopy structure parameters provided extensive information about how crops develop, their health, the accumulation of biomass, and variability within each field. Crop vegetation indices such as VARI, RGBVI, GLA, and NGRDI exhibited strong correlation with wheat yields and could be utilized to evaluate vegetation condition and crop vigor. Canopy structure parameters such as canopy area and height also were important predictors of yield. Numerous machine learning algorithms including KNN, Decision Tree, SVM, Random Forest, and LASSO Regression were implemented and compared for yield prediction. The SVM and LASSO Regression algorithms achieved the best prediction accuracy and lowest prediction error SVM (RMSE = 1.025; $R^2 = 0.93$); LASSO Regression (RMSE = 1.022; $R^2 = 0.93$) respectively demonstrating high correlation between the estimated and actual wheat yield values. Comparison of effectiveness of UAV image analysis, GIS analysis, vegetation index calculation and ML based analyses in determining crop yield respective to traditional field survey methods for estimating crop yield has improved efficiency and reliability. As a result, this work provides a quick, non-destructive and low cost alternative to existing solutions for precision agriculture and crop monitoring as well as a means to make intelligent agricultural decisions.

5.1 Funding Declaration

This research did not receive any external funding.

References

- [1] McCown, R.; Hammer, G.; Hargreaves, J.; Holzworth, D.; Freebairn, D. APSIM: A novel software system for model development, model testing and simulation in agricultural systems research. *Agric. Syst.* 2021, vol.50, pp. 255–271.
- [2] Sadeh, R.; Avneri, A.; Tubul, Y.; Lati, R.N.; Bonfil, D.J.; Peleg, Z.; Herrmann, Chickpea leaf water potential estimation from ground and VENUS satellite. *Precis. Agric.* 2024, vol. 25, pp.1658–1683.
- [3] Sonam Ahuja, Manjunath HR, Intekhab Alam, Ajay Rastogi “Vegetable crop biomass estimation using hyperspectral and rgb 3d uav data,” *Agronomy*, 2022, vol. 10, pp.1600-1626.
- [4] E Mehdi Raouhi, Mohamed Lachgar, Hamid Hrimech and Ali Kartit “Rice yield prediction model using data mining,” in 2020 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE, 2020, pp.321–326.

- [5] G. Gupta, R. Setia, A. Meena, and B. Jaint, "Environment monitoring system for agricultural application using iot and predicting crop yield using various data mining techniques," in 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2020, pp. 1019–1025.
- [6] A. S. Terliksiz and D. T. Alt'ylar "Development of an Adaptive Approach for Precision Agriculture Monitoring with Drone and Satellite Data", IEEE J Sel Top Appl Earth Obs Remote Sens, vol 10, 2023, pp. 5322-5328.
- [7] Daryl H, Hepting H, Maciag T, Hill H. Web-Based Support of Crop Selection for Climate Adaptation. 45th Hawaii International Conference on System Sciences, IEEE, Maui, HI. 2012. pp. 1227–605.
- [8] V.C. Waikar, S.Y. Thorat, A.A. Ghute, P.P. Rajput, M.S. Shinde, Crop prediction based on soil classification using machine learning with classifier ensembling, International Research Journal of Engineering and Technology 7 (May) (2022). pp. 4857–4861.
- [9] B Sankara Babu, A Suneetha, G Charles Babu, Y. Jeevan Nagendra Kumar, G Karuna, "Medical Disease Prediction using Grey Wolf optimization and Auto Encoder based Recurrent Neural Network", Periodicals of Engineering and Natural Sciences, June (2022) ISSN 2303-4521 Vol.6, No.1, pp. 229-240.
- [10] Alvarez, R. (2023). "Predicting average regional yield and production of wheat in the Argentine pampas by an artificial neural network approach". Eur. J. Agron. Vol.30, pp.70–77.
- [11] Romero, J. R., Roncallo, P. F., Akkiraju, P. C., Ponzoni, I., Echenique, V. C., and Carballido, J. A. (2020). Using classification algorithms for predicting durum wheat yield in the province of buenos aires. Comput. Electron. Agric. Vol.96, pp.173–179.
- [12] Rashid, M.; Bari, B.S.; Yusup, Y.; Kamaruddin, M.A.; Khan, N. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Wheat Yield Prediction. IEEE Access 2021, Vol.9, pp.63406–63439.
- [13] S. Fei, M. A. Hassan, Y. Xiao and Y. Ma, "UAV-Based Multi-Sensor Data Fusion and Machine Learning Algorithm for Yield Prediction in Wheat", *Precision Agriculture*, vol. 23, no. 4, 2022, pp. 1563-1585.
- [14] S. Yang, L. Li, S. Fei and Y. Xiao, "Wheat Yield Estimation Using Machine Learning Method Based on UAV Remote Sensing Data", *Drones*, vol. 8, no. 5, 2024, pp. 1-21.
- [15] T. Liu, Y. Wang and H. Zhang, "The Estimation of Wheat Yield Combined with UAV Canopy Information and Machine Learning Methods", *Frontiers in Environmental Science*, vol. 12, 2024, pp. 1-15.
- [16] R. Tsukahara, Y. Koga and T. Watanabe, "Winter Wheat Yield Prediction Using Convolutional Neural Networks and UAV-Based Multispectral Imagery", *Field Crops Research*, vol. 295, 2023, pp. 108907-108918.
- [17] A. K. Srivastava, N. Safaei, S. Khaki, G. Lopez, W. Zeng and J. Rahimi, "Winter

