Samuel Naik B
Division of Agricultural
Statistics, ICAR-Indian
Agricultural Statistics Research
Institute, Pusa, New Delhi,
India

Harish Nayak GH
Division of Agricultural
Statistics, ICAR-Indian
Agricultural Statistics Research
Institute, Pusa, New Delhi,
India

Dr. S Govinda Rao
Assistant Professor, Department
of Statistics and Computer
Applications, ANGRAU
Agricultural College Naira,
Srikakulam, Andhra Pradesh,
India

Prediction of wheat yield by using UAV RGB drone imagery and advanced machine learning techniques

Samuel Naik B, Harish Nayak GH and Dr. S Govinda Rao

Abstract
Yield prediction before harvest is one of the important issues in terms of managing agricultural policies and making the right decisions for the future. The aim of this study is to predict wheat yield using field phenotypic data obtained from unmanned aerial vehicle (UAV) images and advanced machine learning techniques. A UAV platform carrying RGB cameras was employed to collect images of wheat crop. These images (402 images) were combined to form ortho-mosaic image through image processing software Pix4D mapper and were used to extract the vegetation indices (VIs), canopy volume, canopy area by quantum geographic information system (QGIS) open-source software. The yield prediction was done with the help of green leaf area index (GLA), the excess red index (ExR), the excess green index (ExG), the excess green minus excess red index (ExGR), water index (WI), the normalised green-red difference index (NGRDI), the red green blue VI (RGBVI), and the visible atmospherically resistant index (VARI) obtained from UAV RGB images. In addition, some digital variables were also used to reflect the growth trend of wheat, including G/R, G/B, and R/B. The results show that the models can accurately predict yield before the harvest. Support vector machine (SVM) (RMSE=1.025, R^2=0.93) and least absolute shrinkage and selection operator (LASSO) regression (RMSE=1.022, R^2=0.93) represent the top two best methods for predicting yields among the five typical machine learning models tested in this study. Our findings highlight a potentially powerful tool to predict yield using UAV drone data and advanced machine learning techniques fin other regions and for crops.

Keywords: UAV, drone, RGB cameras, QGIS, PiX4D mapper, machine learning

1. Introduction
Wheat (*Triticum aestivum* L.), as one of the three top grains (wheat, rice, and corn) and one of the most productive cereals in the 21st century, provides the most calories and protein for the global food supply (Ramadas *et al*., 2019) [18]. India is a leading producer of wheat in the world, with a large area of land devoted to the crop. The accurate prediction of crop yields in advance plays an important role in the grain circulation market, famine prevention, and food security. The prediction of crop yield using phenotypic characteristics were significantly improves the accuracy.
The use of Unmanned Aerial Vehicles (UAVs), commonly referred to as drones, has seen a significant increase in recent years, particularly in the field of agriculture. UAVs provide a cost-effective and flexible alternative to traditional remote-sensing platforms such as satellites and manned aircraft. By equipping UAVs with various sensors, including RGB, multispectral, hyperspectral, thermal, and light detection and ranging (LiDAR), farmers and researchers can obtain a wide range of data on crops, including vegetation indices, plant number, plant height, canopy area, and more. Additionally, agricultural statisticians can use data obtained from UAVs to generate accurate and up-to-date statistics on crop yields, crop acreage, and other important agricultural metrics. UAVs with multiple advantages, such as flexibility, non-destructive monitoring, low costs, and high amount of output, has been increasingly used in the field of precision agriculture in recent years (Bending *et al*., 2012; Bending *et al*., 2014). UAVs can carry different kinds of sensors, such as a RGB, multispectral, hyperspectral, thermal, LiDAR. The cost of these sensors is very high, thus the using of these sensors in agriculture is very less. But in these sensors RGB is a cheap sensing device, and it is becoming a promising tool for continuous observation (Fu *et al*., 2021) [8].

Corresponding Author:
Samuel Naik B
Division of Agricultural
Statistics, ICAR-Indian
Agricultural Statistics Research
Institute, Pusa, New Delhi,
India

So that, by using UAV-RGB remote sensing some advances have been made in crop phenotypic characteristic monitoring like Leaf Area Index (Hasan *et al*., 2019) [12], plant growth (Jamil *et al*., 2022) [13], plant density (Jin *et al*., 2017) [14], crop canopy area (Fu *et al*., 2021) [8], etc. Vegetation indices (VI) are generally used to estimate biophysical parameters that can be incorporated in models to predict crop yield (Myneni *et al*., 1995) [15]. Some sample vegetation indices (VIs) such as visible atmospherically resistant index (VARI) (Gitelson *et al*., 2002) [9], normalised green-red difference index (NGRDI) (Elazab *et al*., 2015) [15], red green blue VI (RGBVI) (Bending *et al*., 2015), green leaf area (GLA) (Guijarro *et al*., 2011) [10], excess red (ExR), excess green (ExG), and excess green red (ExGR) (Woebbecke *et al*., 1995) [17] were calculated based on images acquired from the visible wavelengths. In addition, some digital variables were also used to reflect the growth trend of ramie, including g/r, g/b, and r/b. The images were processed to extract information on the vegetation indices, plant heights, canopy volume and crop canopy area etc.

In recent decades, many researchers have been increasingly focused on improving crop yield prediction by different methods, including empirical statistical models and process-oriented crop growth models (Lobell *et al*., 2010) [19]. Conventional statistical models predict yields by developing regression equations between weather variables (temperature, precipitation, solar radiation, etc.) and measured yields at different temporal and spatial scales (Shi *et al*., 2013; Zhang *et al*., 2016) [20, 21]. Such regression results did show distinctly how climatic factors affected yield, however their relative lower explanation ability was commonly debated, and the dominant factors controlling yields often varied by geographical location, crop variety, and growing season (Filippi *et al*., 2019) [22]. Machine learning has demonstrated its powerful performance in data mining (Witten *et al*., 2002) [23] and agricultural analyses, including crop type classification and yield prediction (Cai *et al*., 2019) [24]. Crop yield is a function of the interaction between spatial and temporal changes of variables. Considering the strong ability in treating multi-dimensional datasets mining (Witten *et al*., 2002) [23]. Accordingly, machine learning techniques could provide powerful supports for improving yield prediction models. Several publications conducted recently in the world have substantiated this viewpoint.

