Khodjaev, Shovkat; Bobojonov, Ihtiyor; Kuhn, Lena; Glauben, Thomas

Article — Published Version

Optimizing machine learning models for wheat yield estimation using a comprehensive UAV dataset

Modeling Earth Systems and Environment

Modeling Earth Systems and Environment (2025) 11:15
https://doi.org/10.1007/s40808-024-02188-9

ORIGINAL ARTICLE

Optimizing machine learning models for wheat yield estimation using a comprehensive UAV dataset

Shovkat Khodjaev[1] · Ihtiyor Bobojonov[1] · Lena Kuhn[1] · Thomas Glauben[1]

Received: 10 September 2024 / Accepted: 30 October 2024

Abstract
Timely and accurate wheat yield forecasts using Unmanned Aircraft Vehicles (UAV) are crucial for crop management decisions, food security, and ensuring the sustainability of agriculture worldwide. While traditional machine learning algorithms have already been used in crop yield modelling, previous research used machine learning algorithms with default parameters and did not take into account the complex, non-linear relationships between model variables. Especially, the combination of vegetation indices, soil properties, solar radiation, and wheat height at the field estimation has not been deeply analysed in scientific literature. We present a machine learning based wheat yield estimation model using comprehensive UAV datasets with the implementation of hyperparameter tuning to improve model performance. The performance of the models before and after optimisations was measured using the metrics RMSE, MAE and R2, and the results showed that the models improved after tuning. Furthermore, we find that the Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models outperformed other examined models. Furthermore, a non-parametric Friedman test with a Nemenyi post-hoc test indicates that the best-performing algorithms for wheat yield estimation and prediction are RF and XGBoost models. In the final step, we utilised a SHapley Additive exPlanations approach to identify the direct impact of each input variable on the yield estimation model. Among the input variables, only the Red-Edge Chlorophyll Index, the Normalised Difference Red-Edge Index and wheat height were found to be of high explanatory power in predicting wheat yield. The optimised model is 7–12% more accurate in estimating wheat yields than traditional linear models.

Keywords Machine learning models · Drone sensors · Wheat yield · Comprehensive datasets · Feature importance

Introduction

The advancement of satellite-based remote sensing data has enabled timely and accurate yield estimates on a large scale (Ferencz et al. 2004; Singh et al. 2002). In agriculture, satellite-based crop yield estimation models make a significant contribution to avoiding crop losses, thereby improving food security as well as climate change mitigation (Ashapure et al. 2020). Higher-resolution optical satellites are needed to provide reliable estimations on village, farm, or plot levels, however often don't provide data at high frequency or historical dimension and also suffer from cloud-cover blockage.

✉ Shovkat Khodjaev
khodjaev@iamo.de; khodjaevshovkat@gmail.com

[1] Department of Agricultural Markets, Marketing and World Agricultural Trade, Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Theodor-Lieser- Str. 2, 06120 Halle (Saale), Germany

Agricultural drones (or unmanned aerial vehicles UAVs) are more flexible in terms of image collection time, provide a higher degree of accuracy, and are not affected by cloud cover (Cao et al. 2021). Moreover, the new generation UAVs can fly longer distances, providing opportunities to collect various climates, soil, fertilizers, nutrition levels, and crop height (Lopatin and Poikonen 2023). Recent studies demonstrated that UAV-based yield estimation models are significantly more accurate for predicting crop yields in various areas compared to traditional methods (Gevaert et al. 2015; Honkavaara et al. 2013).

In recent years, the development of artificial intelligence has opened a wide range of applications of machine learning (ML) and deep learning (DL) algorithms, which can significantly outperform other traditional linear methods (Milla-Val et al. 2024; Mwaura & Kenduiywo 2021). ML and DL models are now also successfully applied to estimate crop yield. For instance, Everingham et al. (2016) demonstrated that the random forest (RF) model can estimate sugarcane

Published online: 19 December 2024

yield estimation better than the traditional linear model (LM). Ajith et al. (2024) indicated that the support vector (SV) model is the most powerful for weather-index-based yield estimation among the linear models. You et al. (2017) applied convolutional neural machine (CNN) and SV models for soybean yield prediction in the United States. Their results showed significant superiority compared to linear traditional yield estimation methods. Khaki and Wang (2019) demonstrated prediction accuracy machine learning models based on satellite images for the maize yield, which showed that RFM had more significant outperformance than SV and DT models. Peerlinck et al. (2019) demonstrated that the Gradient Boost model provides more accurate yield prediction compared to other Booster models. Shahhosseini et al. (2021) applied linear regression, LASSO, LightGB, RF, and XGBoost models to estimate corn yield in the US. Their results indicate that the XGBoost hybrid model is more powerful for crop yield prediction than other fundamental linear and non-linear models.

It should be noted that these existing studies were restricted to utilising a single variable or considering linear relations between yield and variables, which underutilises the full potential of machine learning models and fails to capture the complex relationships and interactions which play a crucial role in crop yield estimation (Kiran Kumar et al. 2023; Krishnadoss and Ramasamy 2024). More holistic indicators should take into account a wide range of various influencing factors, such as physical parameters and spectral information (Sun et al. 2020). Vegetation indices, wheat height, solar radiation, and soil properties are important parameters for a comprehensive understanding of crop yield (Khaki and Wang 2019; Mateo-Sanchis et al. 2019). Furthermore, combining crop physical parameters, soil property, solar radiation, and VIs can improve the model quality because each of these indicators has unique biophysical information which can minimise various errors or complement insufficient datasets (Chen et al. 2020; Semeraro et al. 2019; Xu et al. 2019).

However, including a wide range of parameters can increase the complexity of the model and may potentially reduce model metrics (Cai et al. 2019; Mwaura and Kenduiywo 2021) and also raises the potential risk of increasing model errors (Greitzer et al. 2012; Marsh et al. 2014). Previous machine learning models have applied the Importance Method to determine the importance of predictor variables. However, this method is limited to comprehensive datasets, which include data from different levels such as regional, district, and farmer (Altmann et al. 2010; Strobl et al. 2008). Moreover, this method is primarily designed to analyse numerical features on a global scale and may not be effective for categorical features. Jones et al. (2022) indicated that the feature importance method might not capture complex relationships, interactions, or non-linear effects and may not completely represent all the nuances impact of each variable's impact on final results. The SHapley Additive exPlanations (SHAP) approach, in contrast, is a deeper analysis of the contribution of all input variables to the final results, which ensures that each player or feature has obtained their actual share of the total contribution (Kim and Kim 2022; Remman and Lekkas 2021; Shapley 1953). Compared to other methods, the SHAP approach assesses the contribution of each variable to the final outcome based not only on the presence or absence of influence of the variable, but also on local accuracy and the stable effect of the variable across different time intervals (De Filippi et al. 2020; Remman and Lekkas 2021).

Most crop yield estimations use traditional machine learning models with default parameters (Krishnadoss and Ramasamy 2024), which does not allow for maximis their performance and, therefore, are suboptimal solutions for specific complex models (Kiran Kumar et al. 2023), such as a yield estimation model. Machine learning-based crop yield estimation models often consist of complex non-linear variables (Ajith et al. 2024), and default parameters may not be capable of fully capturing these interactions. For example, soil properties and solar radiation variables often have non-linear effects on crop yields (Camargo-Alvarez et al. 2023; Kigo et al. 2023). Without optimising machine learning parameters, it is impossible to capture all these linear and non-linear effects. Hyperparameter tuning supports optimising machine learning models and their predictive power for yield estimation. These tunings are done by finding the optimal number of trees in the ensemble, maximising the node depth of each tree and the weights at each iteration (Bischl et al. 2023). Moreover, optimising a model with hyperparameter tuning can reduce overfitting problems for a specific training dataset.

Within the existing research, we identify the following research gaps: First, very few studies applied the SHAP method for crop prediction models which used machine learning models with default parameters (Srivastava et al. 2021; Zhu et al. 2021). These existing researches use machine learning models with default parameters, and their models may suffer from overfitting problems because they are not optimised to the specific characteristics of the dataset. Furthermore, these studies used meteorological parameters, including temperature, precipitation and humidity for crop yield modelling. However, as previous studies have shown, in many cases, meteorological datasets contain noise that might lead to either over- or underfitting of the data, which affects model performance (Kigo et al. 2023; Kiran Kumar et al. 2023). In this circumstance, several studies demonstrated that isolated meteorological variables could not adequately represent crop yields (Chen et al. 2020; Semeraro et al. 2019). There is a general lack of studies optimising the performance of machine learning models for wheat yield estimation using comprehensive datasets such as vegetation indices, wheat height, solar radiation, and soil properties to estimate wheat yield.

