

Yield prediction of rainy season onion crop using UAV-mounted multispectral sensors and machine learning tools.

Sagar M. Wayal¹, Shardul Parab², Anusha, R¹, Kiran Khandagale¹, Sanket Bhegde², Mukund

Dawale¹, Indira Bhangare¹, Mahesh Khaire¹, Yogesh Kadam¹, Zafar Shaikh², V. Karuppaih¹,

Pranjali Ghodke¹, Bhushan Bibwe¹, Suresh J. Gawande¹

¹ICAR-Directorate of Onion and Garlic Research, Pune, India

²TIH Foundation for IoT & IoE, Mumbai, India

Abstract:

Precision agriculture is a management approach that utilizes UAV-assisted remote sensing, combined with IoT and IoE technologies, to analyze spatial and temporal variations in agricultural fields, thereby optimizing profitability, sustainability, and protecting agro-ecological services. Present study focused on rainy season onion trials with four different planting dates, employing multispectral images to develop crop growth models and predict yield. Spectral indices serve as key indicators of crop growth, real-time non-destructive monitoring of these indices and crop morphological observations are essential for diagnosing crop health status as well as forecasting bulb yield. The field spectral images from various trials were collected by UAV-equipped multispectral camera at all onion crop growth stages. We extracted onion canopy reflectance mosaic (to use in pix4D mapper) and calculated multiple vegetation indexes (using the FieldimageR package in R software) for monitoring onion growth and yield. Simple linear regression and machine learning approaches were employed to build a mathematical model for estimating onion yield. From simple linear regression modeling in early planted rainy season onion NDRE and SR (for first trial 65 DAT, second trial 63 DAT and third trial 77 DAT) were the best predictor, but late planted rainy season's onion SAVI, LAI and GNDVI (for trial four 66 DAT) vegetation index was the best predictor. ML method (i.e. linear regression, Decision tree, Random forest, Support vector regression, Xgboost) supported better for estimating onion yield at the bulb development stage of onion crop. SVR consistently performed well on both train and test sets, showing good generalization, while RF models tended to good fit, especially in the test results. Our study finds that rainy season onion trials suggest variability in the data, impacting model performance. Both RF and SVR models handle this variability effectively and generalize well to predict Kharif season onion yields. The present study, demonstrated that the multispectral imagery can be used for reliable growth and yield prediction of rainy season onion.

Keywords: UAV, multispectral images, precision agriculture, machine learning, remote sensing onion yield prediction, spectral indices.

Abbreviations:

UAV – Unmanned Aerial Vehicle.
IoT – Internet of Thing.
IoE – Internet of Everything.
DAT – Date after Transplanting.
ML – Machine Learning.
PCA – Principle component analysis.
LR – Linear regression.
RF – Random Forest.
DT – Decision Tree.
SVR – Support vector Regression.
XGBoost – extreme gradient bootstrapping.
RSME – Root mean square error.
MAE – Mean absolute error.
VI's – Vegetation Indices.

Introduction

Onion is an essential vegetable and spice crop consumed worldwide. It plays a crucial role in various cuisines, adding flavor and depth to countless dishes. Whether used raw in salads, caramelized in savory dishes or as a base for flavorful stocks and sauces, their significance in cooking is undeniable. Additionally, nutritional value and long shelf life make onions a staple in households globally. India stands as the world's largest onion producer, contributing 30.19 million tones to global production in 2022-23 (FAOSTAT, 2024). In India Onion is cultivated during monsoon (June-September) and post-monsoon (October-March) seasons, with the former being rainfed and the latter being irrigated. While the rainy season onion constitutes around 20-25% of India's total onion production, its significance lies in its timely arrival following the depletion of stored post-monsoon or rabi season onions (Gadge and Lawande, 2012). Failures in the rainy season harvest, often caused by a range of biotic and abiotic stresses, lead to dramatic price spikes due to the shortage of fresh supply. Hence, conducting early assessments of yield and losses in rainy season crop becomes crucial. This proactive approach aids researchers and policymakers in anticipating price movements and devising contingency plans to manage price fluctuations effectively. The traditional method for predicting yield and assessing losses is laborious and time-consuming and has low operational efficiency for monitoring over a wide area. Remote sensing offers an alternative approach for swiftly and impartially scouting and measuring crop growth and various stresses (Usha and Singh, 2013). High-altitude remote sensing tools like Landsat satellites are effective for large-scale monitoring, but they suffer from low resolution, longer return periods,

and susceptibility to weather conditions. In recent times, agricultural researchers have increasingly focused on utilizing unmanned aerial vehicles (UAVs) in various applications, including disease detection, plant health monitoring, and precise pesticide application (Faiçal et al. 2014). The UAV's popularity in precision agriculture stems from its ease of use and ability to capture high-resolution imagery at close range, a capability not easily achievable with satellite images (Maes and Steppe, 2019). In the past, UAV-aided multispectral imagery has facilitated the modeling of various biotic and abiotic stresses. For instance, literature reports high-resolution images enabling automatic disease detection in different crops such as wheat yellow rust (Su et al., 2018), potato leaf blight (Rodriguez et al., 2021), sugarcane white leaf disease (Narmilan et al., 2022), fusarium wilt (Panama disease) of banana (Ye et al., 2020), cotton leaf blight (Wang et al., 2023), peanut leaf wilt (Chen et al., 2020) and tomato spot wilt diseases (Abdulridha et al., 2020). Furthermore, UAV-aided remote sensing has proven to be an effective tool for monitoring crop growth and biomass (Fu et al., 2014). It provides high spatio-temporal resolution imagery for agricultural applications and has been widely employed for monitoring and predicting growth and yield using RGB, hyperspectral, and multispectral sensors (Zhang et al., 2012). Vegetation indices (VI) such as normalized difference vegetation index (NDVI), ratio vegetation index (RVI), and leaf area index (LAI) have long been utilized for monitoring crop growth and yield prediction of rice (Zhou et al., 2017; Din et al., 2017), wheat (Fu et al., 2020; Xie et al., 2014), corn (Geipel et al., 2014), barley (Bendig et al., 2015), onion (Córcoles et al., 2013) and other crops. These studies were based on limited datasets, mostly acquired at a single stage, and the models were predominantly based on NDVI, which has limitations; for example, NDVI saturates when the canopy is dense (Liang, 2004). The performance and potential of these VIs for crop yield prediction should be further tested. LAI is an important parameter that indicates crop photosynthesis and growth status and a significant parameter for crop yield prediction (Noureldin et al., 2013; Verger et al., 2014). Some other studies also attempted to predict crop yield using plant height with Crop Surface Models (CSMs) (Bendig et al., 2014, 2015; Geipel et al., 2014). However, plant height alone is not the yield-determining trait for many crops, particularly vegetable crops. Furthermore, the models reported in these studies are based on a single sowing date, and the behavior of vegetation indices (VIs) to intra-seasonal variation was not studied in depth. (Marino et al., 2015) studied the feasibility of using spectroradiometric measures at the field size to evaluate onion yield. In onion crops, though there are few studies about multispectral imagery and crop growth monitoring but

these studies are preliminary and based on a limited dataset. A comprehensive account of multispectral imagery-derived VIs with yield prediction is not available to date. Therefore, in this study, we have given a detailed account of onion crop yield prediction by 1) A systematic analysis of various VI's in relation to yield and 2) Dynamics of various VIs to intra-season variation.

2. Materials and methods:

2.1. Experiment design and study area:

The experiment was conducted during *Kharif* 2023 at ICAR-Directorate of Onion and Garlic Research, Pune, Maharashtra, India, centered at latitude 18°50'27.99"N, longitude 73°53'12.88"E EPSG:4326 WGS 84/UTM zone 43N. The average temperature and monthly total precipitation during the study period (1 Jul to 31 Dec 2023) are shown in Fig.1. The trials were carried out in a randomized block design with three replications. Such four trials were conducted at 15 days intervals as shown in Fig 2.

Bhima Super variety of red onion recommended for *kharif* season cultivation was used in the present study. The nursery was raised on Broad bed furrows (BBF) of 15 cm height and 120 cm top width with 45 cm furrow at intervals of 15 days. Transplanting was done after 45-50 days on BFF with a spacing of 10x15 cm. Before transplanting, seedlings' were treated with carbendazim solution (1 g/L) and Carbosulfan 2 ml/L. The dates of transplanting and harvesting are mentioned in Table 1. Onions were harvested at maturity (50% neck fall) after 90-115 days of transplanting; marketable yield in each treatment was measured.

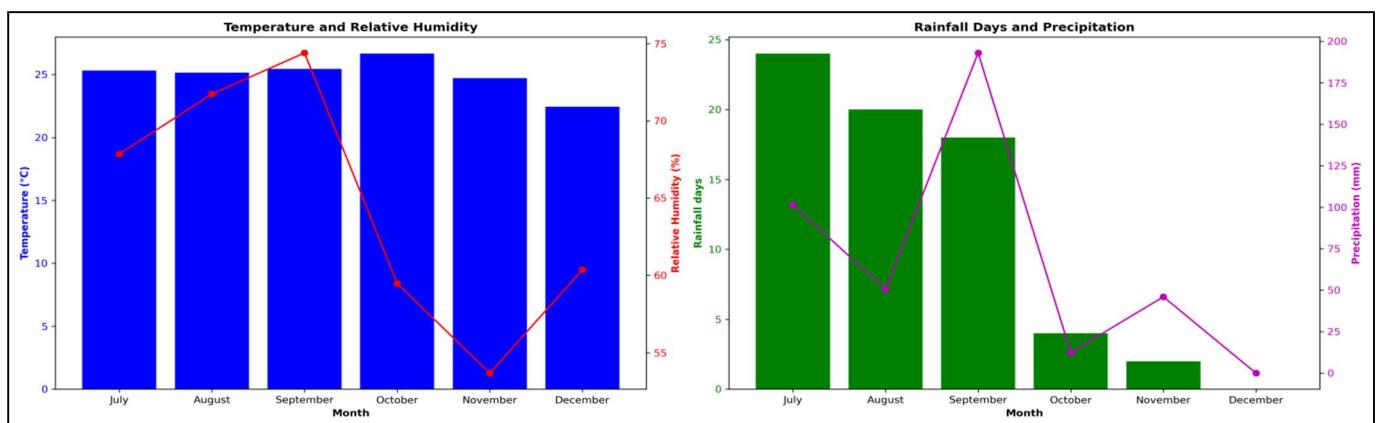


Fig1. Temperature and Relative humidity, Average rainfall days and precipitation of study area during July 2023 to December 2023. Source of data IMD Pune.

Table 1 Cropping details of four different trials

Sr. No.	Date of Transplanting	Date of Harvesting	Cropping Duration
Trial 1	11-Jul-23	5-Oct-23	86
Trial 2	24-Jul-23	25-Oct-23	93
Trial 3	8-Aug-23	2-Dec-23	116
Trial 4	25-Aug-23	17-Dec-23	114

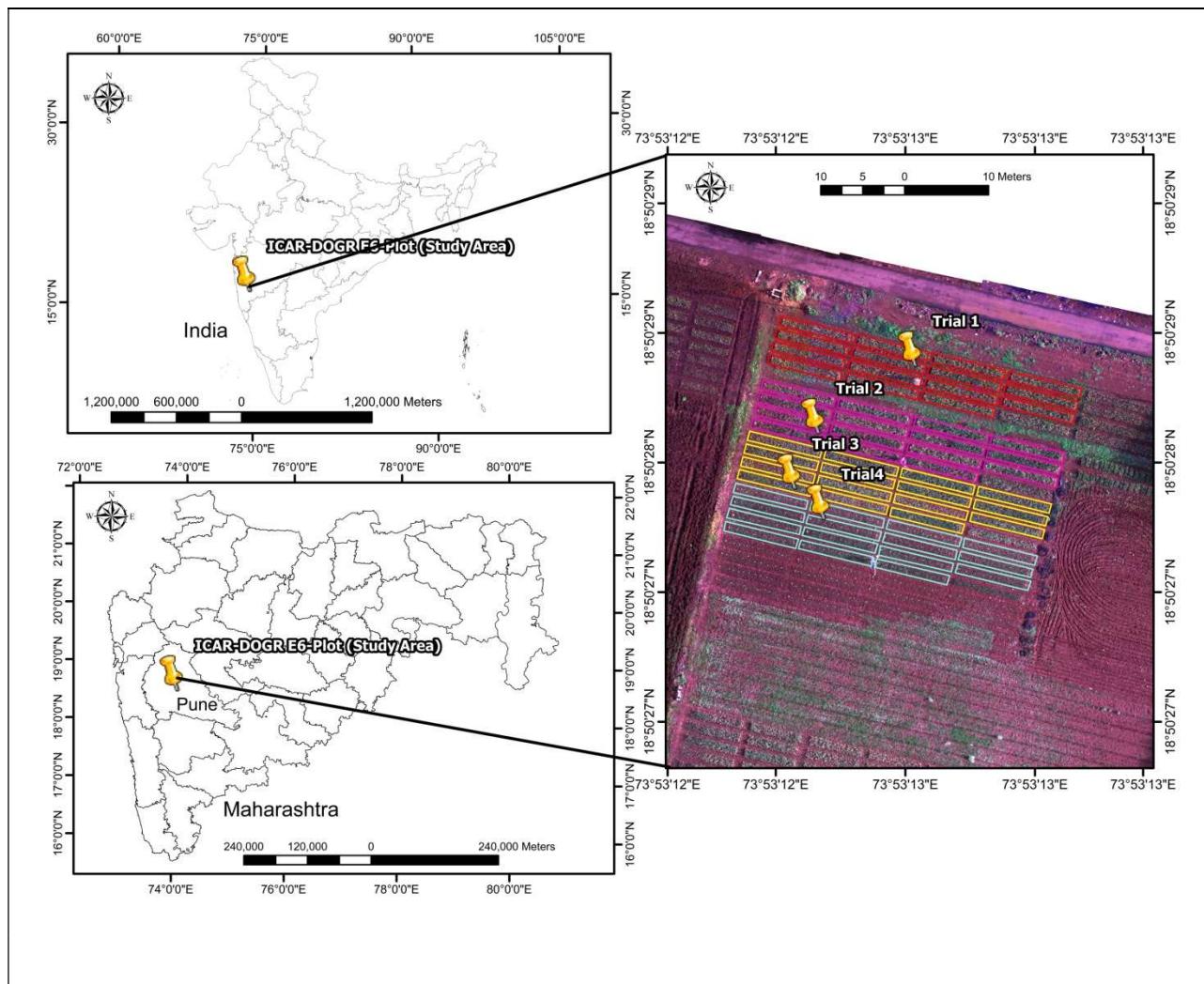


Fig 2. Study area map

2.2. Data collection and processing:

2.2.1. Acquisition of UAV Images: The Unmanned Aerial Vehicle equipped with a Mica Sense Red Edge-p multispectral camera was used to acquire five multispectral bands and one panchromatic onion field image at various growth stages. Mica Sense Red Edge-p camera is

equipped with a 6.3 mm diagonal sensor size for multispectral and 11.1 mm for panchromatic, with a 3.45-pixel size, 1450 x 1088 (1.58 MP) resolution per multispectral bands and 2464 x 2056 (5.1MP) resolution panchromatic band (Table 2.), 4:3 aspect ratio for multispectral and 6:5 for panchromatic. Based on the information provided by the manufacturer focal length is 5.5 mm for a multispectral band and 10.3 mm for a panchromatic band, with a field view is 49.6^0 HFOV x 38.3^0 VFOV (multispectral band) and 44.5^0 HFOV x 38^0 VFOV (panchromatic) and capturing 2-3 images per second, storing the file in 16-bit tiff row format. Each sample data was geo-referenced in the WGS 1984 datum, the UTM zone is N43. Radiometric correction images of standard reflectors were taken on the ground before each flight. All flights were conducted in clear sky between 10 am to 3 pm. Table 3 depicts the details of multispectral data acquisition.