In this study, we integrated 13 indicators derived from UAV RGB images, to build machine learning models for predicting wheat yield. We adopted five machine learning algorithms for predicting wheat yield, including K-nearest neighbor (KNN), decision tree (DT), support vector machine (SVM), random forest (RF), least absolute shrinkage and selection operator (LASSO) regression. The main objective of this study is to apply UAV-based spectral and structural information for the prediction of wheat yield. This study will propose to select the better machine learning algorithms for yield prediction.

2. Materials and Methods
2.1 Study area
In this study, the data of RGB sensors data has been captured from wheat fields located at the Indian Agricultural Research Institute (IARI), (28.6331° N, 77.1525° E). The wheat crop was growing in 0.012 km^2 / 1.2 ha / 2.9652 acres, with 9 X 9rc. The data was processed using PiX4D mapper software, which generated a quality report and a single TIFF image (obtained by mosaicking of 402 images). The quality report provided information on the hardware of the drone and assessed the quality of the processed data. If the quality check was satisfactory, the TIFF image was imported as a raster layer in QGIS.

The study area was divided into two parts, designated as Area 1 and Area 2, to obtain the exact canopy area from each experimental block. Various crop parameters, such as canopy area, were extracted using QGIS software. These spectral parameters were subsequently used for data analysis.

(A)	(B)	(C)
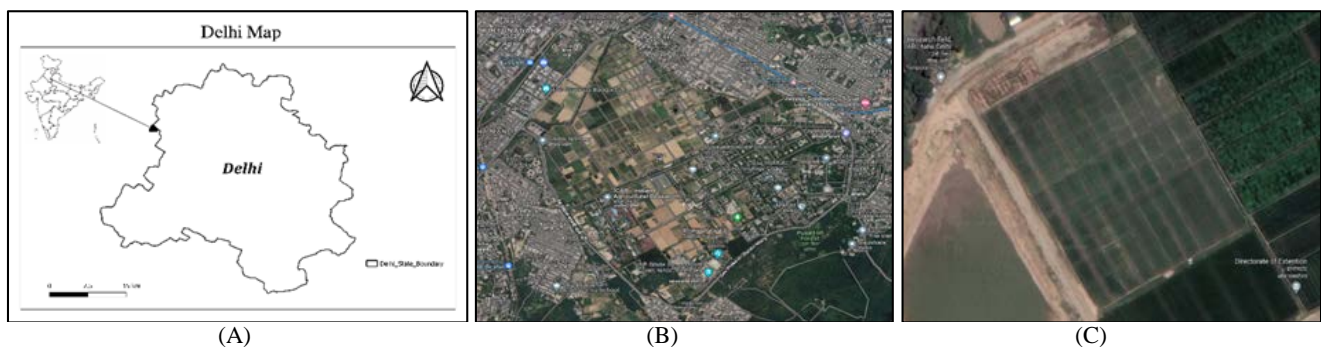
Fig 1: Study area and arrangement of the experimental sites: (a) Geographic map of Delhi, (b) IARI, Pusa, New Delhi, (c) Wheat crop area

2.2 Image Acquisition
The data was captured by drone named DJI Phantom 4 with its impressive flight range of 7 kilometres, 1.3 kg payload capacity, state of the art 4K camera and 30-minute flight time in a windless environment. The field images were acquired at 12:00–14:00 local time under clear and sunny weather conditions. 402 images with pixels of 5472 x 3648 were obtained, and these images were saved in 8-bit unsigned integer data in tag image file format (TIFF). The UAV platform was used to implement a predefined flight plan that was created in advance by a ground control station. During the flight, the flight speed was 6 m/s, with average ground sampling distance (GSD) is 0.79 cm / 0.31 in.
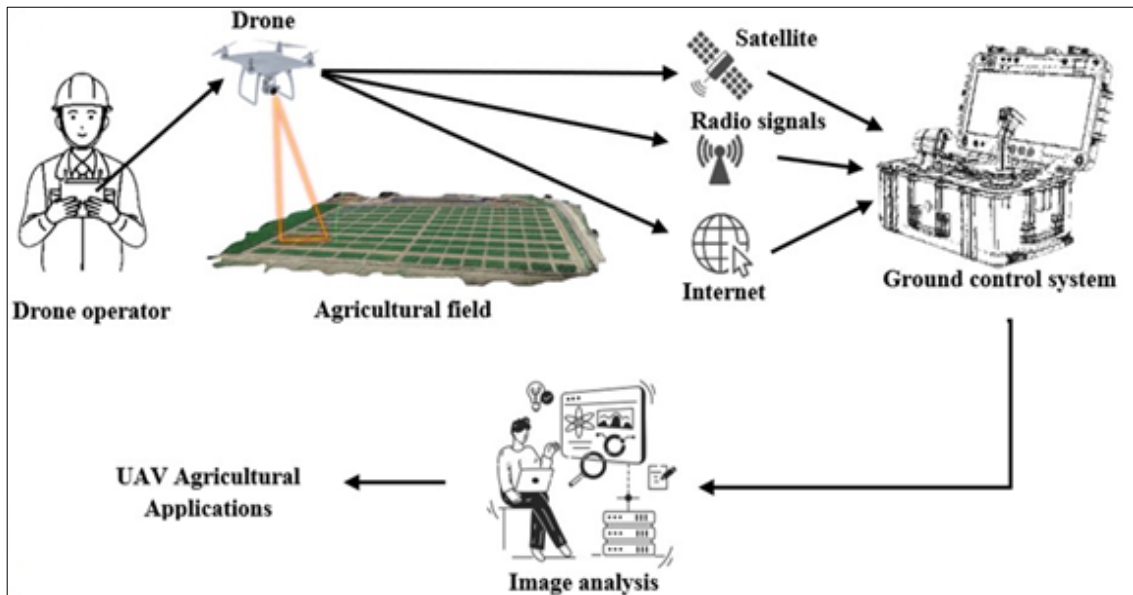
Fig 2: UAV data acquisition methodology