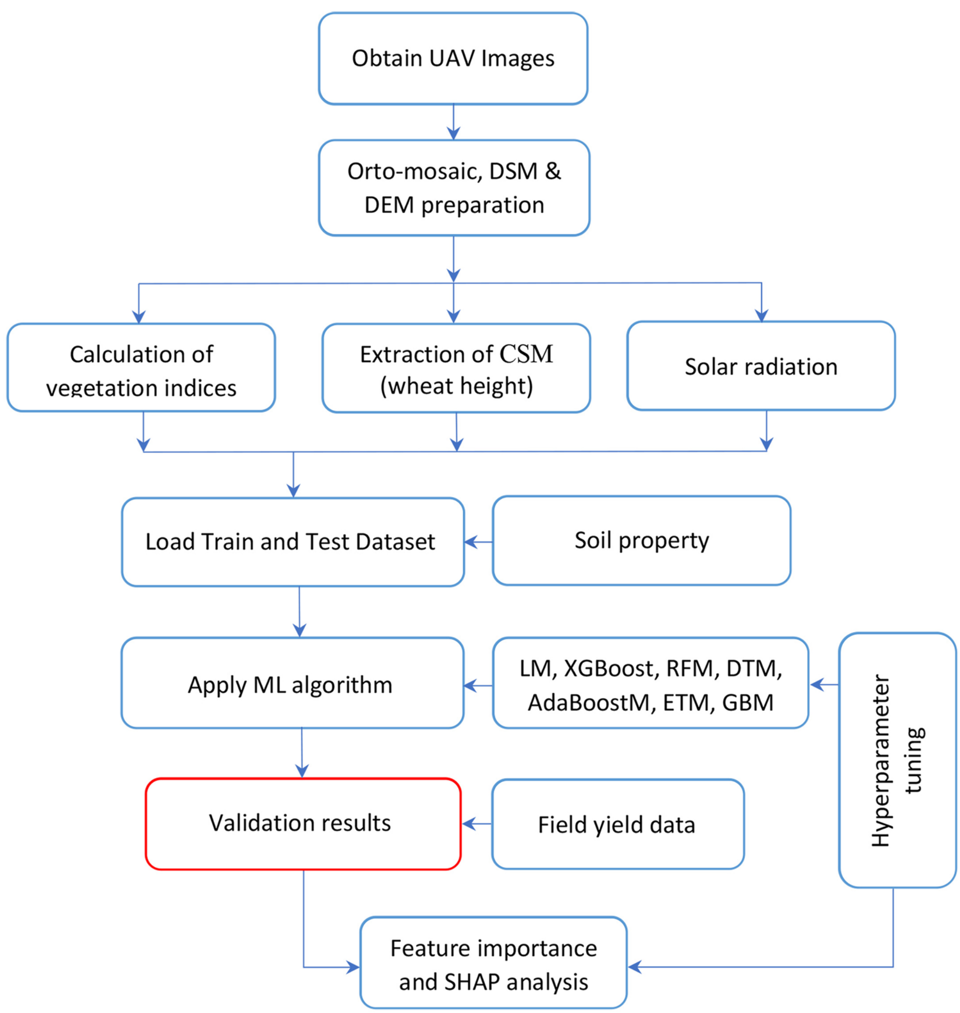
To close this gap, this study aims at identifying an optimising machine learning algorithm for wheat yield estimation using various combinations of input variables to increase the efficiency of yield estimation. This aim is approached by optimising different machine learning algorithms and comparing their performance in terms of predicting true yields. From this perspective, the main contributions of this research are outlined below: First, we utilise a multi-indicator approach by testing a large number of parameters such as crop height, vegetation indices, solar radiation, and soil properties along with machine learning models. Second, we optimise traditional machine learning models using hyperparameter tuning to improve model accuracy and reduce overfitting or underfitting of datasets. Third, we identify optimal input variables for redundancy elimination from the input datasets and calculate the relative significance of variables based on the SHAP approach.

Materials and methods

Analytical framework

The estimation procedure is implemented in several steps. In the first step, we integrated comprehensive datasets to estimate wheat yield at the farm level based on optimised machine learning models. The detailed process of analysis is laid out in the flowchart of Fig. 1. In the first stage, the vegetation indices, wheat height, and solar radiation data were collected through UAV from farmers' fields twice, on 20 May and 16 June 2020 (see the next section on more details). After obtaining the necessary ultra-high-resolution multispectral images of the study areas, the next task is ortho-mosaicking of the raw images using Agisoft PhotoScan Pro software and preparing digital elevation models (DEM) and digital surface models (DSM). Ready ortho-mosaic DEM and DSM data for study areas were used to calculate wheat height parameters for both periods (Fig. 3). The solar radiation parameters for both periods

Fig. 1 Flowchart of analytical tasks



were calculated in the ArcGIS 10.8 software, while information on soil properties was obtained from the Bavarian State Office for Environment at the district/province level.

Overall, we tested the suitability of the Normalized Difference Vegetation Index (NDVI), Normalized Difference Red Edge index (NDRE), Chlorophyll red-edge Index (CIred-edge), Simplified Canopy Chlorophyll Content Index (SCCCI), Green Chlorophyll Vegetation Index GCVI and Enhanced Vegetation Index (EVI2), combined with wheat height, solar radiation, and soil properties indicators, to estimate wheat yield. After that, we examined the potential of all input variables for crop yield estimation, applied six machine learning models such as XGBoost, RF, DT, AdaBoost, ET, and GB models as well as compare their performance. In the next stage, all datasets were randomly split into a 30% testing set and a 70% training set. Based on evaluation metrics of Root Mean Square Error (RMSE), R squared (R2), Mean Absolute Error (MAE), and Mean Squared Error (MSE), we assessed the performance of XGBoost, RF, DT, AdaBoost, ET, and GB machine learning models. In the final stage, we identify the optimal input variables based on the SHAP model and optimise the performance of machine learning models using hyperparameter tuning.

The aforementioned steps are illustrated in more detail in the following flowchart (Fig. 1), and each flowchart step is described in more detail in the next sections.

UAV data collection

UAV data for this study was collected on several farms in the Southern part of Germany in the scope of this study. The total study area of the three plots was approximately 30 hectares, with planted wheat belonging to the same phenotype. Farmers planted wheat in February–March 2020. Yield data, as collected by the combine harvester in July–August 2020, and yield data from the combine harvester were used as actual ('true') yield data. A total of 2548 polygons were utilised in the research, the range of wheat yield values ranging from 0.6 to 12.5 t/ha across the various polygons. The farmer's field was split into many small zones for more in-depth analysis, which can significantly improve the efficiency of the yield estimation method (Fig. 2).

Figure 2 illustrates one of the study fields, which is split into many small plots/cells based on ground yield data. The red line in Fig. 2 represents the wheat crop area, and the yellow rectangles show individual plots.

Extracting wheat height

All crop data in the study areas were collected using a MicaSense multispectral camera mounted on our UAV. The images have a 3D ultra-high resolution of 4.5 cm. Our UAV's multispectral camera has an extra GPS sensor, which allows us to measure the wheat crop height more

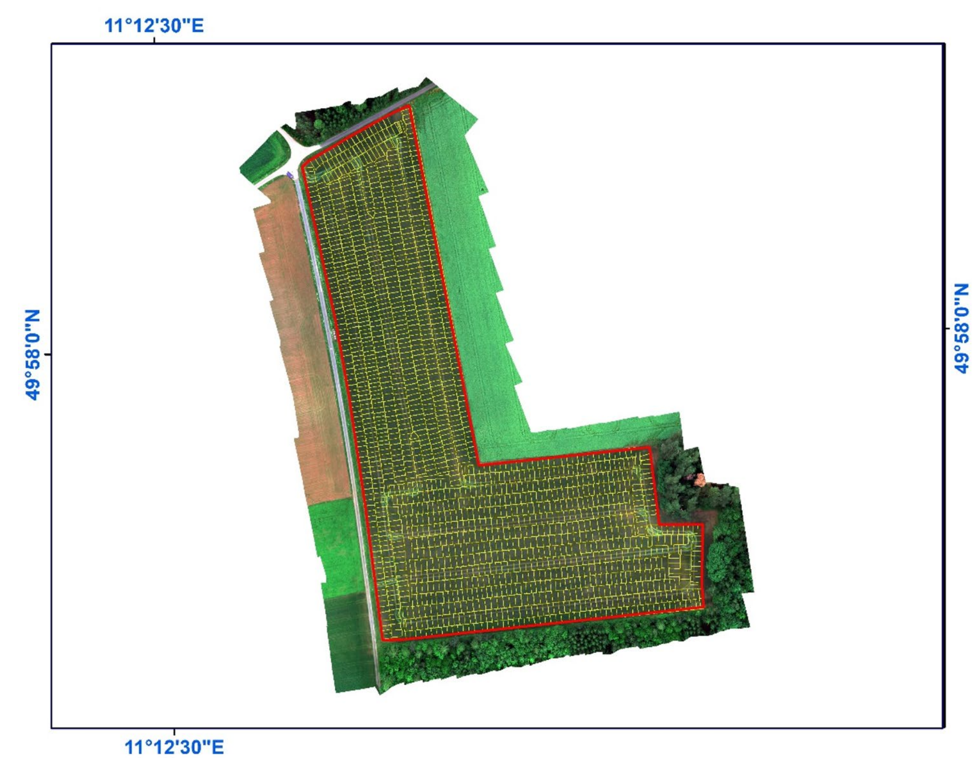


Fig. 2 RGB-band spectral image from the study area at 4.5 cm resolution

accurately and is essential for performing ortho-mosaics to create an accurate 3D shape/form of the fields. The ultra-high-resolution 3D multispectral images from UAV were acquired in the study areas on 20 May and 16 June 2020. All collected 3D multispectral orthophotos were processed using Agisoft Photoscan Pro software to extract the digital elevation model (DEM), digital surface model (DSM), and vegetation indices.

To extract wheat height parameters, a UAV was launched before sowing and after harvest. Both DEM and DSM data were obtained from survey farm fields and processed using Agisoft Metashape software. The absolute wheat height variable or crop height parameter is calculated by subtracting the DEM from each DSM for the two periods, 20 May and 16 June 2020, during the wheat growing season (Fig. 3).