Table2. Details of bands and bandwidth used in data collection.

Band Number	Band name	Center	Bandwidth
1	Blue	475 nm	32 nm
2	Green	560 nm	24 nm
3	Red	668 nm	16 nm
4	Red edge	717 nm	12 nm
5	Near-infrared	842 nm	463 nm
6	Panchromatic	634.5 nm	463 nm

2.2.2 Image processing:

Initially, we selected a set of images that were imported in Pix4D mapper software, to process the images according to the automated processing template clear sky and radiometric correction in Ag-Multispectral and make a reflectance orthomosaic model. ArcGIS10.3 is used for mosaic and composite images. Extracted onion field plot and pre-processing of six bands multispectral image was carried out to remove the soil pixels and edge effect of orthomosaic. The soil pixel was filtered out using the filtered mask with hue index which classified the pixel into soil Vs plants using a threshold value (Matias et al., 2020).

Vegetation indices calculations: seven kinds of vegetation indices were calculated from processed multispectral images.

$$\text{NDVI} = (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red}) \quad (1)$$

$$\text{NDRE} = (\text{NIR} - \text{Red Edge}) / (\text{NIR} + \text{Red Edge}) \quad (2)$$

$$\text{SAVI} = (1.5 * \text{NIR} - \text{Red}) / (0.5 + (\text{NIR} - \text{Red})) \quad (3)$$

$$\text{SR} = \text{NIR}/\text{Red Edge} \quad (4)$$

$$\text{GNDVI} = (\text{NIR} - \text{Green}) / (\text{NIR} + \text{Green}) \quad (5)$$

$$\text{NORM2} = (\text{Red} - \text{Green}) / (\text{Red} + \text{Green}) \quad (6)$$

$$\text{LAI} = (3.618 * 2.5 * (\text{NIR} - \text{R})) / (\text{NIR} + 6 * \text{Red} - 7.5 * \text{Blue} + 1) - 0.118 \quad (7)$$

Table 3. Details of multispectral data acquisition dates and intervals.

Date of planting	Trial	Image acquisition date	Days after transplanting	Growth phases of onion
11-Jul-23	Trial 1	28-Aug-23	48	Bulb initiation phase
		7-Sep-23	58	Bulb development phase
		14-Sep-23	65	
		25-Sep-23	76	
24-Jul-23	Trial 2	28-Aug-23	35	Vegetation phase
		7-Sep-23	45	Bulb initiation phase
		14-Sep-23	52	
		25-Sep-23	63	Bulb development phase
8-Aug-23	Trial 3	28-Aug-23	20	Establishment phase
		7-Sep-23	30	Vegetation phase
		14-Sep-23	37	
		25-Sep-23	48	Bulb initiation phase
		20-Oct-23	73	Bulb development phase
		30-Oct-23	83	
25-Aug-23	Trial 4	7-Sep-23	13	Establishment phase
		14-Sep-23	20	Vegetation phase
		25-Sep-23	31	
		20-Oct-23	56	Bulb initiation phase
		30-Oct-23	66	Bulb development phase
		10-Nov-23	77	
		16-Nov-23	83	

2.3 Statistical analysis:

Quantitative remote sensing is a powerful tool that can be used to identify and classify different crops, assess crop conditions and estimate crop yields. To find out the strength of the linear relationship between each spectral index and yield, we selected statistical methods such as Pearson's correlation analysis, principle component analysis, linear regression and machine learning approach on scaled data. Each spectral index was further analyzed to find its relationship with the yield of each plot. In this study, linear and machine learning models were also used to predict onion yield. To improve model accuracy, we selected several modeling methods including simple linear regression, decision tree, random forest, support vector regression and Xgboost model for each trial.

For onion yield estimation we used a simple linear regression model of vegetation indices and tested the estimation accuracy of each vegetation indices and trials. This was calculated as follows.

$$Y(\text{yield}) = \beta_1 x + \beta_0 + \epsilon \quad (8)$$

Where: Y is the Actual yield and x is spectral indices like NDVI, NDRE, SAVI, GNDVI, SR, LAI, and NORM2. β_1 is the intercept of the independent variable, and β_0 is constant.

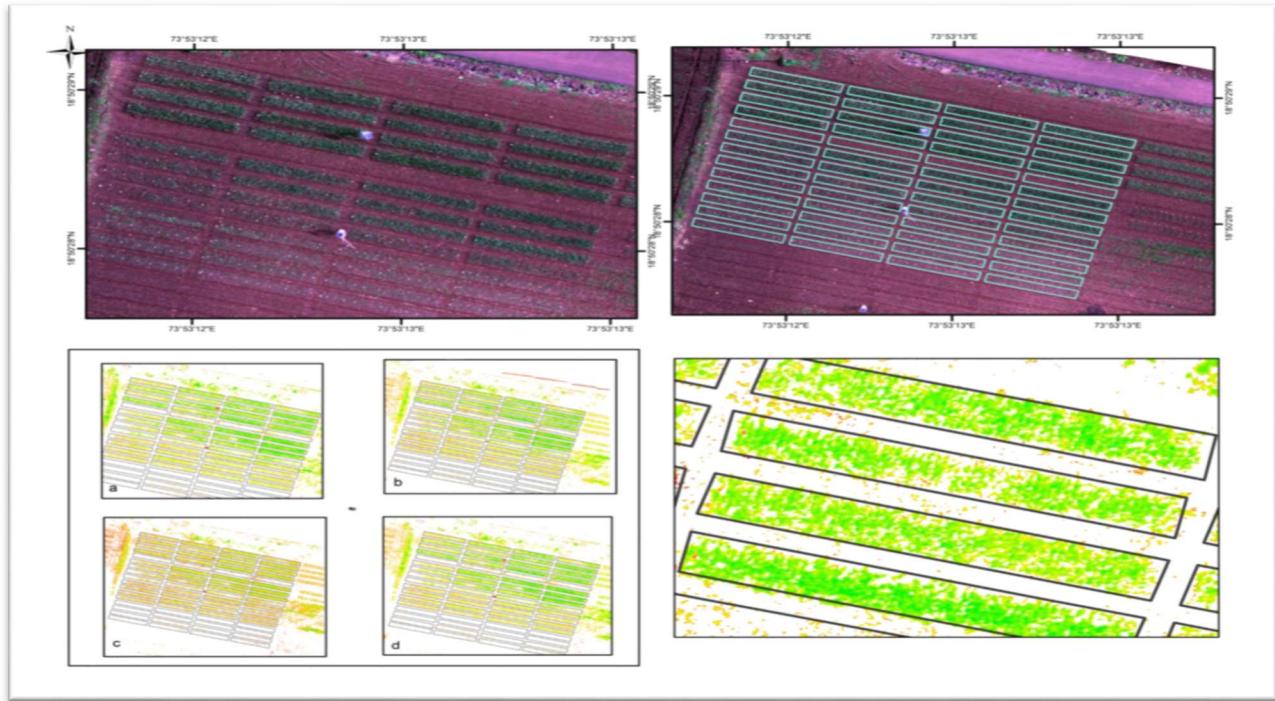


Fig3. Orthomosaic map, shape file, vegetation map, extract zonal statistics vegetation map.

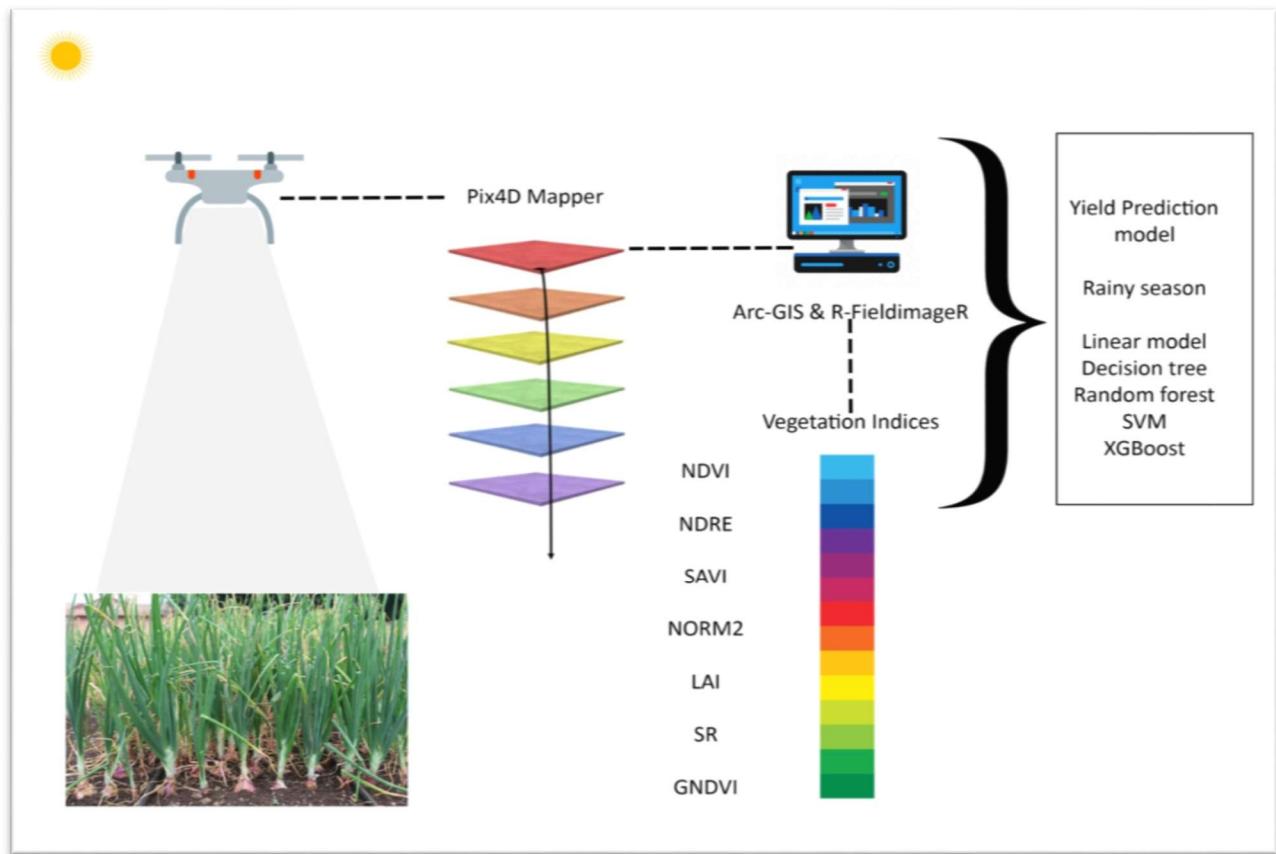


Fig.4 Overview of the process used in this study.

Machine learning modeling.

Principle Component Analysis (PCA) is a data mining multivariate statistical technique used to reduce the dimensionality of a dataset while preserving as much variability as possible. It does this by transforming original variables into a new set of uncorrelated variables called principle components (PC's) (Paul *et al.*, 2013). In this study, we utilized PCA to reduce the dimensionality of data by transforming the original matrix of vegetation indices into 3 principal components. The PCA transformation can be mathematically expressed as follows:

Given matrix $X \in \mathbb{R}^{n \times 7}$ of vegetation indices transformed data, PCA transformed X into $Z \in \mathbb{R}^{n \times 3}$ the transformed is defined as:

$$Z = XW$$

Where:

$X \in \mathbb{R}^{n \times 7}$ is the matrix of vegetation indices n is no of sample and p=7 vegetation indices.
 $Z \in \mathbb{R}^{n \times 3}$ is the matrix containing the first p=3 eigenvectors of the covariance matrix of X-scaled

data. $W \in \mathbb{R}^{7 \times 3}$ is the matrix containing the first 3 eigenvectors of the covariance matrix of X-scaled data.

The principle components were used as independent variables in various machine learning models, including linear regression, Decision Tree, Random forest, support vector regression and Xgboost to predict the rainy season onion yield based on transformed principle components(Nayana et al., 2022; Cheng et al., 2022). To evaluate the performance of the machine learning models, initially, data was split into 80% training set the model and 20% was reserved for testing subsets. Train-Test split data ensures that the model is validated on unseen data and provides a realistic assessment of its generalization capability (Sing et al., 2021).