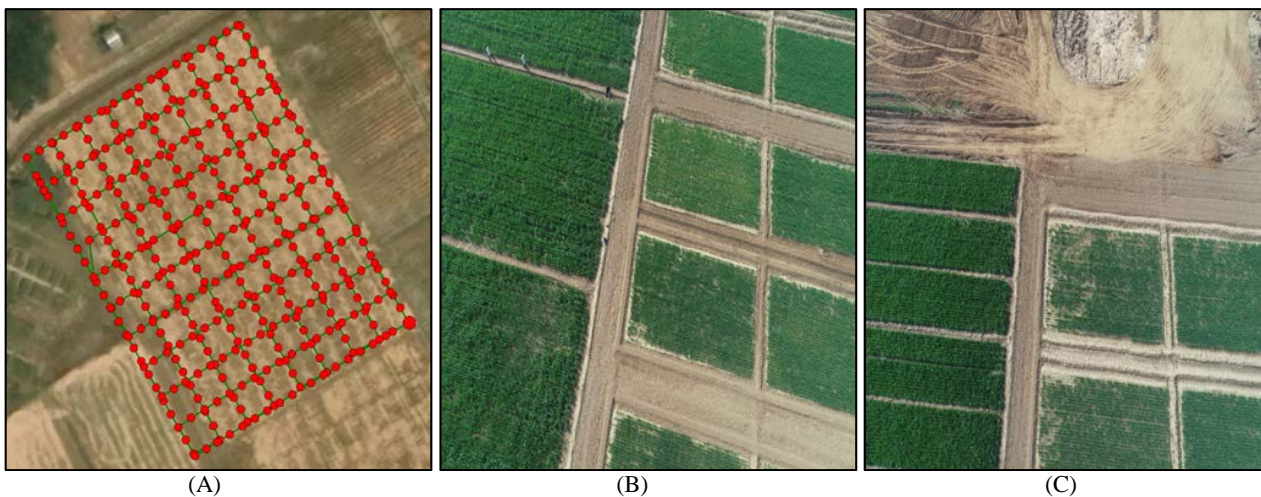**Fig 3:** DJI Phantom 4 drone image

| (A) | (B) | (C) |

**Fig 4:** UAV image acquisition: a) Image captured points, b) Individual images

2.3 Image mosaicking and processing

Image mosaic is a technique that combines several images with overlapping parts (the images may be obtained at different times, different viewing angles or by different sensors) into a large-scale seamless high-resolution image. The captured 402 images, were mosaicked using Pix4Dmapper. The basic operation process includes:

1. Verifying the POS data extracted from the UAV images to ensure the correct geographical location of the images.

2. Extracting the spatial position and altitude information of a certain point X from all the images, creating a point cloud map of the study area.

3. Adding the 3D spatial information of control points to the empty ray editor to align the images captured at different times to the same coordinate system. The final output is an ortho photo image of the wheat crop area.
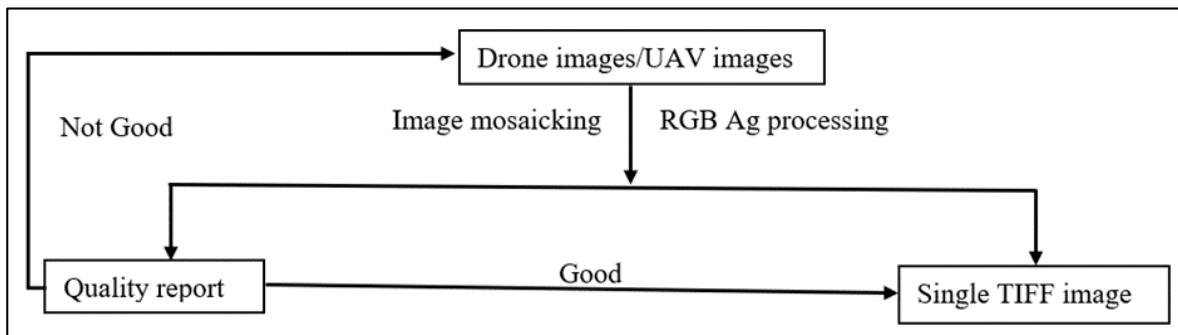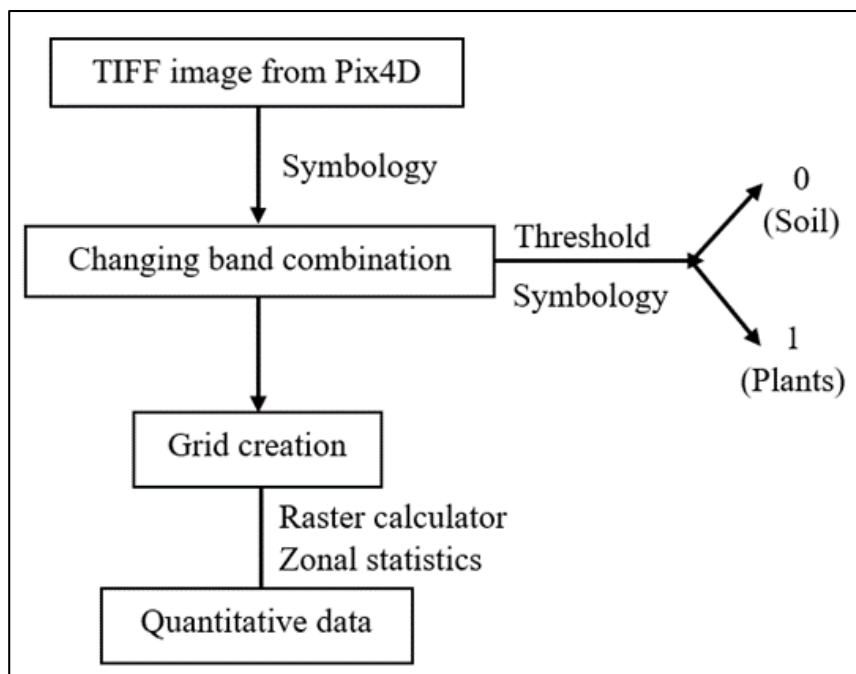
Flow chart 1: Image processing in Pix4D mapper
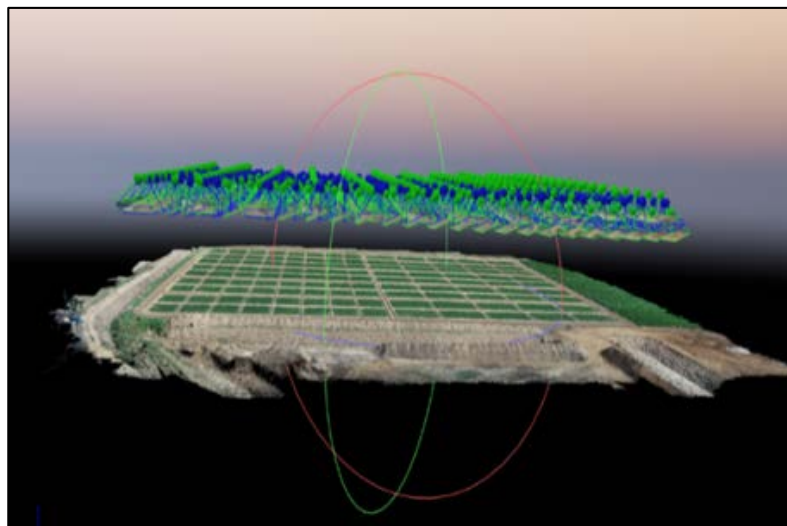
Fig 5: Ortho-mosaic image of wheat field