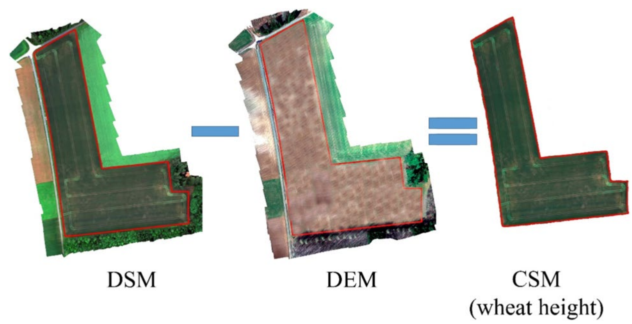


Fig. 3 Extracting wheat height from UAV imagery

Solar radiation

Solar radiation is the essential indicator that directly influences photosynthetic activity, growth, and crop yields (Holzman et al. 2018; Zhang et al. 2010). A few previous yield estimation studies (Holzman et al. 2018; Zhang et al. 2010) demonstrated that solar radiation can provide valuable insight for a comprehensive understanding of crop yield and help to optimise the crop yield estimation model. This parameter was calculated for all study fields based on a digital surface model for 20 May and 16 June 2020 using the Solar Radiation tool of ArcGIS spatial analyst toolbox. This tool allows us to calculate diffuse, reflected radiation and beam, as well as the shadow effect caused by terrain features. Furthermore, this tool takes many important factors affecting solar radiation into account, such as the slope, diffusion, and direct and global solar radiation.

Soil properties

Soil properties provide key insights into nutrient availability, water retention, and root development (Ransom et al. 2019; Tripathi et al. 2022), which directly affect plant health. For all study fields, information on soil properties was obtained from the Bavarian State Office for the Environment. In total, the research fields have seven soil types and they vary within the fields.

Vegetation indices

In recent years, most studies to estimate crop yields have utilised vegetation indices such as the Normalized Difference Vegetation Index (NDVI), Green Chlorophyll Vegetation Index (GCVI), Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), Normalized Difference Water Index (NDWI), Normalized Different Red-Edge (NDRE) index, Simplified Canopy Chlorophyll Content Index (SCCCI), and Chlorophyll red-edge (CIred-edge) index (Ashapure et al. 2020; Bendig et al. 2015). Each of the selected VIs has been found to add important information on crop conditions and contributing to more accurate yield estimations (Vallentin et al. 2022). However, the selection of an effective vegetation index remains a challenge due to the requirement to take into account environmental conditions and crop characteristics (Xu et al. 2021). Therefore, several vegetation indices were selected for a comprehensive understanding of crop health. For instance, the SAVI is more effective in monitoring crops with low vegetation density (Xu et al. 2019), whereas NDWI is more effective for measuring water stress in crops as well as for identifying flooded agricultural lands (Chen et al. 2020; Semeraro et al. 2019). The NDVI is widely utilised for crop classification and crop yield estimation as well as for measuring the density of the vegetation captured (Semeraro et al. 2019; Tucker 1979). Liu and Huete (1995) showed that EVI is sensitive to identifying crop yield changes at different crop growth stages, which is essential for crop yield estimation. Furthermore, NDRE, SCCCI, GCVI, and CIred-edge are widely applied to estimate canopy chlorophyll and nitrogen content, which directly impacts crop health and crop yield (Barnes et al. 2000; Gitelson and Merzlyak 1994; Gitelson et al. 2003a). An overview of the selected indices is provided in Table 1. The examination of different vegetation indices helps us to comprehensively understand which index is more appropriate to integrate with wheat height, solar radiation, and soil properties, as well as improves the accuracy of wheat yield estimation.

Model optimization

The analysis was conducted using machine learning libraries in Python 3.9, which is an optimal tool for data processing, modelling and evaluation. To examine whether optimisation of machine learning algorithms significantly affects the accuracy of wheat yield estimation, we analysed both cases without and with optimisation. The performance of XGBoost, RF, DT, AdaBoost, ET, and GB machine learning

Table 1 Vegetation indices equations

Vegetation Index	Equation	Reference
Normalized difference vegetation index (NDVI)	$NDVI = \frac{P_{NIR}-P_{RED}}{P_{NIR}+P_{RED}}$ (1)	(Tucker 1979)
Normalized difference red edge index (NDRE)	$NDRE = \frac{P_{NIR}-P_{Red-Edge}}{P_{NIR}+P_{Red-Edge}}$ (2)	(Gitelson and Merzlyak 1994)
Chlorophyll red-edge index (CIred-edge)	$CI_{red-edge} = \frac{P_{NIR}}{P_{Red-Edge}} - 1$ (3)	(Gitelson et al. 2003a)
Simplified canopy chlorophyll content index (SCCCI)	$SCCCI = \frac{\frac{P_{NIR}-P_{Red-Edge}}{P_{NIR}+P_{Red-Edge}}}{\frac{P_{NIR}-P_{Red}}{P_{NIR}+P_{Red}}}$ (4)	(Barnes et al. 2000)
Green chlorophyll vegetation index (GCVI)	$GCVI = \frac{P_{NIR}-P_{Green}}{P_{NIR}+P_{Green}}$ (5)	(Gitelson et al. 2003b)
Enhanced vegetation index 2 (EVI2)	$EVI2 = 2.4\frac{P_{NIR}-P_{RED}}{P_{NIR}+P_{RED}+1}$ (6)	(Jiang et al. 2008)
Soil-Adjusted Vegetation Index (SAVI)	$SAVI = \frac{P_{NIR}-P_{RED}}{P_{NIR}+P_{RED}+0.16} * (1 + 0.16)$ (7)	(Huete 1988)
Normalized Difference Water Index (NDWI)	$NDWI = \frac{P_{Green}-P_{NIR}}{P_{Green}+P_{NIR}}$ (8)	(McFeeters 1996)
Normalized Difference Water Index – Blue (NDWI-B)2	$NDWI - B = \frac{P_{Blue}-P_{NIR}}{P_{Blue}+P_{NIR}}$ (9)	(Wei et al. 2011)

Table 2 Equations of the evaluation metrics

Name	Equation
R squared (R2)	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}$ (10)
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$ (11)
Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert}{n}$ (12)
Mean Squared Error (MSE)	$MSE = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}$ (13)

algorithms was evaluated utilising Root Mean Square Error (RMSE), R squared (R2), Mean Absolute Error (MAE) and Mean Squared Error (MSE) (Table 2). The goodness of fit of the machine learning models based on the R2 score, where the value ranges from 0 to 1, and a higher value ≥ 0.70 is considered strong, explaining all the variability in the dependent variable (actual wheat yield) (Huang et al. 2014). The RMSE calculates the discrepancy between modelled and observed data values (Zacharias et al. 1996), while the MAE indicates the average magnitude of errors between two continuous variables without taking into account their direction (Allen 1971). Moreover, another model evaluator metric is MSE, which measures the average of the squares of model errors (Jacovides & Kontoyiannis 1995). MSE and MAE are appropriate metrics for evaluating prediction accuracy or error and comparing the performance of machine learning models. The values of RMSE and MAE range from 0 to 1, with smaller values indicating a better fit between model predictions and actual yields. A value of 0.4 is often considered a critical threshold for a sufficient model fit (Oikonomidis et al. 2022).

In order to conduct a comprehensive evaluation of the performance of machine learning models, we utilised non-parametric statistical tests, as proposed by (Demšar 2006), as non-parametric statistical tests are considered more appropriate and reliable than parametric tests for statistical comparisons of different machine learning models (Feng et al. 2019; García et al. 2015). We selected the Friedman test, which evaluates statistical differences across various models based on their ranks (Friedman 1937). In case the Friedman test rejected the null hypothesis of equal performance between methods (Nemenyi 1963), a Nemenyi post-hoc test can be conducted. The Nemenyi test is mainly used to assess whether differences between models are indeed statistically significant or due to randomness (Kamir et al. 2020). Based on the test results, a critical distance graph is plotted, which reveals the performance rank of each machine learning model and the critical distance.

Another important aspect of developing an accurate yield estimation model is identifying the optimal input variables to avoid redundancy, which can introduce noise and reduce the model's performance. The SHAP method evaluates the importance or contribution of an individual player to the collaborative work of the command (Jones et al. 2022; Shapley 1953; Srivastava et al. 2022). In the machine learning case, the "games" are the different models, such as XGBoost, Random Forest, and Decision Tree models, while the "players" are all input variables. As the input data increases, the number of variations of input variables increases exponentially (Lundberg and Lee 2017). Therefore, marginal effects from the predicted values become more computationally complex (Shendryk et al. 2021). In this respect, the SHAP method can provide a unique solution based on fair share or reasonable reward for each input variable/feature (Remman and Lekkas 2021).The contribution of each feature to the final model output is assessed by natural properties such as symmetry, additivity, and efficiency (Rodríguez-Pérez & Bajorath 2020).

The SHAP feature importance equation is given below:

$$\varphi_i = \frac{1}{|N|!} \sum_{S \subseteq N \setminus \{i\}} |S|!(|N| - |S| - 1)! \left[f(S \cup \{i\} - f(S) \right] \quad (14)$$

Here, i is the conditional value of SHAP, whereas $f(S)$ corresponds to the output of the machine learning model to be explained using a set S of features, N is the full set of all features, and $i(\varphi_i)$ is defined as the average value of its input over all possible rearrangements of the feature dataset.

Results

The results obtained from machine learning models for wheat yield estimation are summarised in Table 3. We observe that the RF and XGBoost models outperformed other examined machine learning models, demonstrating a high $R2 \geq 0.87$ with a low $RMSE \geq 0.27$ and $MAE \geq 0.19$. GB and LM models demonstrated a medium performance (Table 3, $R2 \geq 0.68$, $RMSE \geq 0.63$ and $MAE \geq 0.46$), while the AdaBoost model demonstrated slightly lower performance (Table 3, $R2 = 0.62$, $RMSE = 0.73$ and $MAE = 0.57$).