Decision Tree Regressor is a non-parametric supervised learning algorithm for predicting continuous values. It splits data into subsets based on the feature values, creating tree model-like decisions. Decision tree is a promising approach for regression problems to estimate continuous class proportions within each pixel, facilitating soft classification from remote sensing data (Kushwah et al., 2022; Xu et al., 2005). In this study, we utilized a continuous variable decision tree Regressor to predict onion yield using the principle component Z as independent variables. The decision tree Regressor was configured with the following parameters `max_depth=4`, `min_sample_leaf=3`, `random_state=1`. The mathematical formula of the decision tree Regressor is as follows:

$Z \in \mathbb{R}^{n \times 3}$ is the matrix of the first 3 principal components (explanatory variable),
 $y \in \mathbb{R}^n$ is the vector of onion yield values (dependent variable) here $y > 0$.

The DT Regressor makes predictions by learning a series of decision rules inferred from the principal components. The tree is built through recursive process:

At each node t, the dataset (Z , y) is split into two subsets, left (Z_{left} , y_{left}) and right (Z_{right} , y_{right}), based on condition on features $Z_j \leq \theta$.

Impurity criteria:

Mean Squared Error:

$I(t) = \frac{1}{Z_t} \sum_{i \in Z_t} (y_i - \bar{y}_t)$ the predicted value of terminal nodes is set to the mean value \bar{y}_t of the node. The best split is chosen to minimize the impurity $I(t)$.

Impurity reduction:

The impurity reduction ΔI at node t is calculated as:

$$\Delta I = I(t) - \left(\frac{|Z_{left}|}{|Z_t|} I(left) + \frac{|Z_{right}|}{|Z_t|} I(right) \right)$$

Stopping Criteria:

The recursion stops when one of the following condition is met the maximum depth = 4 is reached and a node contains fewer than minimum sample leaf =3 samples.

Prediction: for a new sample Z_i , the prediction \hat{Y}_i is the average yield of the samples in the leaf node I where Z_i falls.

$$\widehat{Y}_l = \frac{1}{|Z_I|} \sum_{j \in Z_I} y_j$$

Random Forest Regressor is a learning model that integrates decision trees, efficiently processing large-scale information with the ability to obtain more accurate and stable prediction and also reduce the noise (Breiman,2001). In this study, we used RF Regressor to predict onion yield using principal component Z as an independent variable. We select a parameter for random forest Regressor `n_estimator=10, max_depth=3, min_sample_leaf=2, random_state=42`. The mathematical formula of RF Regressor is as follows:

$Z \in \mathbb{R}^{n \times 3}$ is the matrix of the first 3 principal components (explanatory variable),

$y \in \mathbb{R}^n$ is the vector of onion yield values (dependent variable).

The random forest Regressor consist of an ensemble of T decision trees, where $T = n_estimator = 10$. Each DT t in the forest is trained on bootstrap sample of the data and make prediction $h_t(Z)$.

The final prediction \widehat{y} is obtained by averaging the prediction of all individual trees.

Mathematically,

$$\widehat{Y}_i = \frac{1}{T} \sum_{t=1}^T h_t(Z_i)$$

Where: \widehat{Y}_i is the predicted yield for the i^{th} sample, T is the number of trees (10) and $h_t(Z_i)$ is the prediction of i^{th} sample.

Support vector Regressor is the tolerant model that creates an ‘internal band’ with a spacing of epsilon on both sides of a linear function and which does not calculate the loss for all samples falling into the internal band. The model is obtained by minimizing total loss and maximizing interval. We select ‘rbf’ Gaussian kernel function, the parameter epsilon is 0.1 and regularization parameter C is 10 and gamma = ‘scale’. The mathematical formula for SVR model to predict onion yield.

$Z \in \mathbb{R}^{n \times 3}$ is the matrix of the first 3 principal component(explanatory variable),

$y \in \mathbb{R}^n$ is the vector of onion yield values (dependent variable).

The SVR function can be written as:

$$f(Z) = \omega \cdot \phi(Z) + b$$

Where: $\omega \in \mathbb{R}^p$ is the weight vector, $\phi(Z)$ is the mapping function that transforms the input space Z into higher dimension feature space and b is the bias term.

$$\min_{\omega, b, \xi, \xi^*} \quad \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

Where ω, b the model parameter, n is the number of sample points and $\xi_i, \xi_i^* > 0$ is slack variables.

Evaluation metrics

Multispectral and morphological data obtained from the experiment were used to establish the rainy season onion yield estimation model. Widely used parameter we selected three indices coefficient of determination R^2 , root mean error square (RMSE) and mean absolute error (MAE) to evaluate the performance of the model. The coefficient of determination measures the explanation of variation by regression analysis of model MAE and RMSE measure the average absolute of error and the average magnitude of errors between predicted and actual value respectively. \bar{Y}

$$R^2 = 1 - \frac{\sum_1^N (Y_i - \hat{Y}_i)^2}{\sum_1^N (Y_i - \bar{Y})^2}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_1^N (Y_i - \hat{Y}_i)^2}$$

$$MAE = \frac{\sum_1^N |Y_i - \hat{Y}_i|}{N}$$

Where Y_i and \hat{Y}_i is the actual and predicted yield of onion and \bar{Y} is a mean value of yield N is no of sample size.

In this study, we utilized a combination of software and statistical tools for image processing, data analysis and the fit predictive model (pix4Dmapper, Arc GIS, R, python-Jupiter notebook) as shown in the flow chart.

3. Results:

At each experimental trial-wise distribution plot vegetation indices at various growth stages help to understand the variability and patterns of these indices across different conditions. Histogram of each VI's at specific stages of onion growth such as establishment period, vegetation period, bulb initiation period and bulb development period as shown in fig 5.

3.1 Correlation between onion yield and spectral indices

A significant bulb yield variation was observed among the experimental plots transplanted on different dates in *Kharif* 2023. The correlation of onion bulb yield and spectral indices (NDVI, NDRE, NORM2, SR, GNDVI, SAVI and LAI) were studied in four different experiments with different planting dates at different time points. A strong correlation of yield with these indices was seen at the bulb development stage (63, 65, 73 and 66 DAT) in all four experiments. Among these indices, NDRE and SR showed maximal correlation coefficient (R) of 0.89, 0.86 and 0.95 in the first three dates of planting while in the last planting, SAVI, LAI and GNDVI was having higher R of 0.83 as shown in fig 6.

Linear regression analysis was done with each spectral index at a time point where a higher correlation was found with yield at the bulb development period. The linear regression model showed that the SR (R^2 – 0.86, 0.75, 0.92 & RMSE – 1.45, 1.55, 1.29) NDRE (R^2 – 0.84, 0.74, 0.92 & RMSE – 1.56, 1.56, 1.32) had a higher R^2 value and low RMSE value for the first three trials and in the fourth SAVI (R^2 – 0.72 & RMSE - 1.98), LAI (R^2 – 0.71 & RMSE – 2.02) and GNDVI (R^2 – 0.68 & RMSE – 2.02) had highest R^2 and low RMSE respectively. (Table 5). Onion bulb yield is affected by multiple factors; we also constructed multifactor models to increase the accuracy and stability of the model.

Table 5. model summary of simple linear regression models.

Regression model	65 DAT-Trial1			63 DAT-Trial2			73 DAT-Trial3			66 DAT-Trial4		
	RMS E	MA E	R^2									
NDVI	1.86	1.46	0.78	2.04	1.71	0.56	2.91	1.47	0.82	2.21	1.67	0.65
NDRE	1.56	1.24	0.84	1.56	1.28	0.74	1.32	0.98	0.92	2.66	2.32	0.5
SR	1.45	1.45	0.66	1.55	1.26	0.55	1.29	0.97	0.92	2.66	2.32	0.5
GNDVI	1.93	1.93	0.66	1.91	1.55	0.61	1.78	1.35	0.86	2.11	1.65	0.8
NORM2	1.89	1.44	0.79	1.97	1.63	0.59	1.85	1.34	0.84	2.14	1.65	0.7
SAVI	2.12	1.67	0.71	2.26	1.92	0.46	2.15	1.72	0.79	1.98	1.62	0.72
LAI	1.65	1.65	0.72	2.26	1.93	0.46	2.14	1.69	0.79	2.02	1.66	0.71

3.2 Principle component analysis.

In this study, PCA was utilized to reduce the dimensionality of data, in the analysis of spectral indices across four trials. For each trial, seven VI's (NDVI, NDRE, SR, GNDVI, NORM2, SAVI, LAI) were standardized and subjected to PCA, resulting in the variance contributions of the first PC1, second PC2 and third PC3 principal components were presented on the X-, Y- and Z-axes as shown in fig 8. The loading matrices from these PCA transformations revealed the contributions of each original variable to the principal components. In the overall trial, PC1

consistently captured the major variance with substantial loading from all VI's, indicating their strong collective influence. PC1 had negative loading in Trial1, Trial2, Trial3 and Trial4 showed positive loadings, highlighting a shift in the data structure. PC2 and PC3 revealed a more nuanced relationship, with SAVI and LAI frequently showing positive contributions and SR and NDRE often contributing differently across components. Where certain indices like NDVI, GNDVI and NORM2 are mainly influenced by principle components, while SR and NDRE exhibit more trial-specific variations in their contributions. This PCA provides valuable insight into dimension reduction and helps to identify the most significant VI's leading to a more efficient and accurate ML model for rainy season onion yield predictions.

Table 6. Validation of onion bulb yield estimation models with ML methods LR, DT, RF, SVR and XGBoost

ML Model for Onion Yield prediction	Bulb Development Period											
	65 DAT-Trial1			63 DAT-Trial2			73 DAT-Trial3			66 DAT-Trial4		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
Linear Regression	0.67 ^a	0.85 ^a	0.94 ^a	1.09 ^a	0.93 ^a	0.86 ^a	1.06 ^a	0.84 ^a	0.95 ^a	1.95 ^a	1.58 ^a	0.7 ^a
	2.26 ^b	1.96 ^b	0.49 ^b	1.31 ^b	1.17 ^b	0.86 ^b	1.88 ^b	1.55 ^b	0.79 ^b	2.72 ^b	2.22 ^b	0.61 ^b
Decision Tree	1.81 ^a	1.51 ^a	0.81 ^a	1.61 ^a	1.34 ^a	0.7 ^a	1.94 ^a	1.48 ^a	0.84 ^a	2 ^a	1.31 ^a	0.68 ^a
	2.22 ^b	1.69 ^b	0.51 ^b	1.66 ^b	1.58 ^b	0.78 ^b	2.3 ^b	2.24 ^b	0.69 ^b	2.67 ^b	2.15 ^b	0.63 ^b
Random Forest	1.7 ^a	1.42 ^a	0.83 ^a	1.41 ^a	1.12 ^a	0.77 ^a	1.6 ^a	1.24 ^a	0.89 ^a	1.8 ^a	1.38 ^a	0.74 ^a
	1.93 ^b	1.52 ^b	0.63 ^b	2.12 ^b	1.92 ^b	0.65 ^b	2.19 ^b	1.76 ^b	0.72 ^b	1.98 ^b	1.95 ^b	0.8 ^b
Support vector Regression	1.42 ^a	0.97 ^a	0.88 ^a	1.05 ^a	0.5 ^a	0.87 ^a	0.62 ^a	0.41 ^a	0.98 ^a	1.81 ^a	1.07 ^a	0.74 ^a
	1.62 ^b	1.35 ^b	0.74 ^b	2.43 ^b	1.66 ^b	0.54 ^b	1.58 ^b	1.32 ^b	0.85 ^b	2.31 ^b	2.12 ^b	0.72 ^b
Xgboost Regression	1.09 ^a	0.76 ^a	0.93 ^a	0.92 ^a	0.68 ^a	0.71 ^a	2.42 ^a	0.94 ^a	0.93 ^a	0.94 ^a	0.73 ^a	0.93 ^a
	0.76 ^b	2.99 ^b	0.15 ^b	2.56 ^b	2.2 ^b	0.49 ^b	2.25 ^b	1.95 ^b	0.7 ^b	4.38 ^b	3.76 ^b	0 ^b

(Note: ^a indicates train set models result matrix and ^b denotes test set models result matrix.)

In this study, five regression algorithms (LR, DT, RF, SVR and Xgboost) were employed alongside machine learning algorithms to predict rainy season onion yield. These predictions were based on UAV-mounted multispectral sensors to find out spectral indices. As shown in Table 6 and fig 9-A Trial1 predicted result from scaled data PCA transformed as independent variable SVR (train_R² – 0.88, test_R² – 0.74 and train_RMSE -1.42 t/ha, test_RMSE – 1.62 t/ha) and RF (train_R² – 0.83, test_R² – 0.63 and train_RMSE – 1.70 t/ha, test_RMSE – 1.93 t/ha) demonstrated to the test set with highest R² value and lower RMSE and MAE values. LR

(train_R² – 0.94, test_R² – 0.49 and train_RMSE – 0.97 t/ha, test_RMSE – 2.26 t/ha) and Xgboost (train_R² – 0.93, test_R² – 0.15 and train_RMSE – 1.09 t/ha, test_RMSE – 2.91 t/ha) showed significant overfitting, with large disparity between train and test set.