The mosaicked image was used as a raster layer in QGIS, where various crop parameters were extracted.

Flow chart 2: Image processing in QGIS software

2.4 Vegetation Index Calculation
Table 1 presents the VIs selected in this study, their calculations, and the corresponding sources. The VIs included GLA, the excess red index (ExR), the excess green index (ExG), the excess green minus excess red index (ExGR), water index (WI), the normalised green-red difference index (NGRDI), the red green blue VI (RGBVI), and VARI. In addition, some digital variables were also used to reflect the growth trend of ramie, including g/r, g/b, and r/b.
Along with the VIs structural features like canopy volume and canopy height were used in this study.

Table 1: Indices calculated based on the unmanned aerial vehicle (UAV) images in this study

Vegetation Indices	Reference
GLA= (2 × G − R − B) / (2 × G + R + B) WI = (G − B) / (R − G) NGRDI = (G − R) / (G + R) RGBVI = (G × G − R × B) / (G × G + R × B) VARI = (G − R) / (G + R − B) ExR = 1.4R − G ExG = 2×G−R−B ExGR = ExG − 1.4R − G G/R G/B R/B	Xiaoqin et al., 2015 [25] Penuelas et al., 1997 [26] Hunt et al., 2005 [27] Bendig et al., 2015 [3] Gitelson et al., 2002 [9] Woebbecke et al., 1995 [17] Woebbecke et al., 1995 [17] Woebbecke et al., 1995 [17] Fu et al., 2021 [8] Fu et al., 2021 [8] Fu et al., 2021[8]

2.5 Plant Height Estimation

The calculation of the plant height was carried out in QGIS open software. Firstly, the digital surface model (DSM) and the digital terrain model (DTM) were constructed based on the acquired images. DSM represented the absolute height of the crop canopy, while DTM represented the absolute height of the ground. Secondly, crop surface models (CSMs) were obtained by subtracting DTM from DSM with a raster calculator tool in QGIS open software. Subsequently, the average plant height of each plot was calculated. Finally, the plant height calculated according to the UAV images was validated based on the field-measured plant heights.

2.6 Field Date Collection

Measured field data at the maturity period, including plant height and yield. We took the average of 15 measured values as the field-measured plant height in the plot.

2.7 Machine-Learning Methods for Estimating Crop Yield

Five advanced machine learning algorithms were applied here. They are explained below:

2.7.1 K-Nearest Neighbor Regression

The K-nearest neighbor (KNN) approach is a type of instance-based learning, which is based on the distance of the predictor variables to the nearest training group known to the model (Bebie et al., 2022 and Appelhans et al., 2015) [28, 30]. Aha et al. firstly proposed the new framework and methodology for KNN. KNN can tolerate noise and unrelated properties and has a relatively relaxed concept bias.

2.7.2 Decision Tree (DT)

The decision tree is an effective tool for solving classification and regression problems and has been widely used in remote sensing application (Xu et al., 2005) [31]. The tree consists of a root node (containing all data), internal nodes, and several leaves. Each node makes a binary decision to separate different categories until the leaf node is reached. The algorithm is non-parametric and can deal with large and complex datasets effectively without complex parameter structure (Song et al., 2015) [32]. C4.5 decision tree is a method of approximating discrete value function, which is robust to noisy data (Polat et al., 2009) [33]. The confidence factor used for pruning is 0.25 and the minimum number of instances per leaf is 2.

2.7.3 Support Vector Machine (SVM)

SVM is a supervised non-parametric algorithm, which is characterized by using the kernels and acting on the margins (Gunn et al., 1998) [34]. During SVM regression, the input is mapped to a high-dimensional feature space using a kernel function, and then a linear regression model is constructed in the new feature space to balance between minimizing errors and overfitting (Cai et al., 2019 and Hearst et al., 1998) [24, 35]. Kernel functions (linear, polynomial, Gaussian, etc.) are one of the important hyper-parameters that need tuning. By comparing different kernel functions, the Gaussian kernel function performed the best in this study.