Optimisation machine learning algorithms based on hyperparameter tuning exhibited more accurate predictions than traditional machine learning algorithms with default parameters (Table 3). These results demonstrated that taking into account the non-linear effects between variables can improve the accuracy and performance of the model. Table 3 reveals that among machine learning models RF and XGBoost results perform better (Table 3, $R2 \geq 0.89$, $RMSE \geq 0.25$ and $MAE \geq 0.18$) than other machine learning (Table 3, $R2 \geq 0.66$, $RMSE \geq 0.67$ and $MAE \geq 0.49$). The DT and ET models showed interestingly low results in both cases for the test dataset (Table 3). However, the evaluation metrics MAE and RMSE could not be calculated for these models due to certain limitations or challenges encountered during the estimation process. One of the potential reasons for the weak performance of DT and ET models may be their ensemble learning mechanism, which is very sensitive to noise in the data, especially when the data consists of various physical and spectral information (Cao et al. 2021; Oikonomidis et al. 2022; Sun et al. 2020). In other words, small fluctuations or inconsistencies in the data arising from the combination of qualitative and quantitative variables with different scales can have a considerable effect on the accuracy of the model.

As can be seen in Table 3, optimisation of the machine learning algorithm leads to an approximation of the R2 values between the training and test datasets, which indicates better generalisation of the model and control of overfitting. The improvement in the performance of machine learning models after hyperparameter tuning is associated with the fact that during the tuning process, the machine finds the optimal parameters for specific datasets and correcting parameters such as learning rate, maximum depth, and number of estimators. In both cases, the RF and XGBoost models demonstrated superiority over the GB and AdaBoost models on all evaluation metrics for the testing, training, and full datasets alike (Table 3). The potential cause of the higher performance of the RF and XGBoost models is that each node tries to enhance the performance of the previous node in the tree as well as all nodes in the tree learn the errors of the previous nodes and try to correct their mistakes of the next nodes.

As an additional test for goodness-of-fit, we generated prediction error plots for each machine-learning model (see Supplementary Figure S1). The prediction error plots of XGBoost, RFM, GBM, and AdaBoost models demonstrate different high and low values but still present adequate results for predicting wheat yield with $R2 > 0.63$. The RF and XGBoost prediction error plots illustrate that the predicted points are very close to the line of best fit. According to the results of all evaluation metrics, we can assert that the best-performing models for wheat yield estimation are the optimised RF and XGBoost models (Table 3).

Furthermore, we are interested in the goodness-of-fit across the actual yield distribution. Figure 4 illustrates that the prediction accuracy for yields ≥ 4.3 t/ha is higher compared to that for low yields.

Previous yield estimation models found RMSE and MAE found 0.58 and 0.49 values from 0.6 to 1 (Kogan

Table 3 Machine learning models results of R2, MAE and RMSE

Model name	R2 Test	R2 Train	Full	MAE	RMSE	R2 Test	R2 Train	Full	MAE	RMSE
	Traditional					Optimized				
LM	0.70	0.71	0.70	0.50	0.68	0.70	0.71	0.70	0.50	0.68
XGBoost Model	0.71	0.93	0.87	0.23	0.32	0.74	0.94	0.87	0.21	0.31
RF Model	0.71	0.95	0.88	0.19	0.27	0.75	0.95	0.89	0.18	0.25
DT Model	0.44	1.00	0.83	0.00	0.00	0.45	1.00	0.83	0.00	0.00
AdaBoost Model	0.62	0.67	0.65	0.57	0.73	0.65	0.68	0.66	0.55	0.71
ET Model	0.69	1.00	0.91	0.00	0.00	0.70	1.00	0.91	0.00	0.00
GB Model	0.69	0.74	0.73	0.47	0.64	0.71	0.75	0.74	0.45	0.62

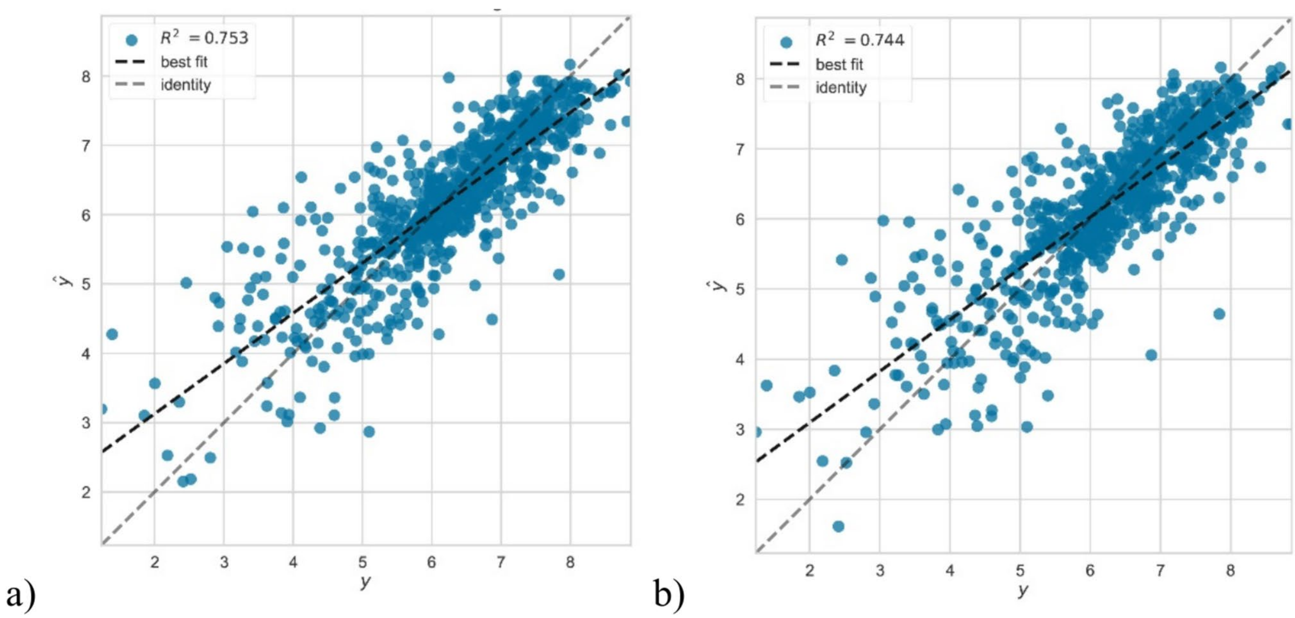


Fig. 4 Scatter plot predicted wheat yield with actual yield a) Random Forest and b) XGBoost

et al. 2013; Naderloo et al. 2012), while our optimised RF and XGBoost models demonstrated RMSE results of 0.25 and MAE of 0.18, which means that optimising machine learning algorithms through hyperparameter tuning can improve model accuracy and performance.

In line with the findings of Elavarasan et al. (2020), residual plots of machine learning models can reveal how effectively a model is capable of predicting crop yields. Figure 5 demonstrates the residual plots of RF and XGBoost models, with residual points concentrating around zero. Moreover, the residual plots of the RF and XGBoost models indicate that no serious deviations occurred except for a few outliers, all residuals of the models being randomly distributed. Also, this test confirms that RF and XGBoost can produce predictions with a close fit to actual wheat yields.

Multiple machine learning algorithms were examined with and without adjustment for multicollinearity. The accuracy before and after removing collinear variables from machine learning models did not exceed 2%, which means that machine learning algorithms can be highly robust even when facing collinearity issues (Supplementary Table S1). However, the accuracy before and after removing collinear variables from the linear model exceeded 5%, which means that the multicollinearity effect has a higher influence on the accuracy of the linear model than on nonlinear machine learning models. These results demonstrate that machine learning algorithms can provide more reliable and robust results than traditional statistical models such as linear regression, which is consistent with the results of previous studies (Arabameri et al. 2020; Ransom et al. 2019).

To examine whether considered machine learning models are statistically different and statistically significantly different from each other, we applied the Friedman test. The test results indicate that the p-value is 0.002, which rejects the null hypothesis, which means all machine learning models are significantly different from each other. The Nemenyi post-hoc compares the performance of the machine learning models with respect to their accuracy in predicting wheat yield (Fig. 6). Based on the test results, all machine learning models were divided into two groups, the first group incorporates the highest ranked RF, XGBoost, GB, and LM models and the second group includes low ranked AdaBoost, ET, and DT models. In this post-hoc, the RF and XGBoost models demonstrated a higher average ranking compared to other models, confirming them as the most reliable machine learning models for wheat yield estimation and prediction within our dataset.

To better understand the distribution of the MSE and to gain insight into its central tendency, spread, and potential outliers, we generated negative mean squared error boxplots for each optimised machine learning algorithm (Fig. 7). High dispersion, i.e. low performance, was observed for the DT model, while XGBoost and RF machine learning models had low dispersion, i.e. superior performance.