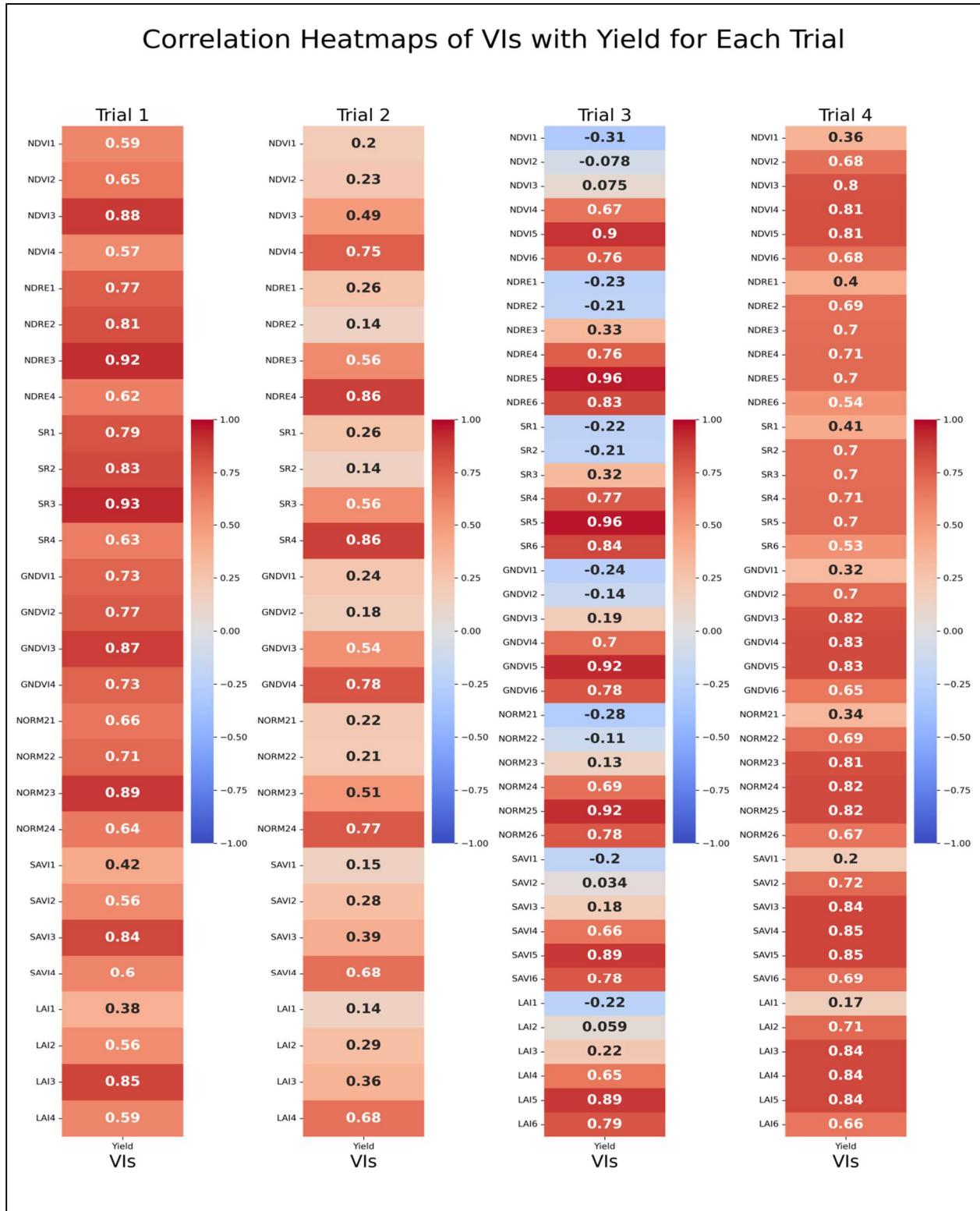
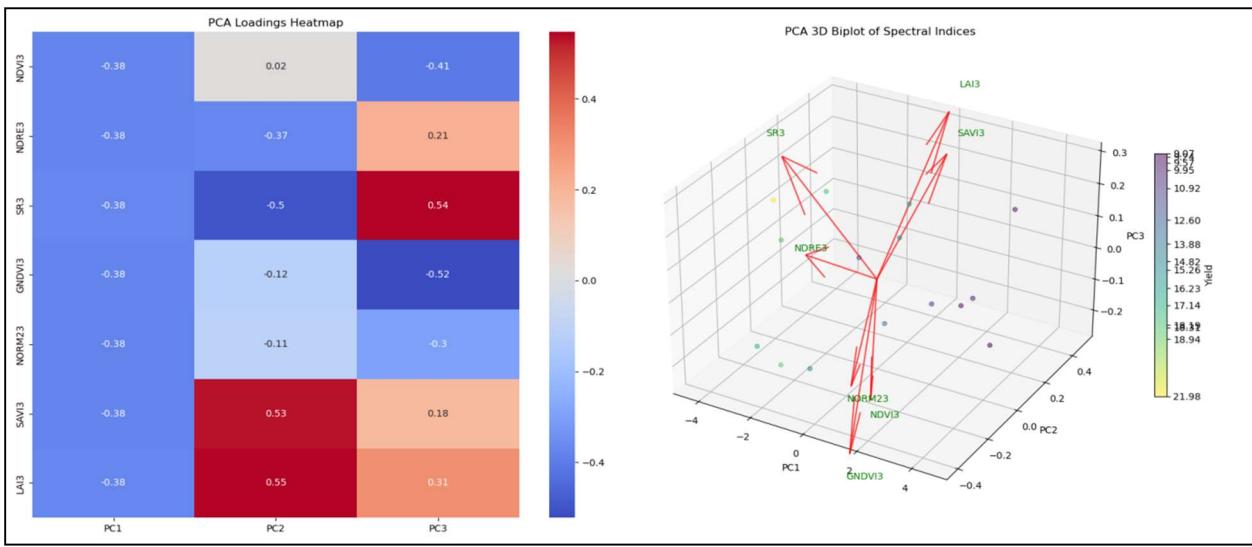
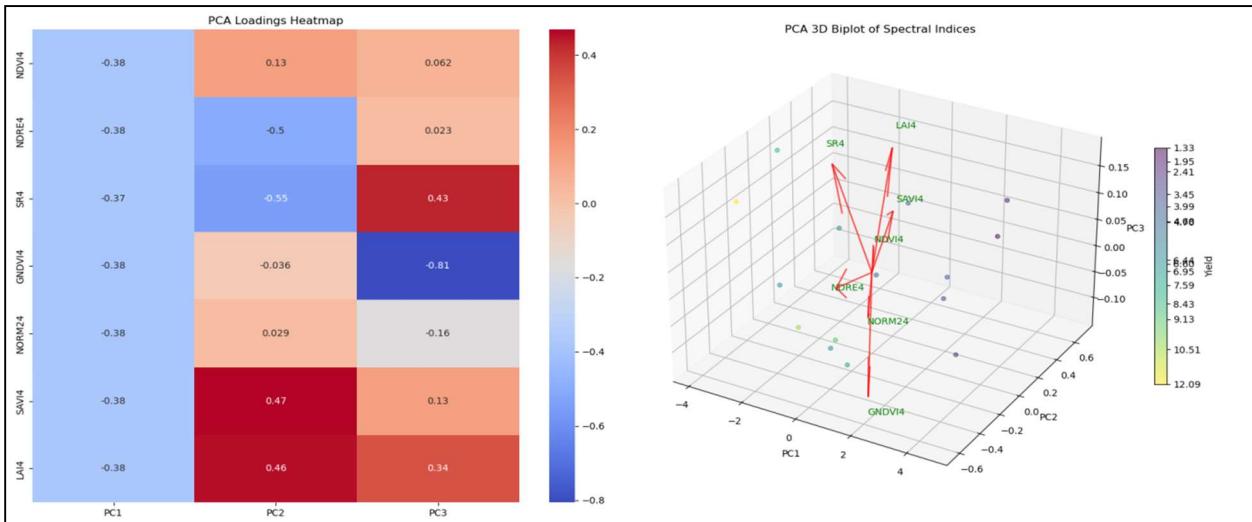


Fig6. Correlations heat-map of onion bulb yield and vegetation indices at different growth stages in onion.

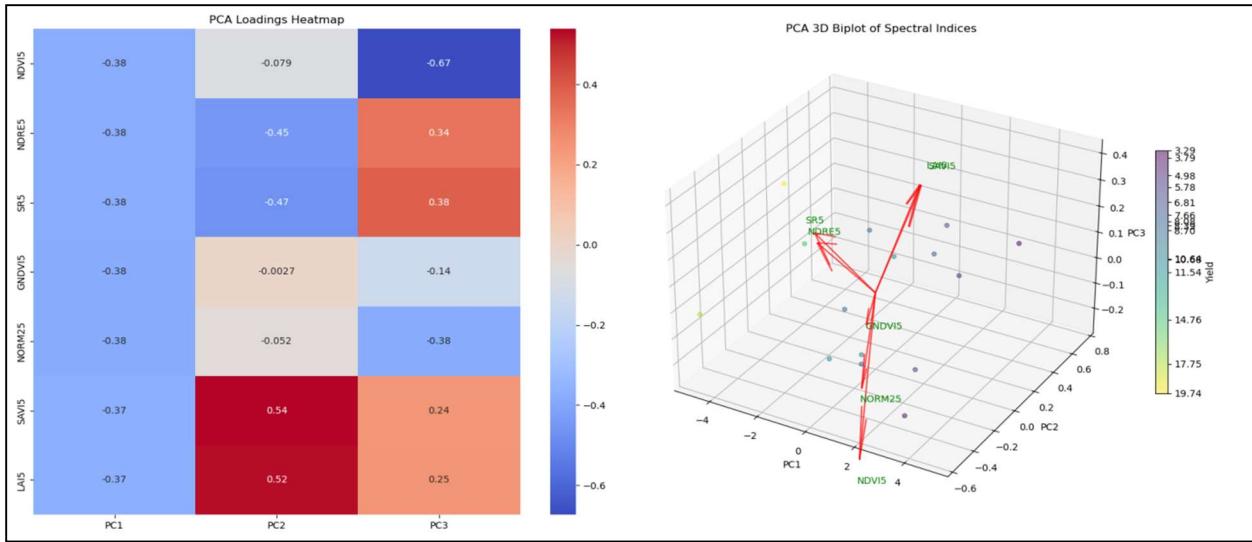
A



B



C



D

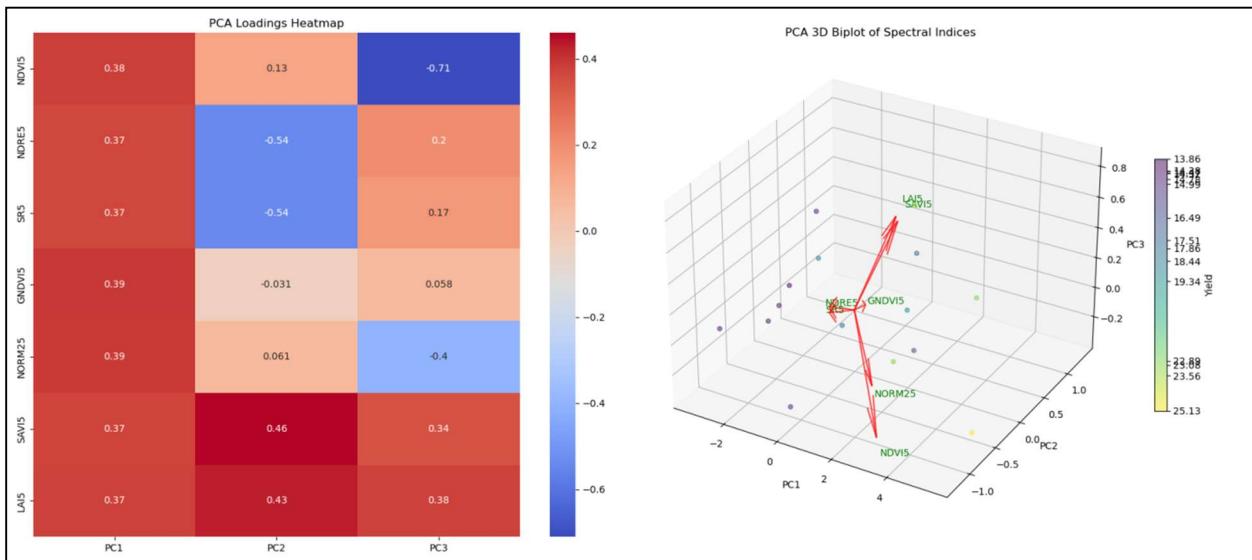
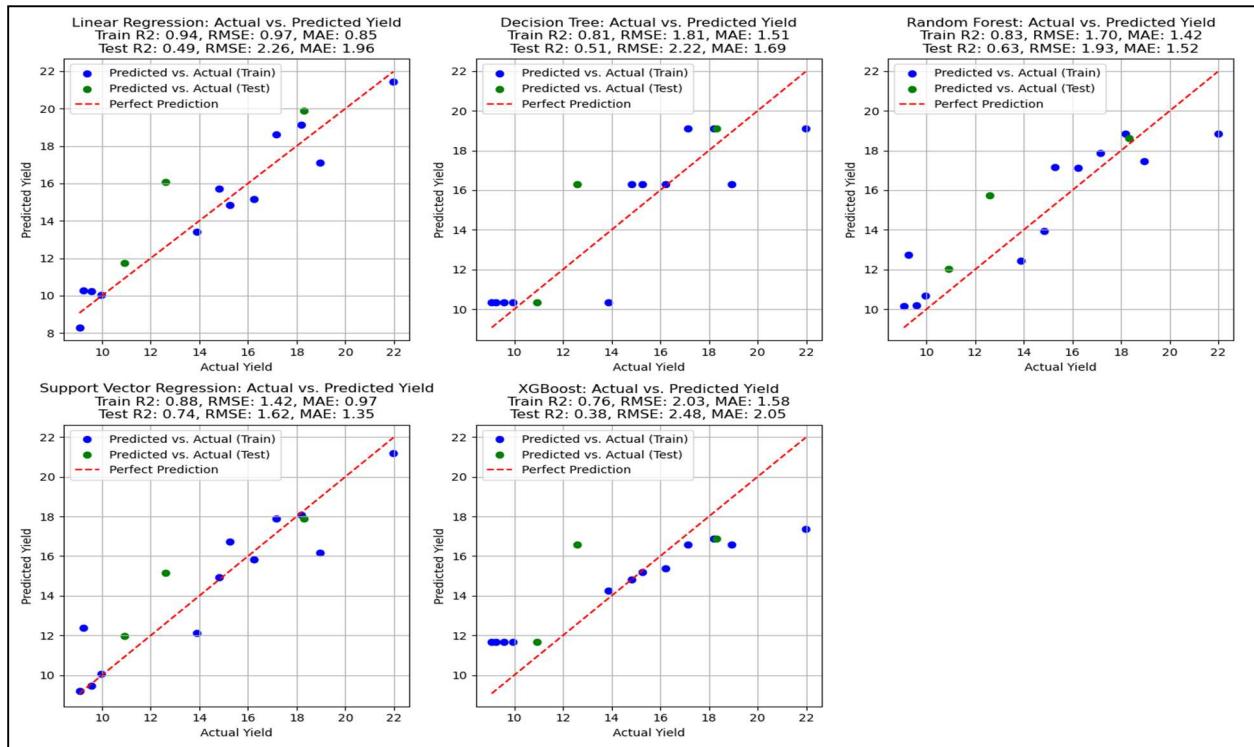


Fig.8. Principle component analysis heat-map of loading matrix biplot for different texture features
A) Trial1, B) Trial2, C) Trial3 and D) Trial4.