Wheat Yield Prediction Using Convolutional Neural Networks from Environmental and Phenological Data”, *Computers and Electronics in Agriculture*, vol. 184, 2021, pp. 106091-106102.

[18] Y. Bansal, D. Lillis and M. T. Kechadi, “Winter Wheat Crop Yield Prediction on Multiple Heterogeneous Datasets Using Machine Learning”, *arXiv Preprint arXiv:2306.11946*, 2023, pp. 1-14.

[19] D. Pathak, M. Miranda, F. Mena and P. Helber, “Predicting Crop Yield With Machine Learning: An Extensive Analysis of Input Modalities and Models on a Field and Sub-Field Level”, *arXiv Preprint arXiv:2308.08948*, 2023, pp. 1-18.

[20] H. dos Santos Felipetto, R. Silva and A. Ferreira, “Combining Machine Learning with UAV Derived Multispectral Imagery for Wheat Grain Yield Estimation”, *Geomatics, Natural Hazards and Risk*, vol. 16, no. 1, 2025, pp. 1-20.

[21] K. Kešelj, M. Novak and D. Petrović, “Machine Learning (AutoML)-Driven Wheat Yield Prediction Using UAV-Based Remote Sensing”, *Agriculture*, vol. 15, no. 14, 2025, pp. 1534-1552.

[22] S. B. Rufaioglu, M. Demir and T. Kaya, “CNN-Based Wheat Yield Prediction Using Multi-Source and Time-Series UAV Data”, *Information Processing in Agriculture*, vol. 13, no. 1, 2026, pp. 1-16.

[23] R. Tsukahara, Y. Koga and T. Watanabe, “Winter Wheat Yield Prediction Using Convolutional Neural Networks and UAV-Based Multispectral Imagery”, *Field Crops Research*, vol. 295, 2023, pp. 108907-108918.

[24] A. K. Srivastava, N. Safaei, S. Khaki, G. Lopez, W. Zeng and J. Rahimi, “Winter Wheat Yield Prediction Using Convolutional Neural Networks from Environmental and Phenological Data”, *Computers and Electronics in Agriculture*, vol. 184, 2021, pp. 106091-106102.

[25] Y. Bansal, D. Lillis and M. T. Kechadi, “Winter Wheat Crop Yield Prediction on Multiple Heterogeneous Datasets Using Machine Learning”, *arXiv Preprint arXiv:2306.11946*, 2023, pp. 1-14.

[26] D. Pathak, M. Miranda, F. Mena, P. Helber and A. Dengel, “Predicting Crop Yield With Machine Learning: An Extensive Analysis of Input Modalities and Models on Field and Sub-Field Level”, *arXiv Preprint arXiv:2308.08948*, 2023, pp. 1-18.

[27] F. Z. Bassine, T. E. Epule, A. Kechchour and A. Chehbouni, “Recent Applications of Machine Learning, Remote Sensing, and IoT Approaches in Yield Prediction: A Critical Review”, *arXiv Preprint arXiv:2306.04566*, 2023, pp. 1-25.

[28] H. dos Santos Felipetto, R. Silva and A. Ferreira, “Combining Machine Learning with UAV Derived Multispectral Imagery for Wheat Grain Yield Estimation”, *Geomatics, Natural Hazards and Risk*, vol. 16, no. 1, 2025, pp. 1-20.

[29] S. A. K. Shariati, R. Miller and D. Jones, “Machine Learning-Based Winter Wheat Yield Prediction Using Multisource Data”, *Agricultural Water Management*, vol. 322, 2025, pp. 109951-109964.

[30] S. B. Rufaioglu, M. Demir and T. Kaya, “CNN-Based Wheat Yield Prediction Using Multi-Source and Time-Series UAV Data”, *Information Processing in Agriculture*, vol. 13, no. 1, 2026, pp. 1-16.

[31] U. V. Nikhil, P. Kumar and A. Sharma, “Machine Learning-Based Crop Yield Prediction in South Asia”, *Computers*, vol. 13, no. 6, 2024, pp. 137-152.