The results of the SHAP analysis on all optimised machine learning models (see Supplementary Figure S2) are summarised within the SHAP plot in Fig. 8, ordering all input variables according to their degree of explanatory power. The vertical axis in the graph of Fig. 8 represents all input variables (vegetation indices, solar radiation, wheat height and soil properties) for both periods,

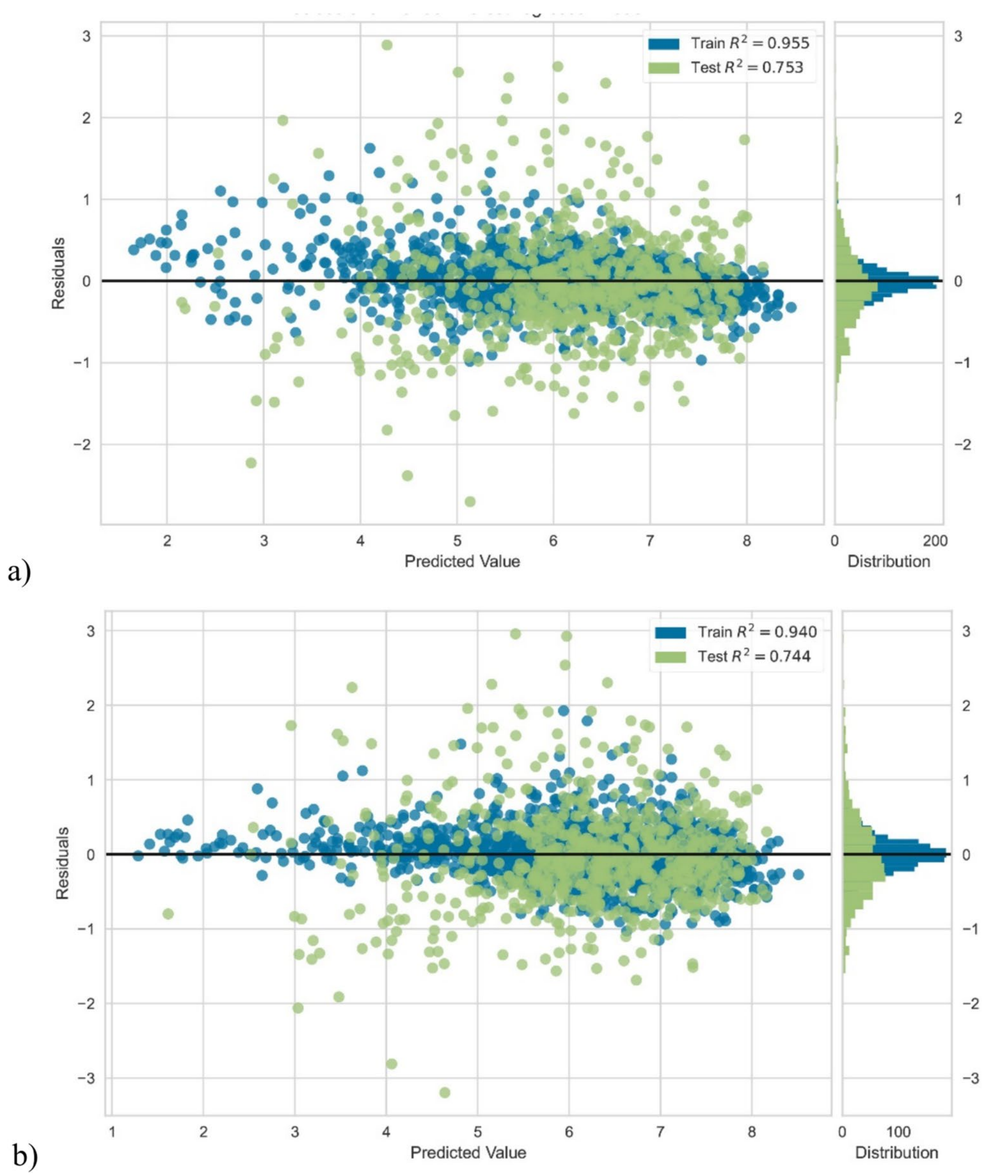


Fig. 5 Residual plots of models for a) Random Forest and b) XGBoost

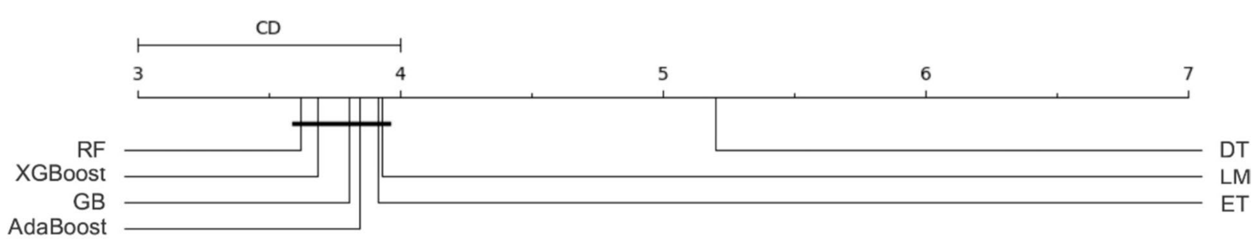


Fig. 6 Nemenyi critical distance diagram for various ML models

while the horizontal axis refers to SHAP values. Each dot on the plot refers to one observation point in the dataset. The SHAP approach revealed that the variables NDRE, CIred-edge, and CSM (wheat height) have high importance for wheat yield, while solar radiation and soil properties

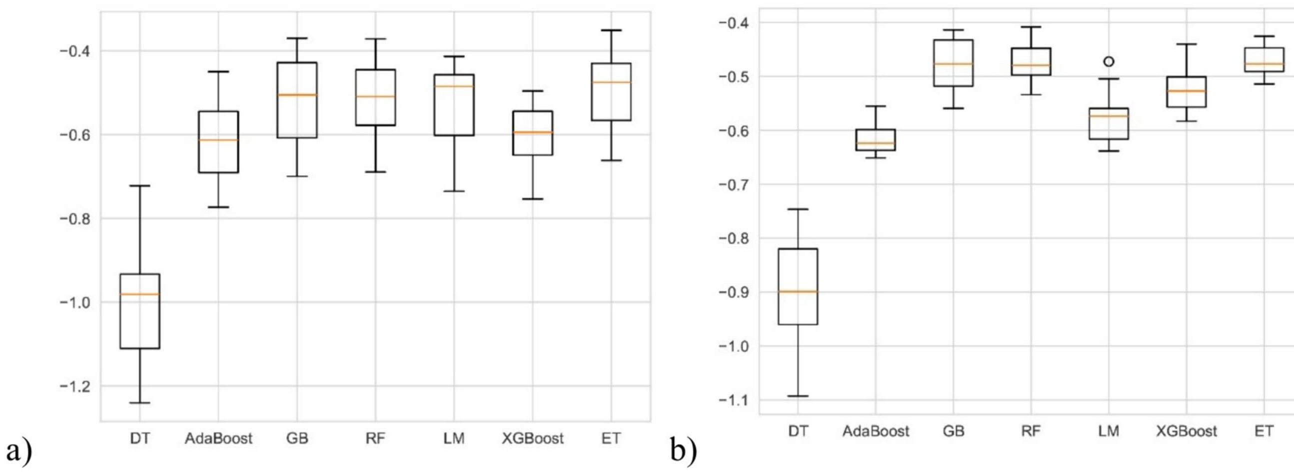


Fig. 7 Negative mean squared error (MSE) plots for various ML models a) testing dataset and b) training dataset

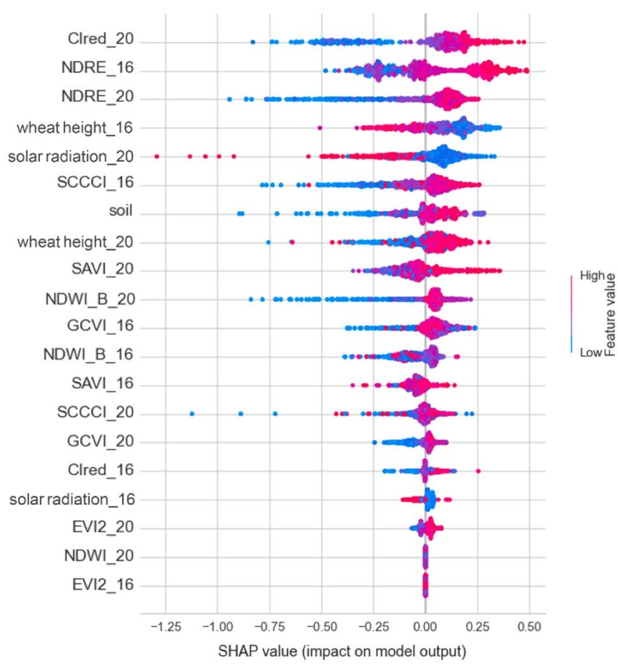


Fig. 8 RF model SHAP plot

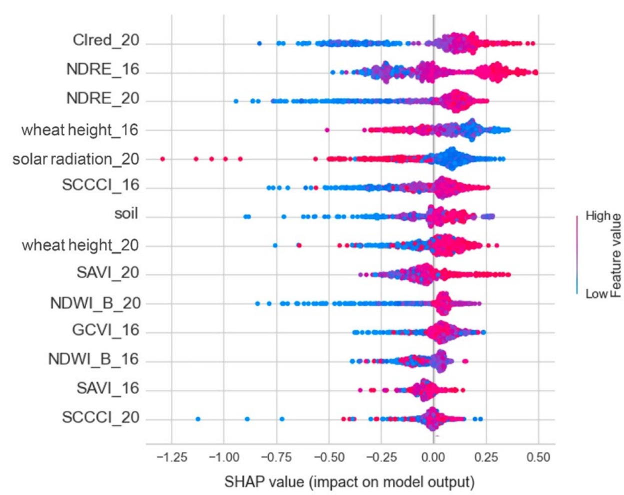


Fig. 9 RF model SHAP plot after exclusion of variables

have medium importance, but the variables NDVI, NDWI, and SAVI had low importance (Fig. 8).