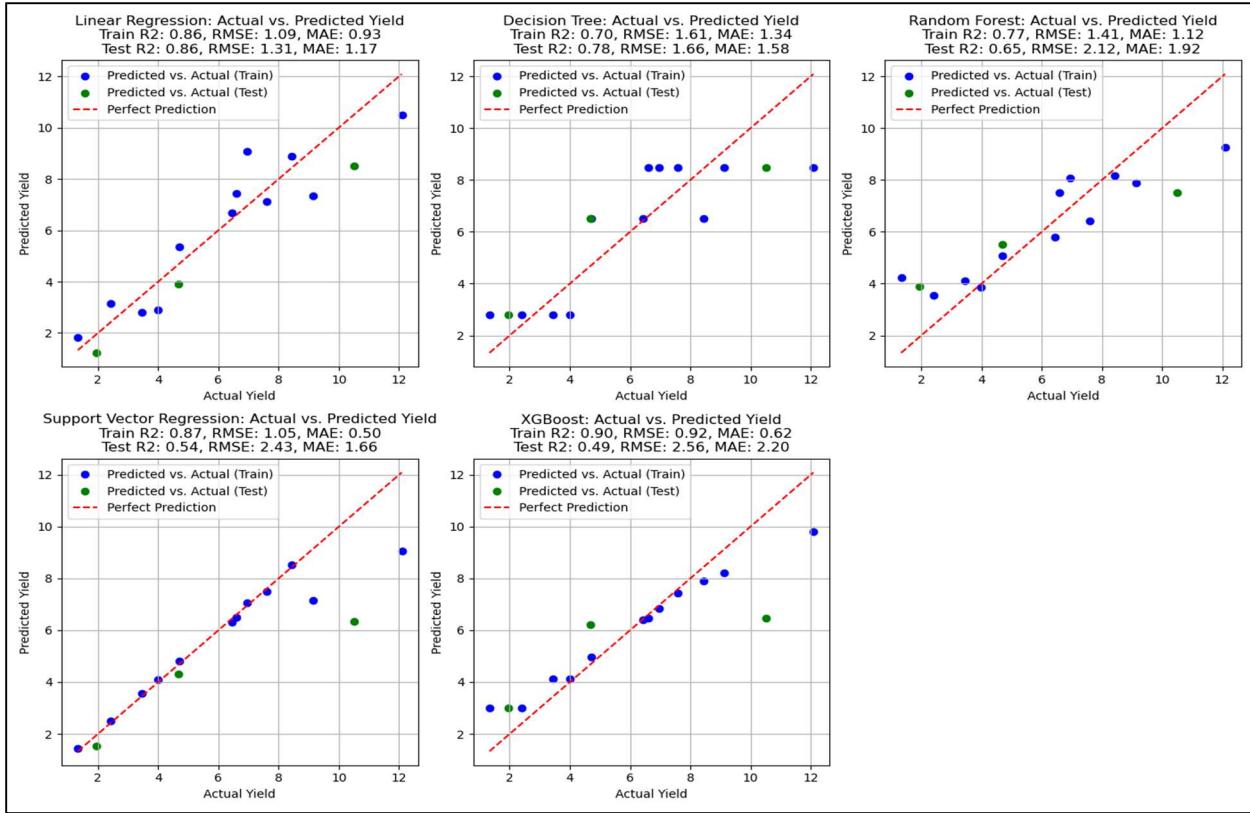
Trial2 linear regression model (train_R² – 0.86, test_R² – 0.86 and train_RMSE – 1.09 t/ha, test_RMSE – 1.31 t/ha) showed consistent performance across training and test set, indicating good model generalization. DT (train_R² – 0.70, test_R² – 0.78 and train_RMSE – 1.61, test_RMSE – 1.66) model showed good fit on the training data set but the performance was slightly better on the test set indicating a good generalization. RF (train_R² -0.77, test_R² – 0.65 and train_RMSE – 1.41, test_RMSE - 2.12) sums up a reasonable fit on the train set and test set. SVR (train_R² – 0.87, test_R²-0.54 and train_RMSE – 1.05 t/ha, train_RMSE – 2.43 t/ha) and XGBoost (train_R² – 0.90, test_R² – 0.49 and train_RMSE - 0.92, train_RMSE – 2.56) both models performed excellent on training set but poor generalization in test set. In trial3 LR (train_R² – 0.95, test_R² – 0.79 and train_RMSE – 1.06 t/ha, test_RMSE – 1.88 t/ha), SVR (train_R² – 0.98, test_R² – 0.85 and train_RMSE – 0.62 t/ha, test_RMSE – 1.58), RF (train_R² – 0.89, test_R² – 0.72 and

train_RMSE – 1.60 t/ha, test_RMSE – 2.19), XGBoost (train_R² – 0.93, test_R² – 0.70 and train_RMSE – 1.29 t/ha, test_RMSE – 2.25 t/ha) and DT (train_R² – 0.84, test_R² – 0.69 and train_RMSE – 1.94 t/ha, train_RMSE – 2.30 t/ha) all five models performed excellent on train sets with good generalization in test set. In trial4 SVR (train_R² – 0.74, test_R² – 0.72 and train_RMSE – 1.81 t/ha, train_RMSE – 2.31 t/ha), RF (train_R² – 0.74, test_R² – 0.80 and train_RMSE – 1.80 t/ha, train_RMSE – 1.98 t/ha), LR (train_R² – 0.70, test_R² – 0.61 and train_RMSE – 1.94 t/ha, train_RMSE – 2.72 t/ha) and DT (train_R² – 0.68, test_R² – 0.63 and train_RMSE – 2 t/ha, train_RMSE – 2.67 t/ha) shown better performance on both sets, indicating good generalization except XGBoost (train_R² – 0.93, test_R² – 0 and train_RMSE – 0.94 t/ha, test_RMSE – 4.38 t/ha) which showed strong training performance but very poor generalization indicating significant over fitting.

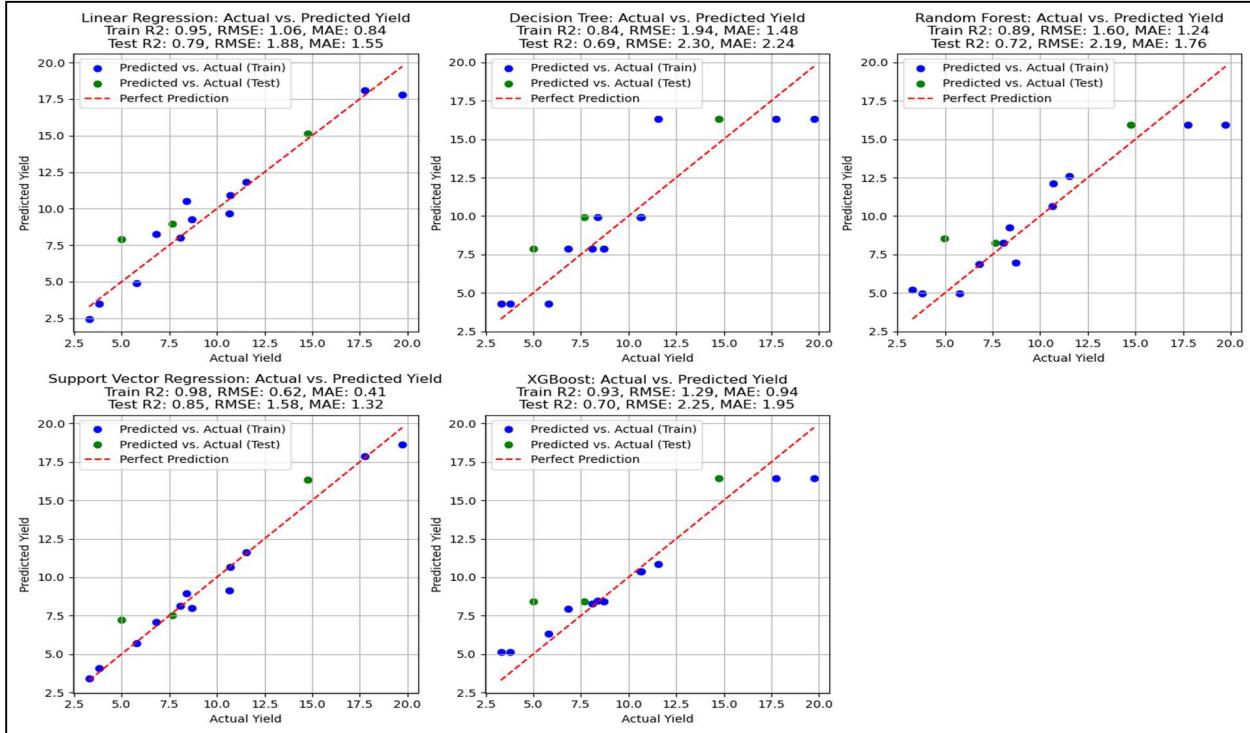
A



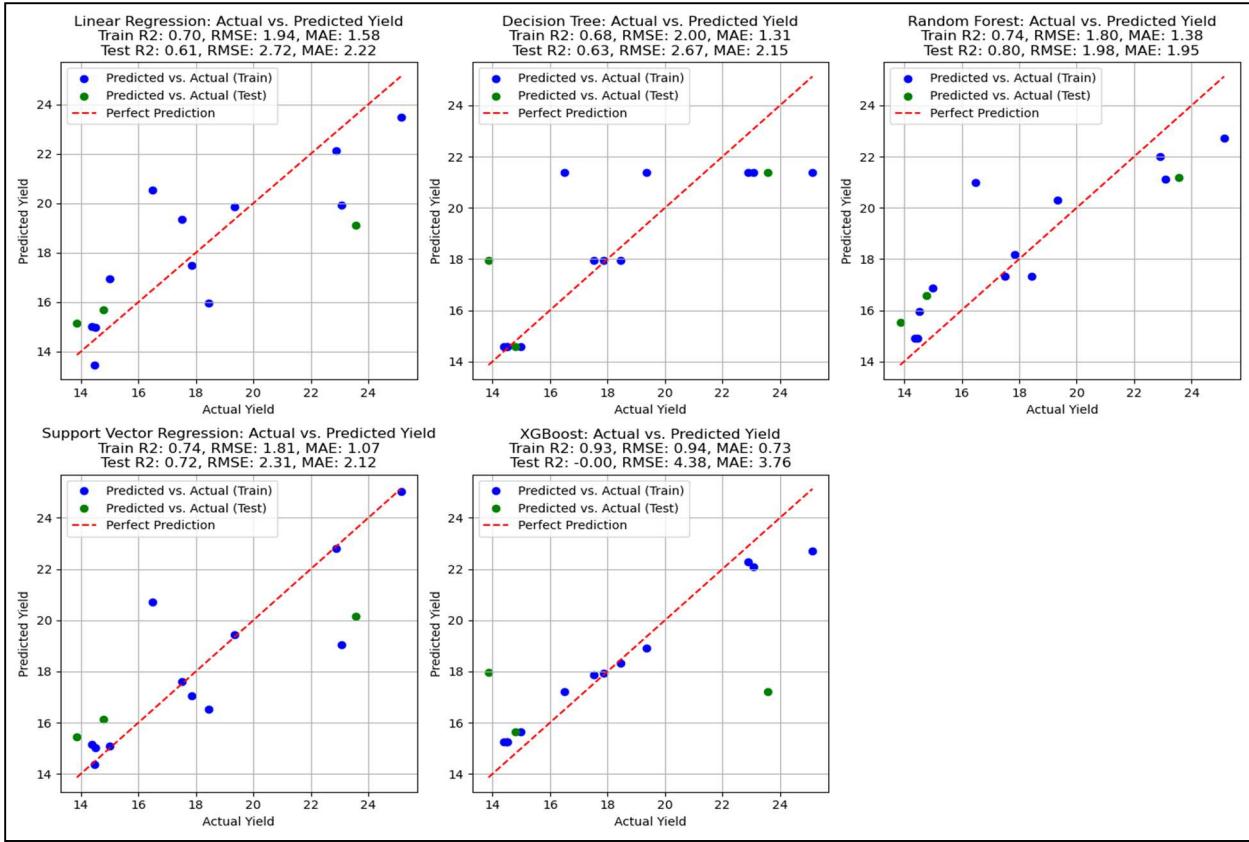
B



C



D



4. Discussion:

Crop monitoring in real-time and prediction of crop yield is crucial for agriculture planning and management. Due to inherent advantages, the use of UAV-based remote sensing platforms for the acquisition of real-time high-resolution spatiotemporal images and vegetation indices for growth modeling (Zhou et al., 2023; Tao et al., 2020), yield prediction (Zhou et al., 2017; Camenzind and Yu 2024) and plant health monitoring (Gu et al., 2023) is an area of active research. Many studies indicate that multispectral images can be good data resources for the prediction of agricultural crop yield and the estimation of spectral characteristics of plants. Multispectral and hyperspectral sensors (visible, near-infrared [nir], mid-infrared) have been used in the agriculture field (Zarco-Tejada et al., 2016). UAV-based Remote sensing technology has great potential in precision agriculture and increases the efficiency of the farmers (Velusamy et al., 2021). The literature search revealed that there are very few attempts made to study the onion crop using UAV-based multispectral imagery for growth monitoring (Jeong et al., 2017; Din et al., 2021), biomass monitoring (Ballesteros et al., 2018), yield prediction (Kang et al.,

2020) and the growth pattern identification (Duarte-Correa et al., 2023). The majority of these studies are based on a limited dataset in terms of indices used, growth stages, and number of plots and absence of morphological data. The present study investigated the linear relationship between onion bulb yield with spectral and color vegetation indices at different growth stages of rainy season crop planted at four different dates.