2.7.4 Random Forest (RF)

Random forests are a combination of tree predictors and are more robust with respect to noise (Breiman et al., 2001) [36]. Each tree is built by selecting random variable sets and dataset samples, and all the trees in the forest have the same distribution characteristic. After generating a large number of individual trees, they will vote for the most popular classes. Therefore, RF shows the efficiency to handle high-dimensional datasets and avoids overfitting during the past decade (Rhee et al., 2017; Vincenzi et al., 2011) [37, 38]. Additionally, RF can quantify the relative importance of measured variables and is a reasonable method for variable selection (Breiman et al., 2001 and Strobl et al., 2007) [36, 39].

2.7.5 Least Absolute Shrinkage and Selection Operator (LASSO) Regression

Lasso is a regularization technique which reduces the number of predictors in a regression model and identifies important predictors. Lasso is a shrinkage estimator with potentially lower predictive errors than ordinary least squares and it also includes a penalty term that constrains the size of the estimated coefficients. Therefore, it resembles ridge regression. It generates coefficient estimates that are biased to be small. Nevertheless, a lasso estimator can have smaller mean squared error than an ordinary least-squares estimator when you apply it to new data. Unlike ridge regression, as the penalty term increases, lasso sets more coefficients to zero. This means that the lasso estimator is a smaller model, with fewer predictors. As such, lasso is an alternative to stepwise regression and other model selection and dimensionality reduction techniques (Singh et al., 2019) [29].

2.8 Model Evaluation

In order to evaluate the five ML techniques, we adopted the root-mean-square error (RMSE), the coefficient of determination (R^2), and the mean absolute error (MAE) to evaluate the performance of the machine learning model, which can be calculated as follows:

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

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

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

where n (i = 1, 2, . . ., n) is the number of samples used for machine learning model, y_i is the observed winter wheat yield, $\bar{y}_i$ is the corresponding mean value, f_i is the predict winter

wheat yield, $\bar{f_i}$ is the corresponding mean value. The closer R^2 is to 1, the higher the prediction performance of the model is small RMSE and MAE values indicate less discrepancy within the observed yield and predicted yield.

3. Results
3.1 Comparison of Training Accuracy of Wheat Yield Prediction Models:
In this study, five machine learning models were trained with the observed yields and 13 variables of wheat taken from UAV imagery. The evaluated results, based on R^2, RMSE and MAE. Comprehensively SVM and LASSO models showed the higher accuracy, with higher R^2 (0.93) and lower RMSE (1.025 and 1.022) and MAE (0.807 and 0.826). Although, R^2 of other models are above 0.8, all their RMSEs were > 1.0. Figure 6 shows the evaluated results, based on R^2, RMSE and MAE.
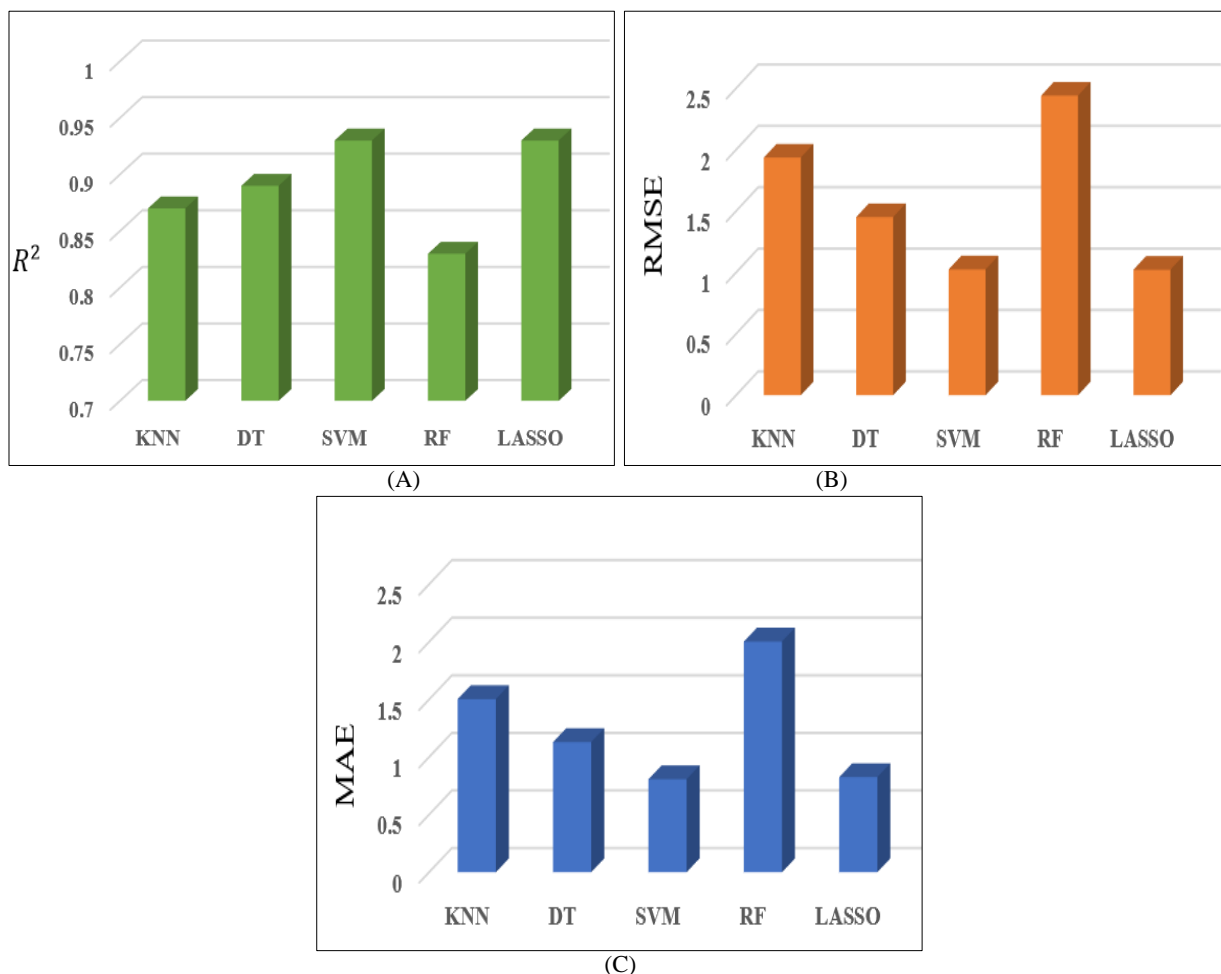