Variables with positive SHAP values can increase the accuracy of wheat yield prediction, while variables with negative SHAP values can decrease the accuracy of the model. Within the lower part of the SHAP plot, we find variables with the lowest contribution to overall explanatory power, such as NDWI, EVI2 and SAVI vegetation indices.

According to the finding of our first SHAP analysis, we dropped variables located at the bottom of the SHAP plot (EVI2_16, NDWI_20, EVI2_20, solar radiation_16, CIred_16 and GCVI_20). Figure 9 shows the SHAP plot after the exclusion of these variables. As the last step, we re-ran all optimised machine learning models to provide a comprehensive assessment of the stability of our model.

According to previous works (Liakos et al. 2018; Srivastava et al. 2021), a stable prediction model should be robust to exclusion of individual input variables. As shown in Table 4, our changes in the input variables did not have major effect on our model results.

Table 4 Machine learning model results after SHAP analysis

Model name	R2 Testing	R2 Training	R2 Full	MAE	RMSE
LM	0.68	0.68	0.68	0.51	0.70
XGBoost Model	0.72	0.92	0.86	0.22	0.32
RF Model	0.73	0.93	0.88	0.18	0.26
DT Model	0.43	1.00	0.82	0.00	0.00
AdaBoost Model	0.64	0.67	0.66	0.57	0.72
ET Model	0.70	1.00	0.89	0.00	0.00
GB Model	0.70	0.72	0.73	0.46	0.63

Discussion

In this study, we present a machine learning based wheat yield estimation model using comprehensive UAV datasets with implementation of hyperparameter tuning to improve model performance. Furthermore, in this research, we utilised various UAV datasets and qualitative data (soil properties), which significantly improved the accuracy of the model (Ashapure et al. 2020; Khoshroo et al. 2018). Comprehensive datasets such as vegetation indices, wheat height, solar radiation, and soil properties were used as the input variables for machine learning models.

Our optimised machine learning algorithms using hyperparameter tuning demonstrated higher performance compared to traditional machine learning with default parameters. Among the optimised machine learning models, the best performance was observed for RF and XGBoost models, which outperformed other examined machine learning models. The comparison results indicate that the optimised machine learning model with comprehensive datasets incorporating physical and spectral information are an effective way to estimate wheat yield. The findings revealed that traditional machine learning models with default parameters are constrained in their ability to accurately estimate yields when analysing complex, non-linear variables, which is consistent with the results of the studies (Bischl et al. 2023; Kiran Kumar et al. 2023). Many crop yield estimation studies (Kastens et al. 2005; Nebiker et al. 2016; Panday et al. 2020) utilised various ground-based sensors and ground-based data to obtain comprehensive datasets, which is costly to replicate at a larger scale. Meanwhile, our approach demonstrated the possibility of obtaining comprehensive datasets based on UAVs, which is cost-effective and capable of collecting highly accurate data. The results reveal that machine-learning-based yield estimation using high-quality UAV data can provide accurate yield estimation, which can be a useful tool for agriculture systems, especially in low-income or developing countries. Moreover, compared to previous machine-learning-based yield estimation approaches achieving an R2 coefficient of 0.65–0.75 (Ashapure et al. 2020; Jones et al. 2022; Srivastava et al. 2022) compared to our approach R2 coefficient of 0.70–0.89. Our results revealed that the measured wheat yield based on the comprehensive data was an excellent fit with the actual yield, which indicates the importance of proper model selection and the optimisation of model parameters.

All the evaluation metrics of MSE and RMSE confirmed that RF and XGBoost machine learning models are the most efficient algorithms for all training, testing, and full datasets. The model comparison based on the Nemenyi post-hoc test demonstrated that the average ranking of RF and XGBoost machine learning models was higher than other machine learning models; oppositely, the linear model revealed a lower average ranking. Furthermore, we compared all optimised machine learning algorithms based on the negative MSE of the box plot, and results revealed that linear model, extra tree, and decision tree models have a negative MSE with a very high mean deviation, which indicates a lower accuracy of these models compared to the other examined models. The RF and XGBoost models revealed opposite results, the negative MSE values of the box plots showing very close mean deviation, indicating their higher predictive ability compared to the other machine learning models. We also observed a considerable difference in yield estimates between the machine learning (RF and XGBoost) models and the linear model. However, several previous studies have demonstrated that non-linear models or machine learning models do not have significantly different advantages from linear models, on the contrary, they have difficulties with the algorithm (Aghighi et al. 2018; Johnson 2016; Kamir et al. 2020). This might be caused by many factors, such as input variables, machine learning algorithms were not optimised for the input datasets of the model, the number of training and test sets. Each of these listed factors plays a major role in the accuracy of the model and directly affects the final results, for instance, selecting the correct input variable is a critical point at the initial stage of modelling because irrelevant input variables can lead to inefficient utilisation of resources. On the other hand, the correct selection of the number of training and test data, as well as the appropriate estimating period are crucial factors in obtaining reliable/trustworthy machine learning results. Improper taking into account the listed parameters/factors may lead to biased and unpredictable modelling results as well as can be misleading in solving the research problem (Shendryk et al. 2021; Xu et al. 2021).

This study also identified the importance of each variable based on the SHAP approach for all machine learning models. All the input variables impact wheat yield estimates, some greater and others lesser degree. However, only the variables NDRE, CIred-edge vegetation indices, wheat height (CSM), solar radiation, and soil properties were observed to be of high prediction power in both periods in the considered region. These NDRE and CIred-edge vegetation indices are calculated based on the red-edge spectrum, which is sensitive to changes in chlorophyll concentration in the crop. Chlorophyll concentration is the main parameter demonstrating the level of crop health and photosynthetic activity in crops (Gitelson & Merzlyak 1994). The SHAP results also indicated that SAVI, NDWI, and NDVI for both periods of 20 May and 16 June had a very low impact on wheat yield estimates. These indices may be related to the utilisation of the red spectrum, which is more suitable for identifying green biomass and vegetation density but less effective for estimating chlorophyll content concentration in wheat crops (Xu et al. 2019). In our case NDRE, CIred-edge vegetation indices and wheat height (CSM) variables demonstrated a higher contribution to the final model output than solar radiation, soil properties, and other vegetation indices (Fig. 8). In other words, solar radiation, soil properties, and other vegetation indices predict less variation than NDRE, CIred-edge, and wheat height (CSM) indicators. Identifying the most important vegetation indices and variables for crop yield estimation can help farmers, downstream value-chain actors, and decision-makers avoid unnecessary costs and labour for data collection and risk management (Ashapure et al. 2020; Khoshroo et al. 2018). Furthermore, SHAP analysis can suggest which input variables have less contribution to final results and might be removed from the model to optimise machine learning analysis time. Furthermore, we found that the accuracy of our model demonstrates the same robust results after removing variables, which were found to have less impact on the prediction accuracy of the SHAP analysis (Table 4).

Accurate wheat yield estimation can be very helpful in farm management. This information could be used for the optimisation of crop growth and manage climate risks as well as developing agricultural insurance products. For instance, accurate monitoring and forecasting of crop yields can increase farmers crop insurance enrolment by 10–15%, which is an important factor in farmers' financial security (Badani et al. 2020). Furthermore, improving the accuracy of yield prediction by 5% can reduce farmers' input costs by up to 10%, especially reducing water and fertiliser costs (Lobell et al. 2009; Wang et al. 2021). This optimised machine learning algorithm using comprehensive datasets is focused on wheat yield estimation, but it can be easily modified for other crops with changes in the specificity of the input attributes. Moreover, this model is one of the simple, inexpensive, and very effective approaches for wheat yield estimation and prediction based on an optimised machine learning algorithm.

The presented approach is flexible and easier data processing than analysing satellite images (e.g., WorldView and GeoEye) (Whitehead and Hugenholtz 2014). Over the past two to three years, many companies have produced low-cost ($5000–7000 USD) UAV technologies (Lopatin and Poikonen 2023; Whitehead and Hugenholtz 2014), which are analogous to the UAVs we utilise in this approach. These low-cost UAVs can collect information on crop height, solar radiation, and crop conditions from multispectral imagery to analyse and monitor crops. Furthermore, UAV technology is multifunctional and more flexible than ground-based sensors, satellites, and other technologies, as well as easily upgradable. In many countries such as the USA, China, and Korea, the sharing method is practised where several farmers can buy one UAV (Lopatin and Poikonen 2023; Whitehead and Hugenholtz 2014) and use it instead of buying high-resolution satellite imagery for crop analysis/monitoring. Furthermore, a new generation of UAVs has a more automated flight system, image processing, and a simple interface that does not require extensive training or technical knowledge from farmers/users.

This approach revealed very good accuracy results (R2 coefficient of 0.89 and RMSE of 0.18) in wheat yield estimation by combining comprehensive datasets and optimised machine learning algorithms. However, several limitations should be considered in future works to improve the accuracy of this approach. Firstly, data collection was limited to one year, while multiple years of datasets and datasets from sowing to harvesting could have significantly improved the robustness of the model. For future research, we suggest the use of long-term datasets that include between-year variability to fully capture the effects of annual yield fluctuations. Secondly, we collected data from different farmers and soil properties, sample fields were all located in the same climatic zone. For future research, increasing the number of sample fields, including those from different climatic zones, as well as adding more sensors for UAVs, would be advantageous. Thirdly, acquiring additional parameters such as fertiliser amount, precipitation, and maximum and minimum temperature variables can improve machine learning capabilities and improve wheat yield estimation accuracy. Moreover, we utilised hyperparameter tuning techniques to improve the performance of machine learning models, but future research may consider cross validation or regularisation techniques.