The onion bulb formation, development and nutrient transportation of allium crops are closely related to photosynthesis dynamics (Bachie et al., 2019). In this study, we assessed the predictive capability of different vegetation indices for the estimation of onion yield across *kharif* (rainy season) onion among four planting dates (11-Jul, 24-Jul, 8-Aug, 25-Aug). Regarding the importance of the explanatory variable, it was found that spectral indices most contributed to yield predictions. Considering the single factor to construct the yield estimation model we only adopted a simple linear regression model, we selected the explanatory variable at a time point where a higher correlation was found with yield (Marino and Alvino, 2015). For the experimental trials, we selected the bulb development stage. These selections were found to be significantly correlated with yield. We conclude that it is feasible to estimate onion yield. In a simple linear model, we found that SR and NDRE are the best predictors of onion yield in the first three trials while SAVI, LAI and GNDVI are found to be the best predictors of onion yield in the fourth trial. SR and NDRE's superior performance in predicting yield can be attributed to their ability to capture subtle changes in leaf chlorophyll content and overall plant health (Zhang et al.,2019; Cheng et al., 2022). SAVI that reduced the impact of soil have a high weight in predicting crop yield. When SAVI transformed it showed the highest performance in predicting onion yield (Marino and Alvino, 2015). In the *Kharif* seasonal window, our simple linear regression analysis model suggests that for early planting of onions, the SR and NDRE spectral vegetation index most effective predictors of onion yield. Additionally, our findings indicate that for the fourth trial within this season, the SAVI, GNDVI and LAI index has superior predictive capability.

While considering multiple factors used to develop the yield estimation models, rice yield forecasting over Vietnam improved by an average level of ML-only 18.5% up to 45 % to PCA-ML combination outperforming 20% to 60% (Pham et al., 2022). The benefit of the application of PCA was the overall reduction between 12 and 30.2% in the maize yield prediction (Croci et al., 2022). We selected seven spectral indices to establish the model by PCA-ML methods. Different modeling methods may marginally improve model accuracy. We also analyzed and evaluated the

onion yield estimation accuracy of LR, DT, RF, SVR and XGBoost (Table 6). Random forest regression generally performs well across all trials with good generalization, indicating significantly good fitting evidenced by small RMSE and MAE values and relatively high R^2 , indicating that the model is capable of making precise and accurate predictions while effectively explaining the variance in the dependent variable, similar results were obtained by (Ramos et al., 2023) while working on UAV-based vegetations spectral indices and maize yield prediction. The SVR model shows strong performance on the training dataset across all trials, with high R^2 values (0.74 to 0.98) indicating that the model captures the majority of the variance in the data but the test set performance generally good but exhibits some degree of over fitting in trial 2 show moderate performance with test R^2 values indicating that model still capture significant variance in test set, this results are in similitude with (Kumar et al., 2023). This underscores the model's reliability and robustness in those specific trials, making it a valuable tool for yield prediction in *Kharif* onion. Providing insights into the validation of the onion yield estimation model highlights the importance of assessing the model performance of different experimental conditions.

5. Conclusion

In the ever-evolving landscape of agricultural technology, the utilization of UAVs for predicting the yield of *Kharif* onion crops represents a paradigm shift in precision agriculture. With the ability to capture high-resolution imagery and employ advanced analytical techniques, UAVs offer a revolutionary approach to assessing crop health and estimating yields. The *Kharif* onion crop, a cornerstone of India's agricultural sector, stands to benefit significantly from this innovative application, promising farmers enhanced efficiency, productivity, and sustainability. By leveraging UAV technology, researchers and farmers alike are pioneering new methods that not only optimize resource allocation but also mitigate risks associated with market uncertainties and climatic variability. As UAV-based yield prediction continues to demonstrate promising results, it heralds a new era of data-driven decision-making in agriculture, empowering stakeholders to navigate the complexities of modern farming with precision and foresight.

Yield prediction using UAV and satellite imagery could help in identifying yield gaps during particular growth stages and seasonal fluctuations in production and supply that can be counterbalanced with a robust monitoring system. SVR and RF models performed best in yield

prediction during all four trials, similarly, comparative studies can be conducted through length and breadth across the localities to predict cumulative yield. ML tools including linear regression, Decision tree, Random forest, Support vector regression, Xgboost supported better for estimating onion yield at the bulb development stage of onion crop. RF and SVR showed good generalization, especially in the test set results. Both RF and SVR models were efficient in handling variability effectively and well generalized to predict bulb yield. This study says that multispectral imagery could be constructively used for reliable growth and yield prediction of rainy season onion crop.

6. References

- Abdulridha, J., Ampatzidis, Y., Kakarla, S.C. and Roberts, P., 2020. Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. *Precision Agriculture*, 21, pp.955-978.
- Bachie, O. G., Santiago, L. S., & McGiffen, M. E. 2019. Physiological responses of onion varieties to varying photoperiod and temperature regimes. *Agriculture*, 9(10), 214.
- Ballesteros, R.; Ortega, J.F.; Hernandez, D.; Moreno, M.A. 2018. Onion biomass monitoring using UAV-based RGB imaging. *Precis. Agric.*, 1–18.
- Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S. and Bareth, G., 2014. Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. *Remote sensing*, 6(11), pp.10395-10412.
- Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M.L. and Bareth, G., 2015. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, 39, pp.79-87.
- Breiman, L. 2001. Random Forests. *Machine Learning* **45**, 5–32
<https://doi.org/10.1023/A:1010933404324>
- Camenzind MP and Yu K. 2024. Multi temporal multispectral UAV remote sensing allows for yield assessment across European wheat varieties already before flowering. *Front. Plant Sci.* 14:1214931. doi: 10.3389/fpls.2023.1214931
- Cheng E, Zhang B, Peng D, Zhong L, Yu L, Liu Y, Xiao C, Li C, Li X, Chen Y, Ye H, Wang H, Yu R, Hu J and Yang S, 2022. Wheat yield estimation using remote sensing data based on machine learning approaches. *Front. Plant Sci.* 13:1090970. doi: 10.3389/fpls.2022.1090970

Chen, T., Yang, W., Zhang, H., Zhu, B., Zeng, R., Wang, X., Wang, S., Wang, L., Qi, H., Lan, Y. and Zhang, L., 2020. Early detection of bacterial wilt in peanut plants through leaf-level hyperspectral and unmanned aerial vehicle data. *Computers and Electronics in Agriculture*, 177, p.105708.

Córcoles, J.I., Ortega, J.F., Hernández, D. and Moreno, M.A., 2013. Estimation of leaf area index in onion (*Allium cepa* L.) using an unmanned aerial vehicle. *Biosystems engineering*, 115(1), pp.31-42.

Croci, M.; Impollonia, G.; Meroni, M.; Amaducci, S. Dynamic Maize Yield Predictions Using Machine Learning on Multi-Source Data. *Remote Sens.* **2023**, *15*, 100. <https://doi.org/10.3390/rs15010100>

Din, M., Zheng, W., Rashid, M., Wang, S. and Shi, Z., 2017. Evaluating hyperspectral vegetation indices for leaf area index estimation of *Oryza sativa* L. at diverse phenological stages. *Frontiers in plant science*, 8, p.237162.

Din, N., Naz, B., Zai, S., Bakhtawer, & Ahmed, W. 2021. Onion Crop Monitoring with Multispectral Imagery using Deep Neural Network. *International Journal of Advanced Computer Science and Applications*, 12.

Duarte-Correa, D., Rodríguez-Reséndiz, J., Díaz-Flórez, G., Olvera-Olvera, C. A., & Álvarez-Alvarado, J. M. 2023. Identifying Growth Patterns in Arid-Zone Onion Crops (*Allium Cepa*) Using Digital Image Processing. *Technologies*, 11(3), 67.

Faiçal, B.S., Costa, F.G., Pessin, G., Ueyama, J., Freitas, H., Colombo, A., Fini, P.H., Villas, L., Osório, F.S., Vargas, P.A. and Braun, T., 2014. The use of unmanned aerial vehicles and wireless sensor networks for spraying pesticides. *Journal of Systems Architecture*, 60(4), pp.393-404.

Fu, Y., Yang, G., Wang, J., Song, X. and Feng, H., 2014. Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. *Computers and Electronics in Agriculture*, 100, pp.51-59.

Fu, Z., Jiang, J., Gao, Y., Krienke, B., Wang, M., Zhong, K., Cao, Q., Tian, Y., Zhu, Y., Cao, W. and Liu, X., 2020. Wheat growth monitoring and yield estimation based on multi-rotor unmanned aerial vehicle. *Remote Sensing*, 12(3), p.508.

Gadge, S.S. and Lawande, K.E., 2012. Crop damage due to climatic change: a major constraint in onion farming. *Indian Res. J. Ext. Educ*, 2, pp.38-41.

- Geipel, J., Link, J. and Claupein, W., 2014. Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. *Remote sensing*, 6(11), pp.10335-10355.
- Gu, C., Cheng, T., Cai, N., Li, W., Zhang, G., Zhou, X.G. and Zhang, D., 2023. Assessing narrow brown leaf spot severity and fungicide efficacy in rice using low altitude UAV imaging. *Ecological Informatics*, p.102208.
- Jeong, S., Kim, D., Yun, H., Cho, W., Kwon, Y., & Kim, H. 2017. Monitoring the growth status variability in Onion (*Allium cepa*) and Garlic (*Allium sativum*) with RGB and multi-spectral UAV remote sensing imagery. In Proceedings of the 7th Asian-Australasian Conference on Precision Agriculture, Hamilton, New Zealand (pp. 16-18).
- Kang, Y. S., Jang, S. H., Park, J. W., Song, H. Y., Ryu, C. S., Jun, S. R., & Kim, S. H. 2020. Yield prediction and validation of onion (*Allium cepa* L.) using key variables in narrowband hyperspectral imagery and effective accumulated temperature. *Computers and electronics in agriculture*, 178, 105667.
- Kumar, C.; Mubvumba, P.; Huang, Y.; Dhillon, J.; Reddy, K. Multi-Stage Corn Yield Prediction Using High-Resolution UAV Multispectral Data and Machine Learning Models. *Agronomy* **2023**, *13*, 1277. <https://doi.org/10.3390/agronomy13051277>
- Kushwah, J.S., Kumar, A., Patel, S., Soni, R., Gawande, A., Gupta, S., 2022. Comparative study of regressor and classifier with decision tree using modern tools. *Materials Today: Proceedings* **56**, 3571–3576. <https://doi.org/10.1016/j.matpr.2021.11.635>
- Liang, S., 2004. Quantitative remote sensing of land surfaces. John Wiley & Sons.
- Paul, L. C., Suman, A. A., & Sultan, N. 2013. Methodological analysis of Principal Component Analysis (PCA) method. *International Journal of Computational Engineering & Management*, **16**(2), 32-38.
- Liu, Y., Sun, L., Liu, B., Wu, Y., Ma, J., Zhang, W., ... & Chen, Z. 2023. Estimation of Winter Wheat Yield using multiple temporal vegetation indices derived from UAV-Based multispectral and hyperspectral imagery. *Remote Sensing*, **15**(19), 4800.
- Maes, W. H. and Steppe, K. 2019. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends in plant science*, **24**(2), pp.152-164.

Marino, S. and Alvino, A. 2015. ‘Hyperspectral vegetation indices for predicting onion (*Allium cepa* L.) yield spatial variability’, *Computers and Electronics in Agriculture*, 116, pp. 109–117. Available at: <https://doi.org/10.1016/j.compag.2015.06.014>.

Matias, F.I., Caraza-Harter, M.V. and Endelman, J.B. 2020. ‘FIELDimageR: An R package to analyze orthomosaic images from agricultural field trials’, *The Plant Phenome Journal*, 3(1), p. e20005. Available at: <https://doi.org/10.1002/ppj2.20005>.

Narmilan, A., Gonzalez, F., Salgadoe, A.S.A. and Powell, K., 2022. Detection of white leaf disease in sugarcane using machine learning techniques over UAV multispectral images. *Drones*, 6(9), p.230.

Nayana, B. M., Kolla Rohit Kumar, and Christophe Chesneau. 2022. Wheat Yield Prediction in India Using Principal Component Analysis-Multivariate Adaptive Regression Splines (PCA-MARS), *AgriEngineering* 4, no. 2: 461-474. <https://doi.org/10.3390/agriengineering4020030>

Noureldin, N.A., Aboelghar, M.A., Saudy, H.S. and Ali, A.M., 2013. Rice yield forecasting models using satellite imagery in Egypt. *The Egyptian Journal of Remote Sensing and Space Science*, 16(1), pp.125-131.

Pham, H.T.; Awange, J.; Kuhn, M.; Nguyen, B.V.; Bui, L.K. Enhancing Crop Yield Prediction Utilizing Machine Learning on Satellite-Based Vegetation Health Indices. *Sensors* **2022**, 22, 719. <https://doi.org/10.3390/s22030719>

Ramos, A.P.M., Osco, L.P., Furuya, D.E.G., Gonçalves, W.N., Santana, D.C., Teodoro, L.P.R., da Silva Junior, C.A., Capristo-Silva, G.F., Li, J., Baio, F.H.R. and Junior, J.M., 2020. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, 178, p.105791.

Rodriguez, J., Lizarazo, I., Prieto, F. and Angulo-Morales, V., 2021. Assessment of potato late blight from UAV-based multispectral imagery. *Computers and Electronics in Agriculture*, 184, p.106061.

Singh, V., Pencina, M., Einstein, A.J. *et al.* 2021. Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging. *Sci. Rep.* **11**, 14490. <https://doi.org/10.1038/s41598-021-93651-5>

Su, J., Liu, C., Coombes, M., Hu, X., Wang, C., Xu, X., Li, Q., Guo, L. and Chen, W.H., 2018. Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. *Computers and electronics in agriculture*, 155, pp.157-166.