Fig 6: a) R^2, b) RMSE, c) MAE of five different machine learning models

3.2 Wheat Yield Prediction using UAV-based image data
Wheat yield prediction based on UAV image data has been done using the all five machine learning techniques. The scatter diagrams of predicted and actual yields of the models are shown in figure 7. We found that the two machine learning models can predict the yield of wheat with high accuracy, in the order is LASSO>SVM>DT>RF>KNN. Although all predicted yields were much closer to the line
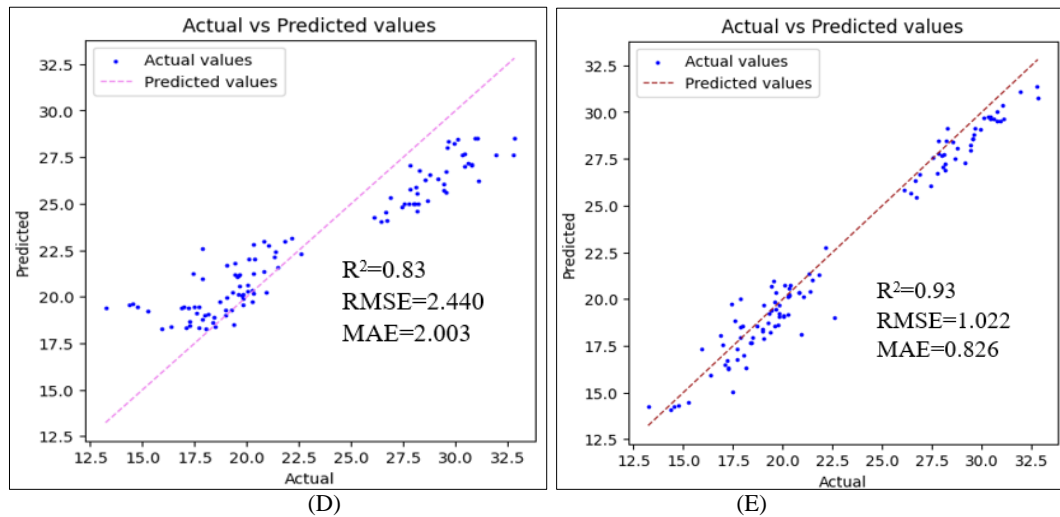
(D)　　　　　　　　　　　　　　　　　(E)
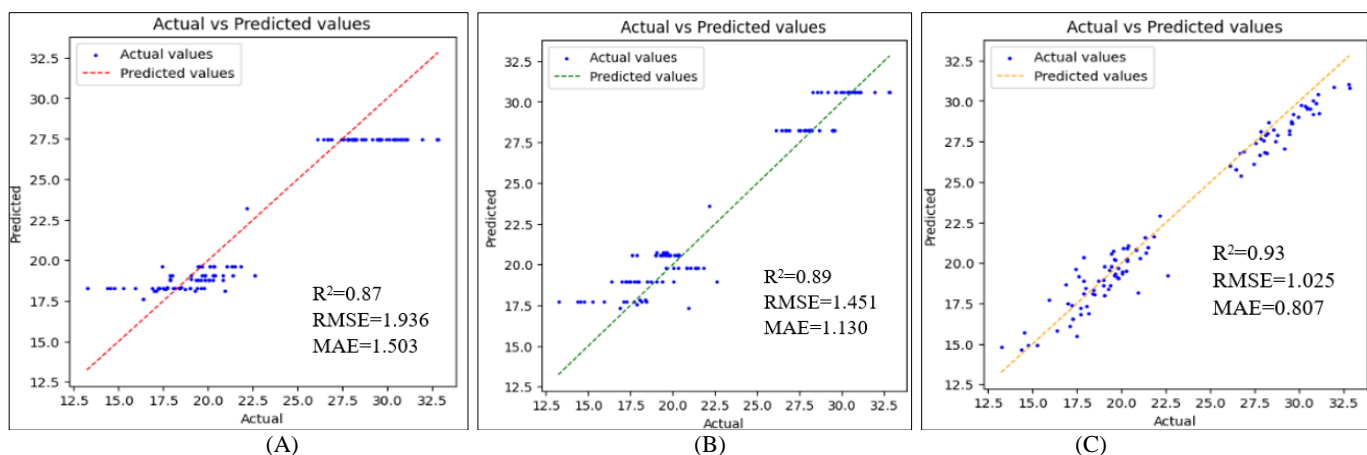
Fig 7: Scatter plots of observed yield and predicted yield of a) KNN, b) DT, c) SVM, d) RF, e) LASSO

Table 2: Correlation analysis of the UAV-based spectral features and yield

	Yield (kg/plot)	Canopy area sq m	Mean canopy height (m)	GLA	WI	RGBVI	VARI	NGRDI	ExG	ExR	ExGR	G/R	G/B	R/B
Yield(kg/plot)	1													
Canopy area sq m	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:
** means significant at 0.01 and
*** means significant at 0.05

3.3 Relationship between VIs and wheat Yield
Table 2 shows the results of the correlation analysis between the UAV-based spectral features and wheat yield. Our results indicated that the correlation relationship between the wheat yield and all of the visible spectrum indices calculated based on the UAV images was weak, of which WI, ExR, and R/B were negatively correlated with the wheat yield, while canopy area, canopy height, NGRDI, GLA, ExG, ExGR, G/R, G/B, and RGBVI were positively correlated with the wheat yield. Among all the VIs, VARI was the most significantly correlated with the WHEAT yield, and the correlation coefficient was 0.391.