Conclusions

In this study, we integrated comprehensive (complex) datasets for wheat yield estimation to minimise the risk of misinterpretation or overreliance on a single metric. The application of an optimised machine learning algorithm enabled us the overcoming of the limitations associated with individual variables and thereby facilitating a more robust estimate of wheat yield. Our approach can be valuable insight for the improvement of crop management and decision-making processes. The increased accuracy was achieved through hyperparameter tunning of machine learning algorithms and the use of comprehensive datasets obtained from UAVs. This model can potentially contribute to the development of methodologies for wheat yield prediction and field-scale crop monitoring. Our approach identified the impact of each input variable on yield, which is important information for farmers and agricultural scientists and can be useful in improving farm productivity and agricultural sustainability. Furthermore, our approach has demonstrated that optimised machine learning algorithms and comprehensive datasets may allow estimating wheat yields 7–12% more accurately than traditional models.

Findings from this research could be a valuable tool for the early prediction of wheat yields and might be used in governmental programs for sustainable agriculture development. If agriculture specialists can implement this or similar approaches into practice, there is a potential opportunity to decrease yield losses and increase farmers' incomes. We conclude that comprehensive datasets and optimised machine learning are promising methods for acquiring accurate estimates of actual yields at the farm and field scales.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40808-024-02188-9.

Acknowledgements The authors express their thanks to the team members of the project "Increasing Climate Resilience via Agricultural Insurance – Innovation Transfer for Sustainable Rural Development in Central Asia (KlimALEZ)", implemented by the Leibniz Institute for Agricultural Development in Transformation Economies (IAMO)."

Funding Open Access funding enabled and organized by Projekt DEAL. This research was supported by the German Federal Ministry of Education and Research (BMBF) [01LZ1705A] and WGL Forschungsverbünde [D0093/3].

Data availability The data are available upon request from the corresponding author.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Aghighi H, Azadbakht M, Ashourloo D, Shahrabi HS, Radiom S (2018) Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 OLI. IEEE J Selected Topics Appl Earth Obs Remote Sens 11(12): 4563-4577. https://doi.org/10.1109/JSTARS.2018.2823361

Ajith S, Debnath MK, Karthik R (2024) Statistical and machine learning models for location-specific crop yield prediction using weather indices. Int J Biometeorol. https://doi.org/10.1007/s00484-024-02763-w

Allen DM (1971) Mean square error of prediction as a criterion for selecting variables. Technometrics 13(3):469–475. https://doi.org/10.1080/00401706.1971.10488811

Altmann A, Toloşi L, Sander O, Lengauer T (2010) Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Arabameri A, Chen W, Loche M, Zhao X, Li Y, Lombardo L, Cerda A, Pradhan B, Bui DT (2020) Comparison of machine learning models for gully erosion susceptibility mapping. Geosci Front 11(5): 1609–1620. https://doi.org/10.1016/j.gsf.2019.11.009

Ashapure A, Jung J, Chang A, Oh S, Yeom J, Maeda M, Maeda A, Dube N, Landivar J, Hague S (2020) Developing a machine learning based cotton yield estimation framework using multi-temporal UAS data. ISPRS J Photogrammetry Remote Sens 169: 180–194. https://doi.org/10.1016/j.isprsjprs.2020.09.015

Badani C, Chen JS, Gil L, Kilavuz A, Sharma T, Zhou W (2020) Using Index Insurance to Promote Climate-Smart Agriculture. Columbia University.

Barnes E, Clarke T, Richards S, Colaizzi P, Haberland J, Kostrzewski M, Waller P, Choi C, Riley E, Thompson T (2000) Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA,

Bendig J, Yu K, Aasen H, Bolten A, Bennertz S, Broscheit J, Gnyp ML, Bareth G (2015) Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. Int J Appl Earth Observation Geoinform 39: 79–87. https://doi.org/10.1016/j.jag.2015.02.012

Bischl B, Binder M, Lang M, Pielok T, Richter J, Coors S, Thomas J, Ullmann T, Becker M, Boulesteix AL (2023) Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. Wiley Interdisciplinary Reviews 13(2); e1484. https://doi.org/10.48550/arXiv.2107.05847

Cai Y, Guan K, Lobell D, Potgieter AB, Wang S, Peng J, Xu T, Asseng S, Zhang Y, You L (2019) Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and Forest Meteorology, 274: 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010

Camargo-Alvarez H, Elliott RJR, Olin S, Wang X, Wang C, Ray DK, Pugh TAM (2023) Modelling crop yield and harvest index: the role of carbon assimilation and allocation parameters. Modeling Earth Syst Environ 9(2):2617–2635. https://doi.org/10.1007/s40808-022-01625-x

Cao J, Zhang Z, Luo Y, Zhang L, Zhang J, Li Z, Tao F (2021) Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. Eur J Agron 123:126204

Chen N, Yu L, Zhang X, Shen Y, Zeng L, Hu Q, Niyogi D (2020) Mapping paddy rice fields by combining multi-temporal vegetation index and synthetic aperture radar remote sensing data using google earth engine machine learning platform. Remote Sensing 12(18):2992

Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30

Elavarasan D, Vincent PM, DR, Srinivasan K, Chang C-Y (2020) A hybrid CFS filter and rf-rfe wrapper-based feature extraction for enhanced agricultural crop yield prediction modeling. Agriculture, 10(9): 400. https://www.mdpi.com/2077-0472/10/9/400

Everingham Y, Sexton J, Skocaj D, Inman-Bamber G (2016) Accurate prediction of sugarcane yield using a random forest algorithm. Agronomy Sustain Dev 36(2), 1–9. https://doi.org/10.1007/s13593-016-0364-z

Feng X, Xiao Z, Zhong B, Dong Y, Qiu J (2019) Dynamic weighted ensemble classification for credit scoring using Markov Chain. Appl Intell 49: 555–568. https://doi.org/10.1007/s10489-018-1253-8

Ferencz C, Bognar P, Lichtenberger J, Hamar D, Tarcsai G, Timár G, Molnár G, Pásztor S, Steinbach P, Székely B (2004) Crop yield estimation by satellite remote sensing. Int J Remote Sens 25(20): 4113–4149. https://doi.org/10.1080/01431160410001698870

De Filippi P, Mannan M, Reijers W (2020) Blockchain as a confidence machine: The problem of trust & challenges of governance. Technol Soc 62: 101284. https://doi.org/10.1016/j.techsoc.2020.101284

Friedman M (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc 32(200): 675–701. https://doi.org/10.1080/01621459.1937.10503522

García V, Marqués AI, Sánchez JS (2015) An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. J Intell Inform Syst 44: 159–189. https://doi.org/10.1007/s10844-014-0333-4

Gevaert CM, Suomalainen J, Tang J, Kooistra L (2015) Generation of spectral–temporal response surfaces by combining multispectral satellite and hyperspectral UAV imagery for precision agriculture applications. IEEE J Selected Topics Appl Earth Obs Remote Sens 8(6): 3140–3146. https://doi.org/10.1109/JSTARS.2015.2406339

Gitelson A, Merzlyak MN (1994) Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. J Photochem Photobiol B 22(3), 247–252. https://doi.org/10.1016/1011-1344(93)06963-4

Gitelson AA, Gritz Y, Merzlyak MN (2003a) Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J Plant Physiol 160(3): 271–282. https://doi.org/10.1078/0176-1617-00887

Gitelson AA, Viña A, Arkebauer TJ, Rundquist DC, Keydan G, Leavitt BJGrl (2003b) Remote estimation of leaf area index and green leaf biomass in maize canopies. 30(5).

Greitzer FL, Kangas LJ, Noonan CF, Dalton AC, Hohimer RE (2012) Identifying at-risk employees: Modeling psychosocial precursors of potential insider threats. 2012 45th Hawaii International Conference on System Sciences,

Holzman ME, Carmona F, Rivas R, Niclòs R (2018) Early assessment of crop yield from remotely sensed water stress and solar radiation data. ISPRS J Photogrammetry Remote Sens 145: 297–308. https://doi.org/10.1016/j.isprsjprs.2018.03.014

Honkavaara E, Saari H, Kaivosoja J, Pölönen I, Hakala T, Litkey P, Mäkynen J, Pesonen L (2013) Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture. Remote Sens 5(10): 5006–5039. https://doi.org/10.3390/rs5105006

Huang J, Wang H, Dai Q, Han D (2014) Analysis of NDVI data for crop identification and yield estimation. IEEE J Selected Topics Appl Earth Obs Remote Sens 7(11), 4374–4384. https://doi.org/10.1109/JSTARS.2014.2334332

Huete AR (1988) A soil-adjusted vegetation index (SAVI). Remote Sens Environ 25(3):295–309

Jacovides C, Kontoyiannis H (1995) Statistical procedures for the evaluation of evapotranspiration computing models. Agricult Water Manage 27(3–4): 365–371. https://doi.org/10.1016/0378-3774(95)01152-9

Jiang Z, Huete AR, Didan K, Miura TJRsoE (2008) Development of a two-band enhanced vegetation index without a blue band. 112(10), 3833–3845.