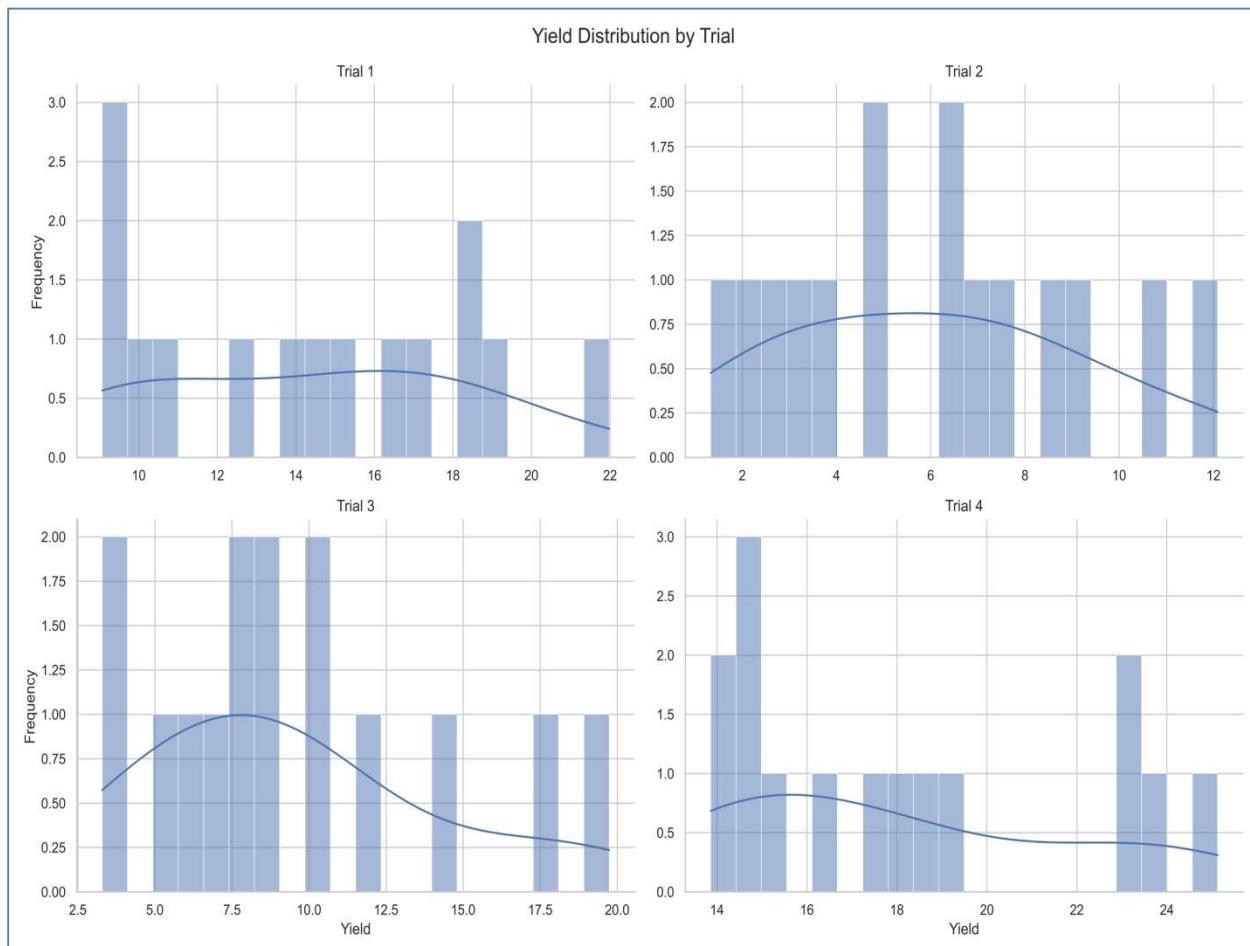
- Tao, H., Feng, H., Xu, L., Miao, M., Long, H., Yue, J., ... & Fan, L. 2020. Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data. *Sensors*, 20(5), 1296.
- Usha, K. and Singh, B., 2013. Potential applications of remote sensing in horticulture: A review. *Scientia horticulturae*, 153, pp.71-83.
- Velusamy, P., Rajendran, S., Mahendran, R.K., Naseer, S., Shafiq, M. and Choi, J.G., 2021. Unmanned Aerial Vehicles (UAV) in precision agriculture: Applications and challenges. *Energies*, 15(1), p.217.
- Verger, A., Vigneau, N., Chéron, C., Gilliot, J.M., Comar, A. and Baret, F., 2014. Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sensing of Environment*, 152, pp.654-664.
- Wang, C., Chen, Y., Xiao, Z., Zeng, X., Tang, S., Lin, F., Zhang, L., Meng, X. and Liu, S., 2023. Cotton blight identification with ground framed canopy photo-assisted multispectral UAV images. *Agronomy*, 13(5), p.1222.
- Xie, Q., Huang, W., Liang, D., Chen, P., Wu, C., Yang, G., Zhang, J., Huang, L. and Zhang, D., 2014. Leaf area index estimation using vegetation indices derived from airborne hyperspectral images in winter wheat. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8), pp.3586-3594.
- Xu, Min., Watanachaturaporn, Pakorn., Varshney, P.K. and Arora, Manoj. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*. 97. 322-336. <https://doi.org/10.1016/j.rse.2005.05.008>.
- Ye, H., Huang, W., Huang, S., Cui, B., Dong, Y., Guo, A., Ren, Y. and Jin, Y., 2020. Recognition of banana fusarium wilt based on UAV remote sensing. *Remote sensing*, 12(6), p.938.
- Zarco-Tejada, P.J., González-Dugo, M.V. and Fereres, E., 2016. Seasonal stability of chlorophyll fluorescence quantified from airborne hyperspectral imagery as an indicator of net photosynthesis in the context of precision agriculture. *Remote Sensing of Environment*, 179, pp.89-103.
- Zhang, C. and Kovacs, J.M., 2012. The application of small unmanned aerial systems for precision agriculture: a review. *Precision agriculture*, 13, pp.693-712.
- Zhang, K., Ge, X., Shen, P., Li, W., Liu, X., Cao, Q., ... & Tian, Y. 2019. Predicting rice grain yield based on dynamic changes in vegetation indexes during early to mid-growth stages. *Remote sensing*, 11(4), 387.

Zhou, X., Zheng, H. B., Xu, X. Q., He, J. Y., Ge, X. K., Yao, X., ... & Tian, Y. C. 2017. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 130, 246-255.

Zhou H, Yang J, Lou W, Sheng L, Li D and Hu H. 2023. Improving grain yield prediction through fusion of multitemporal spectral features and agronomic trait parameters derived from UAV imagery. Front. Plant Sci. 14:1217448. doi: 10.3389/fpls.2023.1217448

<https://www.fao.org/faostat/en/#data/QCL>. Accessed 20 December 2023.

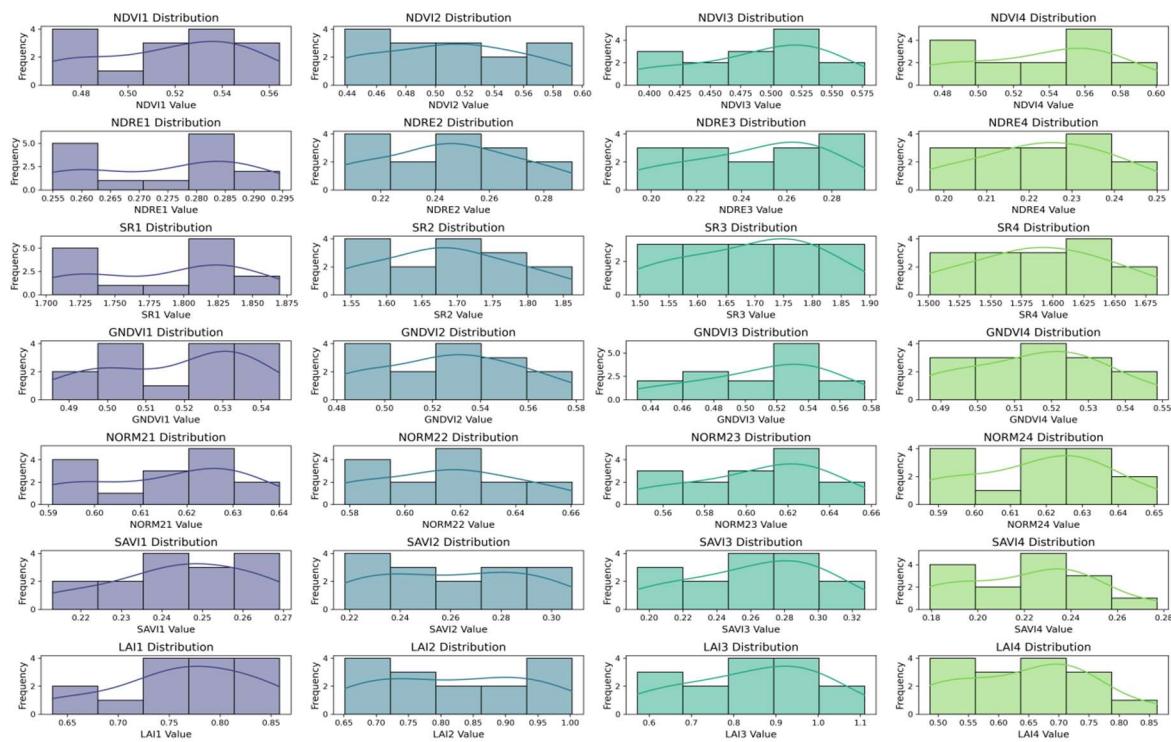
https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation



Trial	Image acquisition date	VI's	Growth phases of onion
Trial 1	28-Aug-23	NDVI1, NDRE1, SR1, GNDVI1, SAVI1, NORM1 LAI1	Bulb initiation phase
	7-Sep-23	NDVI2, NDRE2, SR2, GNDVI2, SAVI2, NORM2, LAI2	
	14-Sep-23	NDVI3, NDRE3, SR3, GNDVI3, SAVI3, NORM3, LAI3	Bulb development phase
	25-Sep-23	NDVI4, NDRE4, SR4, GNDVI4, SAVI4, NORM4, LAI4	
Trial 2	28-Aug-23	NDVI1, NDRE1, SR1, GNDVI1, SAVI1, NORM1 LAI1	Vegetation phase
	7-Sep-23	NDVI2, NDRE2, SR2, GNDVI2, SAVI2, NORM2, LAI2	
	14-Sep-23	NDVI3, NDRE3, SR3, GNDVI3, SAVI3, NORM3, LAI3	Bulb initiation phase
	25-Sep-23	NDVI4, NDRE4, SR4, GNDVI4, SAVI4, NORM4, LAI4	Bulb development phase
Trial 3	28-Aug-23	NDVI1, NDRE1, SR1, GNDVI1, SAVI1, NORM1 LAI1	Establishment phase
	7-Sep-23	NDVI2, NDRE2, SR2, GNDVI2, SAVI2, NORM2, LAI2	Vegetation phase
	14-Sep-23	NDVI3, NDRE3, SR3, GNDVI3, SAVI3, NORM3, LAI3	
	25-Sep-23	NDVI4, NDRE4, SR4, GNDVI4, SAVI4, NORM4, LAI4	Bulb initiation phase
	20-Oct-23	NDVI5, NDRE5, SR5, GNDVI5, SAVI5, NORM5, LAI5	
	30-Oct-23	NDVI6, NDRE6, SR6, GNDVI6, SAVI6, NORM6, LAI6	Bulb development phase
Trial 4	7-Sep-23	NDVI1, NDRE1, SR1, GNDVI1, SAVI1, NORM1 LAI1	Establishment phase
	14-Sep-23	NDVI2, NDRE2, SR2, GNDVI2, SAVI2, NORM2, LAI2	Vegetation phase

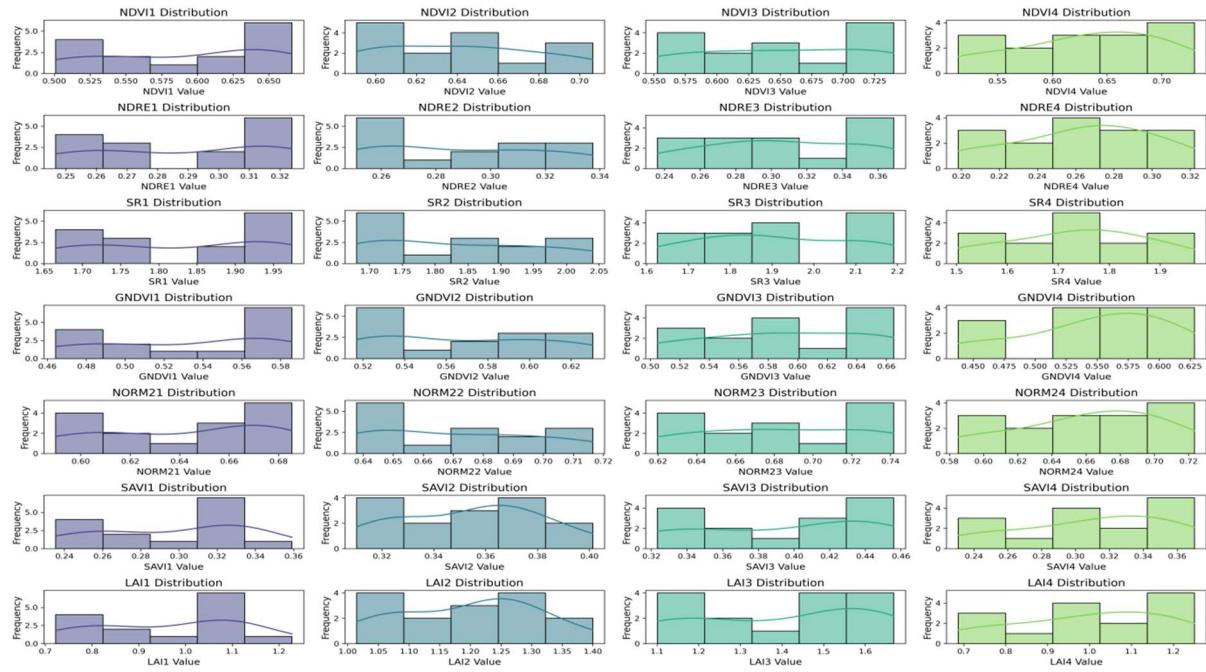
25-Sep-23	NDVI3, NDRE3, SR3, GNDVI3, SAVI3, NORM3, LAI3	
20-Oct-23	NDVI4, NDRE4, SR4, GNDVI4, SAVI4, NORM4, LAI4	Bulb initiation phase
30-Oct-23	NDVI5, NDRE5, SR5, GNDVI5, SAVI5, NORM5, LAI5	Bulb development phase
10-Nov-23	NDVI6, NDRE6, SR6, GNDVI6, SAVI6, NORM6, LAI6	