3.4 Prediction variable and the order of relative importance
To investigate the importance of different variables for the wheat yield prediction, we calculated the decreased accuracy (mean square error) from RF model. Figure 3 represents the predictor variable importance based on the RF model.
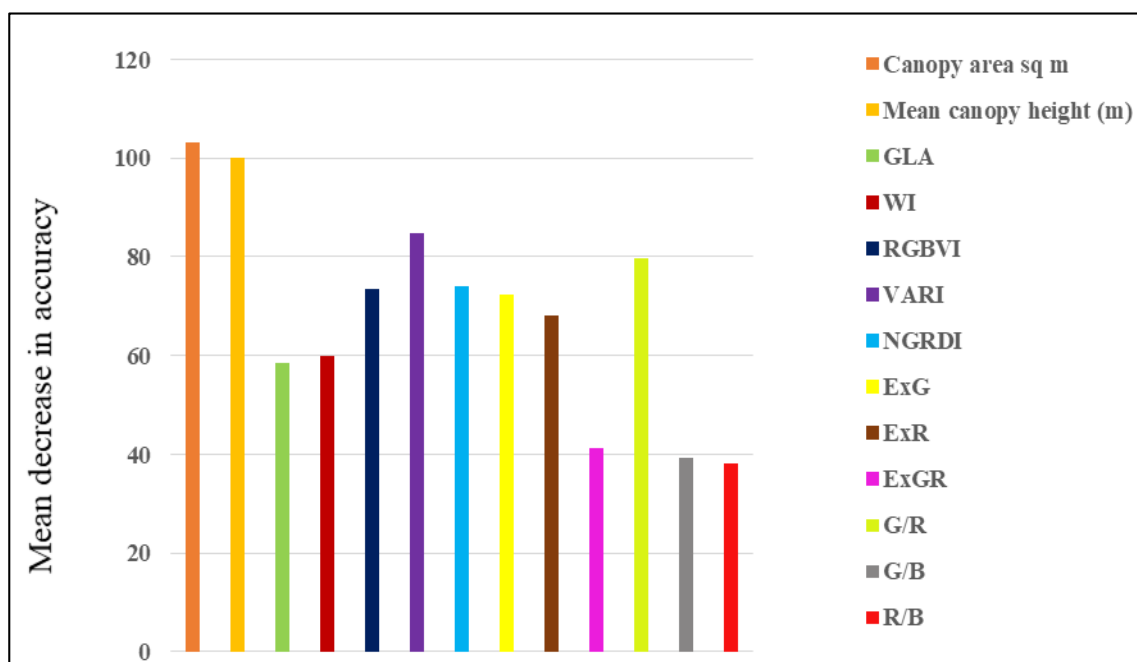
Fig 8: Predictor variable importance based on the RF model

4. Discussion
4.1 Model Performance for Estimating Yields
Some studies have shown that the spectral information obtained from UAV-based images has a certain potential for crop yield prediction. In this study, we proposed a framework for yield prediction through structural and spectral features of UAV imagery and using advanced machine learning techniques. The differences in various machine learning algorithms have been discussed above. Among all the machine learning algorithms SVM and LASSO are performing better.

4.2 Relationship between UAV-Based Image Data and wheat Yield
Many studies have demonstrated that there is a significant correlation between spectral and structural features obtained by UAV-based images and crop yields (Tilly *et al*., 2015) [16]. Our study confirmed that the canopy volume, canopy height, GLA, RGBVI, VARI, NGRDI, ExG, ExGR G/R, G/B obtained by UAV-based images could provide a good characterization of the wheat yield, but they possessed different potentials. Not every vegetation index can accurately reflect the crop yield. If the correlation between the VIs and yield are unacceptable, the fusion of the VIs will result in a saturation of the spectral information and reduce the accuracy of the yield estimation model. In general, wheat yield prediction by both spectral and structural features improved the accuracy. In future research, other information can be considered to improve the reliability and accuracy of the model.

5. Conclusions
In this study, a UAV platform carrying RGB cameras was employed to collect several images of wheat crop. These images were combined to form single orthophoto image using Pix4D mapper software and were used to extract fifteen field parameters like vegetation indices (VIs), canopy volume, canopy height etc., by QGIS open-source software. These variables were incorporated in the predicted wheat yield based multiple advanced machine learning models. It was found that the models can accurately predict yield before the harvest. SVM (RMSE=1.025, R^2=0.93) and LASSO regression

(RMSE=1.022, R^2=0.93) represent the top two best methods for predicting yields among the five typical machine learning models tested in this study. Our findings highlight a potentially powerful tool to predict yield using UAV drone data and advanced machine learning techniques fin other regions and for crops. The study reveals that it is feasible to monitor crop growth and development based on UAV images and that the combination of phenotypic data with it can further improve its accuracy. The study reveals that Drone images may be used for obtaining various crop parameters like VIs and crop canopy area etc., which can be used for obtaining more efficient and reliable parameters of agricultural statistics like crop area, crop yield and production.

6. References
1. Bendig J, Bolten A, Bareth G. Introducing a low-cost mini-UAV for thermal-and multispectral-imaging. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2012;39:345-349.
2. Bendig J, Bolten A, Bennertz S, Broscheit J, Eichfuss S, Bareth G. Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. Remote sensing. 2014;6(11):10395-10412.
3. Bendig J, Yu K, Aasen H, Bolten A, Bennertz S, Broscheit J, *et al*. 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. 2015;39:79-87.
4. Beriya A. Application of drones in Indian agriculture. ICT India Working; c2022. p. 73.
5. Choudhary SS, Biswal S, Saha R, Chatterjee C. A non-destructive approach for assessment of nitrogen status of wheat crop using unmanned aerial vehicle equipped with RGB camera. Arabian Journal of Geoscience. 2021;14(17):1739.
6. Elazab A, Bort J, Zhou B, Serret MD, Nieto-Taladriz MT, Araus JL. The combined use of vegetation indices and stable isotopes to predict durum wheat grain yield

under contrasting water conditions. Agricultural Water Management. 2015;158:196-208.