Distribution of VIs for Trial 1



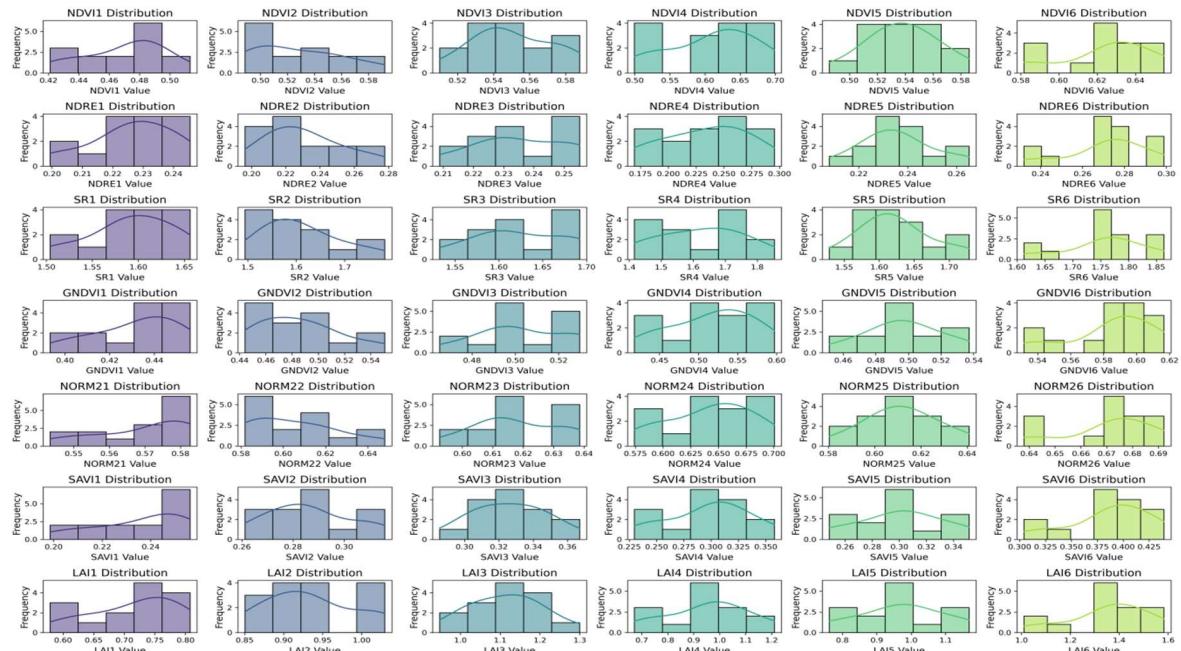
A

Distribution of VIs for Trial 2



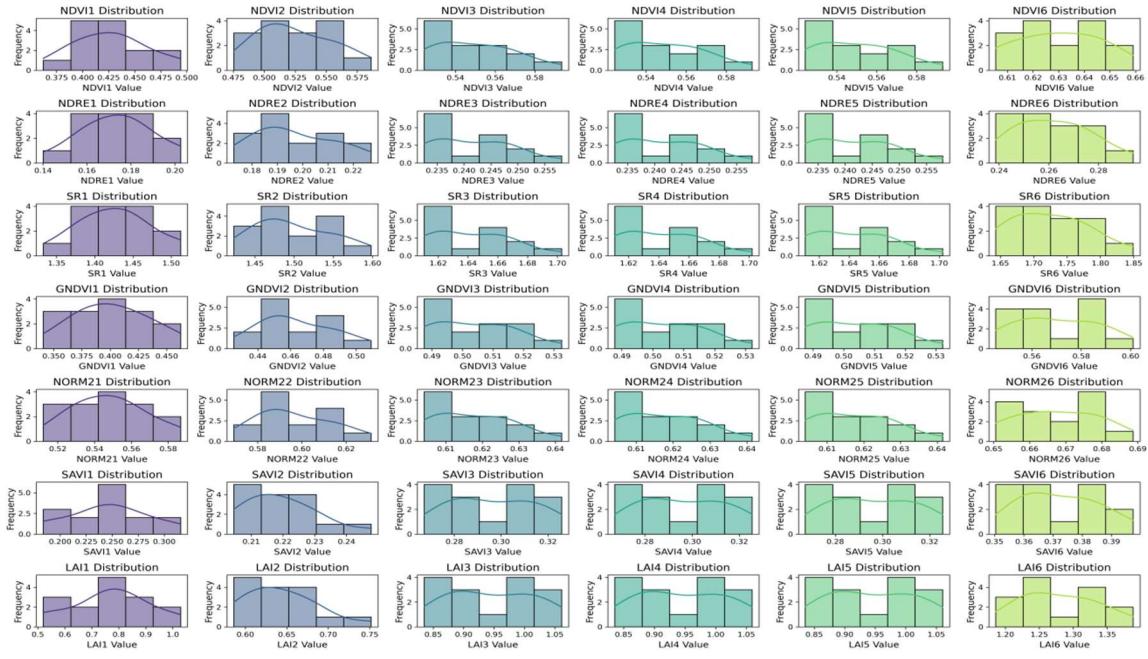
B

Distribution of VIs for Trial 3



C

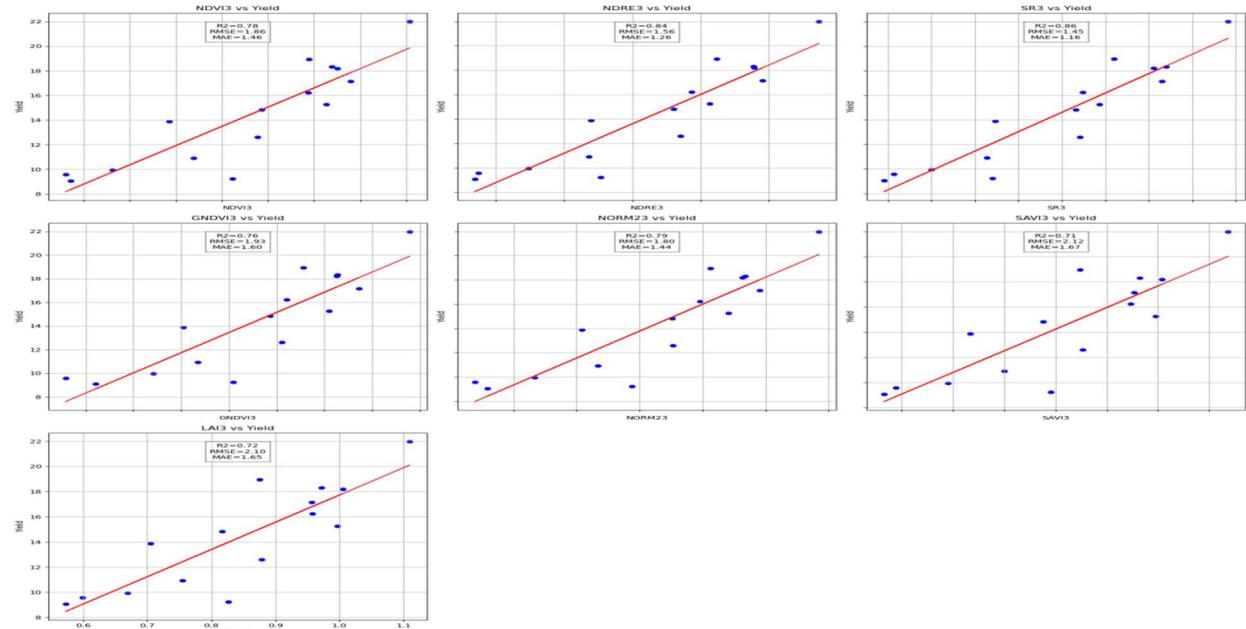
Distribution of VIs for Trial 4



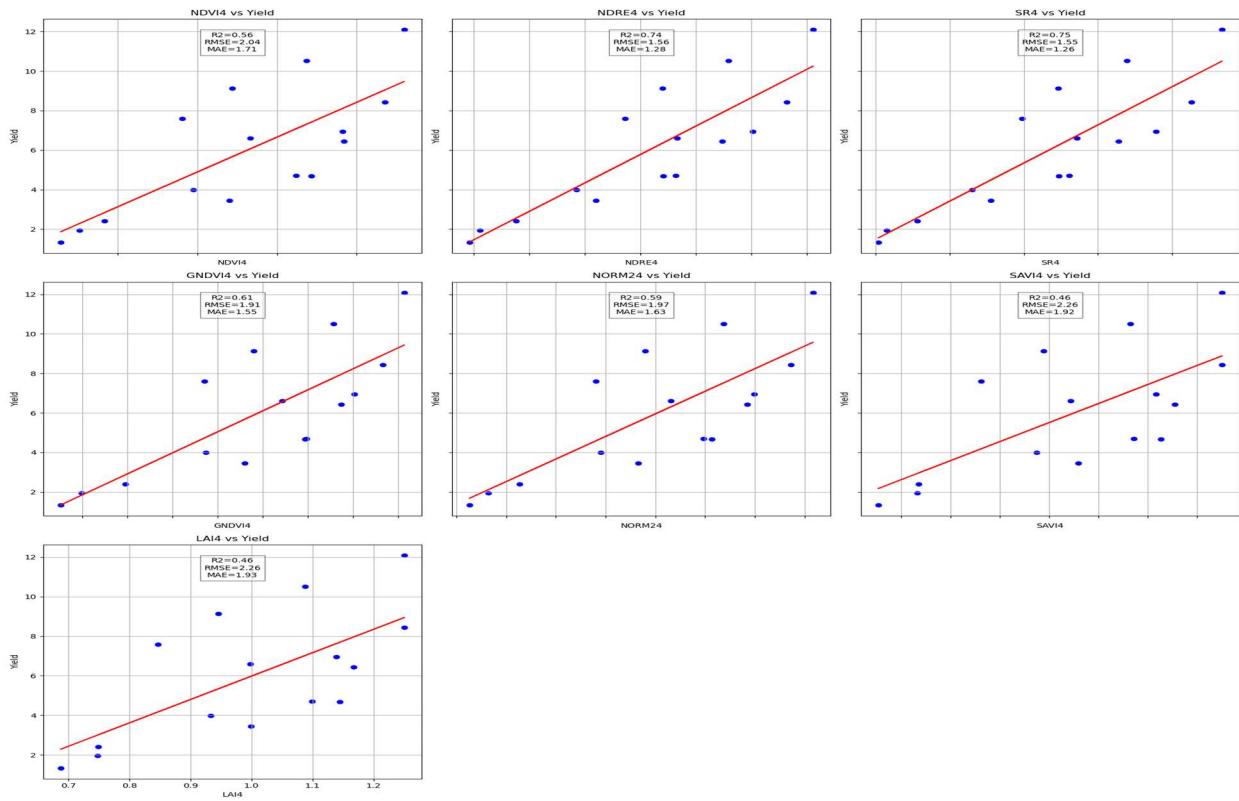
D

Fig.5. Histogram for 7 VI's A) experimental trial 1 B) experimental trial2 C) experimental trial3 D) experimental trial 4.

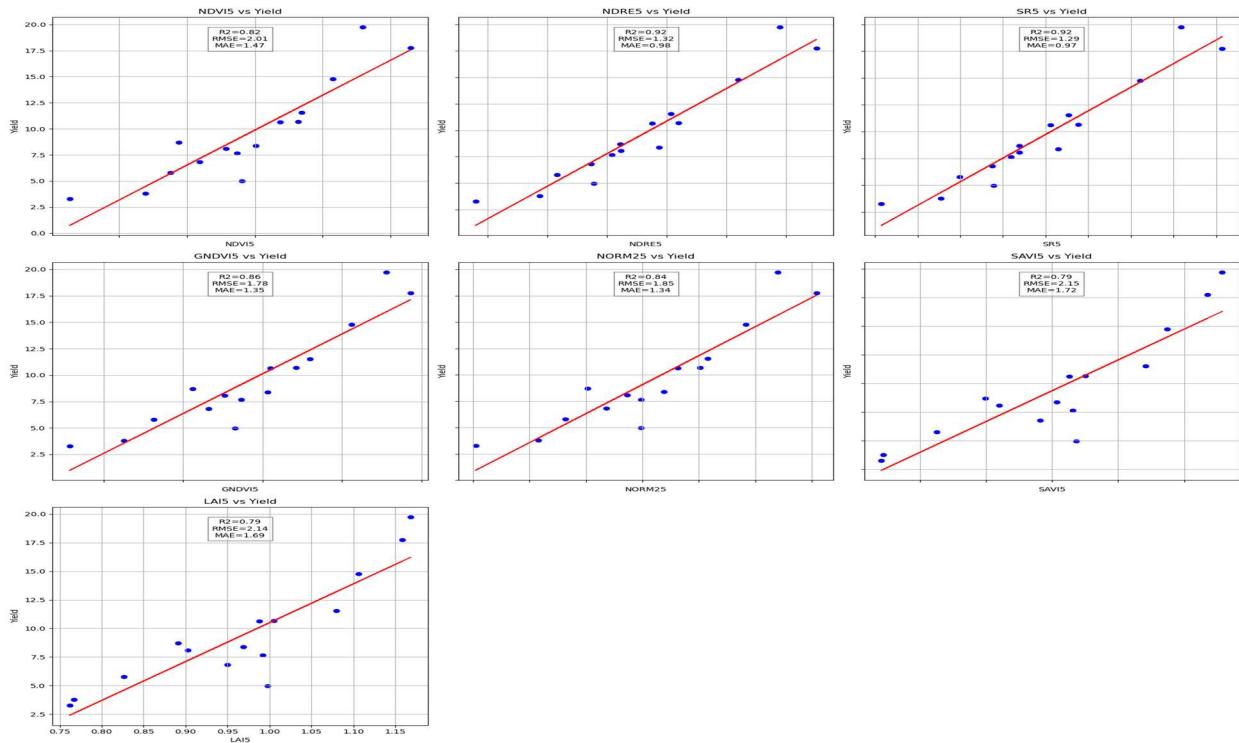
Fig7. The linear association between onion yield and NDVI, NDRE, NORM2, SR, GNDVI, SAVI and LAI measured on bulb development period. A) Trial1, B) Trial2, C) Trial3 and D) Trial4.



B.



C.



D.