7. El-Hendawy SE, Hassan WM, Al-Suhaibani NA, Schmidhalter U. Spectral assessment of drought tolerance indices and grain yield in advanced spring wheat lines grown under full and limited water irrigation. Agricultural Water Management. 2017;182:1-12.

8. Fu H, Wang C, Cui G, She W, Zhao L. Ramie yield estimation based on UAV RGB images. Sensors. 2021;21(2):669.

9. Gitelson AA, Kaufman YJ, Stark R, Rundquist D. Novel algorithms for remote estimation of vegetation fraction. Remote sensing of Environment. 2002;80(1):76-87.

10. Guijarro M, Pajares G, Riomoros I, Herrera PJ, Burgos-Artizzu XP, Ribeiro A. Automatic segmentation of relevant textures in agricultural images. Computers and Electronics in Agriculture. 2011;75(1):75-83.

11. Han L, Yang G, Dai H, Xu B, Yang H, Feng H, *et al*. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. Plant methods. 2019;15(1):1-19.

12. Hasan U, Sawut M, Chen S. Estimating the leaf area index of winter wheat based on unmanned aerial vehicle RGB-image parameters. Sustainability. 2019;11(23): 6829.

13. Jamil N, Kootstra G, Kooistra L. Evaluation of Individual Plant Growth Estimation in an Intercropping Field with UAV Imagery. Agriculture. 2022;12(1):102.

14. Jin X, Liu S, Baret F, Hemerlé M, Comar A. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. Remote Sensing of Environment. 2017;198:105-114.

15. Myneni RB, Hall FG, Sellers PJ, Marshak AL. The interpretation of spectral vegetation indexes. IEEE Transactions on Geoscience and remote Sensing. 1995;33(2):481-486.

16. Tilly N, Aasen H, Bareth G. Fusion of plant height and vegetation indices for the estimation of barley biomass. Remote Sensing. 2015;7(9):11449-11480.

17. Woebbecke DM, Meyer GE, Von Bargen K, Mortensen DA. Color indices for weed identification under various soil, residue, and lighting conditions. Transactions of the ASAE. 1995;38(1):259-269.

18. Ramadas S, Kumar TK, Singh GP. Wheat production in India: Trends and prospects. In Recent Advances in Grain Crops Research. IntechOpen: Rijeka, Croatia, Chapter 6; c2019.

19. Lobell DB, Burke MB. On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology. 2010;150(11):1443-1452.

20. Shi W, Tao F, Zhang Z. A review on statistical models for identifying climate contributions to crop yields. Journal of Geographical Sciences. 2013;23:567-576.

21. Zhang Z, Song X, Tao F, Zhang S, Shi W. Climate trends and crop production in China at county scale, 1980 to 2008. Theoretical and Applied Climatology. 2016;123:291-302.

22. Filippi P, Jones EJ, Wimalathunge NS, Somarathna PD, Pozza LE, Ugbaje SU, *et al*. An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. Precision Agriculture. 2019;20:1015-1029.

23. Witten IH, Frank E. Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record. 2002;31(1):76-77.

24. Cai Y, Guan K, Lobell D, Potgieter AB, Wang S, Peng J, *et al*. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and Forest Meteorology. 2019;274:144-159.

25. Xiaoqin W, Miaomiao W, Shaoqiang W, Yundong W. Extraction of vegetation information from visible unmanned aerial vehicle images. Transactions of the Chinese Society of Agricultural Engineering. 2015;31:5.

26. Peñuelas J, Pinol J, Ogaya R, Filella I. Estimation of plant water concentration by the reflectance water index WI (R900/R970). International Journal of Remote Sensing. 1997;18(13):2869-2875.

27. Hunt ER, Cavigelli M, Daughtry CS, Mcmurtrey JE, Walthall CL. Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. Precision Agriculture. 2005;6:359-378.

28. Bebie M, Cavalaris C, Kyparissis A. Assessing Durum Wheat Yield through Sentinel-2 Imagery: A Machine Learning Approach. Remote Sensing. 2022;14(16):3880.

29. Singh KN, Singh KK, Kumar S, Panwar S, Gurung B. Forecasting crop yield through weather indices through LASSO. Indian Journal of Agricultural Sciences. 2019;89(3):540-544.

30. Appelhans T, Mwangomo E, Hardy DR, Hemp A, Nauss T. Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. Spatial Statistics. 2015;14:91-113.

31. Xu M, Watanachaturaporn P, Varshney PK, Arora MK. Decision tree regression for soft classification of remote sensing data. Remote Sensing of Environment. 2005;97(3):322-336.

32. Song YY, Ying LU. Decision tree methods: applications for classification and prediction. Shanghai archives of Psychiatry. 2015;27(2):130.

33. Polat K, Güneş S. A novel hybrid intelligent method based on C4. 5 decision tree classifier and one-against-all approach for multi-class classification problems. Expert Systems with Applications. 2009;36(2):1587-1592.

34. Gunn SR. Support vector machines for classification and regression. ISIS technical report. 1998;14(1):5-16.

35. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intelligent Systems Applications. 1998;13:18-28.

36. Breiman L. Random forests. Machine learning. 2001;45:5-32.

37. Rhee J, Im J. Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. Agricultural and Forest Meteorology. 2017;237:105-122.

38. Vincenzi S, Zucchetta M, Franzoi P, Pellizzato M, Pranovi F, De Leo GA, *et al*. Application of a Random Forest algorithm to predict spatial distribution of the potential yield of Ruditapes philippinarum in the Venice lagoon, Italy. Ecological Modelling. 2011;222(8):1471-1478.

39. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioin-formatics. 2007;8(1):1-21.