Johnson DM (2016) A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products. Int J Appl Earth Obs Geoinformation, 52, 65–81. https://doi.org/10.1016/j.jag.2016.05.010

Jones EJ, Bishop TF, Malone BP, Hulme PJ, Whelan BM, Filippi P (2022) Identifying causes of crop yield variability with interpretive machine learning. Comput Electron Agric 192:106632

Kamir E, Waldner F, Hochman Z (2020) Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS J Photogrammetry Remote Sens 160: 124–135. https://doi.org/10.1016/j.isprsjprs.2019.11.008

Kastens JH, Kastens TL, Kastens DL, Price KP, Martinko EA, Lee R.-Y (2005) Image masking for crop yield forecasting using AVHRR NDVI time series imagery. Remote Sens Environ 99(3): 341–356. https://doi.org/10.1016/j.rse.2005.09.010

Khaki S, Wang L (2019) Crop yield prediction using deep neural networks. Frontiers in plant science, 10: 621. https://doi.org/10.3389/fpls.2019.00621

Khoshroo A, Emrouznejad A, Ghaffarizadeh A, Kasraei M, Omid M (2018) Sensitivity analysis of energy inputs in crop production using artificial neural networks. J Clea prod, 197: 992–998. https://doi.org/10.1016/j.jclepro.2018.05.249

Kigo SN, Omondi EO, Omolo BO (2023) Assessing predictive performance of supervised machine learning algorithms for a diamond pricing model. Sci Rep 13(1): 17315. https://doi.org/10.1038/s41598-023-44326-w

Kim Y, Kim Y (2022) Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. Sustain Cities Soc 79: 103677. https://doi.org/10.1016/j.scs.2022.103677

Kiran Kumar V, Ramesh KV, Rakesh V (2023) Optimizing LSTM and Bi-LSTM models for crop yield prediction and comparison of their performance with traditional machine learning techniques. Appl Intell 53(23):28291–28309. https://doi.org/10.1007/s10489-023-05005-5

Kogan F, Kussul N, Adamenko T, Skakun S, Kravchenko O, Kryvobok O, Shelestov A, Kolotii A, Kussul O, Lavrenyuk A (2013) Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. Int J Appl Earth Obs Geoinf 23:192–203

Krishnadoss N, Ramasamy LK (2024) Crop yield prediction with environmental and chemical variables using optimized ensemble predictive model in machine learning. Environ Res Commun 6(10): 101001. https://doi.org/10.1088/2515-7620/ad7e81

Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D (2018) Machine learning in agriculture: a review. Sensors 18(8): 2674. https://doi.org/10.3390/s18082674

Liu HQ, Huete A (1995) A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Trans Geosci Remote Sens 33(2): 457–465. https://doi.org/10.1109/TGRS.1995.8746027

Lobell DB, Cassman KG, Field CB (2009) Crop yield gaps: their importance, magnitudes, and causes. Ann Rev Environ Resources 34(1): 179–204. https://doi.org/10.1146/annurev.environ.041008.093740

Lopatin E, Poikonen P (2023) Cost-effective aerial inventory of spruce seedlings using consumer drones and deep learning techniques with two-stage UAV flight patterns. Forests 14(5): 973. https://doi.org/10.3390/f14050973

Lundberg, S. M., & Lee, S.-I. (2017) A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*. https://doi.org/10.48550/arXiv.1705.07874

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014) Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual review of clinical psychology*, *10*, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Mateo-Sanchis A, Piles M, Muñoz-Marí J, Adsuara JE, Pérez-Suay A, Camps-Valls G (2019) Synergistic integration of optical and microwave satellite data for crop yield estimation. Remote Sens Environ 234:111460

McFeeters SK (1996) The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int J Remote Sens 17(7):1425–1432

Milla-Val J, Montañés C, Fueyo N (2024) Economical microscale predictions of wind over complex terrain from mesoscale simulations using machine learning. Modeling Earth Systems and Environment 10(1):1407–1421. https://doi.org/10.1007/s40808-023-01851-x

Mwaura JI, Kenduiywo BK (2021) County level maize yield estimation using artificial neural network. Modeling Earth Systems and Environment 7(3):1417–1424. https://doi.org/10.1007/s40808-020-00943-2

Naderloo, L., Alimardani, R., Omid, M., Sarmadian, F., Javadikia, P., Torabi, M. Y., & Alimardani, F. (2012) Application of ANFIS to predict crop yield based on different energy inputs. *Measurement*, *45*(6), 1406–1413. https://doi.org/10.1016/j.measurement.2012.03.025

Nebiker, S., Lack, N., Abächerli, M., & Läderach, S. (2016) Lightweight multispectral UAV sensors and their capabilities for predicting grain yield and detecting plant diseases. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *41*, 963-970. 10.5194/isprs-archives-XLI-B1-963-2016

Nemenyi, P. B. (1963) *Distribution-free multiple comparisons*. Princeton University.

Oikonomidis A, Catal C, Kassahun A (2022) Hybrid deep learning-based models for crop yield prediction. Applied artificial intelligence 36(1): 2031822. https://doi.org/10.1080/08839514.2022.2031823

Panday US, Pratihast AK, Aryal J, Kayastha RB (2020) A review on drone-based data solutions for cereal crops. Drones 4(3): 41. https://doi.org/10.3390/drones4030041

Peerlinck A, Sheppard J, Senecal J (2019) Adaboost with neural networks for yield and protein prediction in precision agriculture. 2019 International Joint Conference on Neural Networks (IJCNN),

Ransom CJ, Kitchen NR, Camberato JJ, Carter PR, Ferguson RB, Fernández FG, Franzen DW, Laboski CA, Myers DB, Nafziger ED (2019) Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. Comput Electron Agricult 164: 104872. https://doi.org/10.1016/j.compag.2019.104872

Remman SB, Lekkas AM (2021) Robotic lever manipulation using hindsight experience replay and shapley additive explanations. 2021 European Control Conference (ECC),

Rodríguez-Pérez R, Bajorath, J. (2020) Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. J Comput-Aided Mol Design 34: 1013–1026. https://doi.org/10.1007/s10822-020-00314-0

Semeraro T, Mastroleo G, Pomes A, Luvisi A, Gissi E, Aretano R (2019) Modelling fuzzy combination of remote sensing vegetation index for durum wheat crop analysis. Comput Electron Agric 156:684–692

Shahhosseini M, Hu G, Huber I, Archontoulis SV (2021) Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific reports, 11(1): 1–15. https://doi.org/10.1038/s41598-020-80820-1

Shapley A (1953) value for n-person games, Contrib. Theory Games(2), 307. https://doi.org/10.1515/9781400881970-018

Shendryk Y, Davy R, Thorburn P (2021) Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. Field Crops Res 260: 107984. https://doi.org/10.1016/j.fcr.2020.107984

Singh R, Semwal D, Rai A, Chhikara RS (2002) Small area estimation of crop yield using remote sensing satellite data. Int J Remote Sens 23(1): 49–56. https://doi.org/10.1080/01431160010014756

Srivastava AK, Safaei N, Khaki S, Lopez G, Zeng W, Ewert F, Gaiser T, Rahimi J (2021) Comparison of machine learning methods for predicting winter wheat yield in Germany. arXiv preprint arXiv:2105.01282. https://doi.org/10.48550/arXiv.2105.01282

Srivastava AK, Safaei N, Khaki S, Lopez G, Zeng W, Ewert F, Gaiser T, Rahimi J (2022) Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. Scientific reports, 12(1): 3215. https://doi.org/10.1038/s41598-022-06249-w

Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis, A. (2008) Conditional variable importance for random forests. *BMC bioinformatics*, *9*, 1–11. https://doi.org/10.1186/1471-2105-9-307

Sun J, Lai Z, Di L, Sun Z, Tao J, Shen Y (2020) Multilevel deep learning network for county-level corn yield estimation in the us corn belt. IEEE J Selected Topics Appl Earth Obs Remote Sens 13: 5048–5060. https://doi.org/10.1109/JSTARS.2020.3019046

Tripathi A, Tiwari RK, Tiwari SP (2022) A deep learning multi-layer perceptron and remote sensing approach for soil health based crop yield estimation. Int J Appl Earth Obs Geoinform 113:102959. https://doi.org/10.1016/j.jag.2022.102959

Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens Environ 8(2): 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Vallentin C, Harfenmeister K, Itzerott S, Kleinschmit B, Conrad C, Spengler D (2022) Suitability of satellite remote sensing data for yield estimation in northeast Germany. Precision Agricult 23(1): 52–82. https://doi.org/10.1007/s11119-021-09827-6

Wang X, Müller C, Elliot J, Mueller ND, Ciais P, Jägermeyr J, Gerber J, Dumas P, Wang C, Yang H, Li L, Deryng D, Folberth C, Liu W, Makowski D, Olin S, Pugh TAM, Reddy A, Schmid E, Piao S (2021) Global irrigation contribution to wheat and maize yield. Nat Commun 12(1):1235. https://doi.org/10.1038/s41467-021-21498-5

Wei Q, Jingxuan L, Lin L, Xiao-Wen L (2011) Research on automatic extraction of water bodies and wetlands on HJ satellite CCD images. Remote Sensing Information 4:28–33

Whitehead K, Hugenholtz CH (2014) Remote sensing of the environment with small unmanned aircraft systems (UASs), part 1: A review of progress and challenges. J Unmanned Vehicle Syst 2(3): 69–85. https://doi.org/10.1139/juvs-2014-0006

Xu M, Liu R, Chen JM, Liu Y, Shang R, Ju W, Wu C, Huang W (2019) Retrieving leaf chlorophyll content using a matrix-based vegetation index combination approach. Remote Sens Environ 224: 60–73. https://doi.org/10.1016/j.rse.2019.01.039

Xu W, Chen P, Zhan Y, Chen S, Zhang L, Lan Y (2021) Cotton yield estimation model based on machine learning using time series UAV remote sensing data. Int J Appl Earth Observation

Geoinformation 104: 102511. https://doi.org/10.1016/j.jag.2021.102511

You J, Li X, Low M, Lobell D, Ermon S (2017) Deep gaussian process for crop yield prediction based on remote sensing data. Thirty-First AAAI conference on artificial intelligence,

Zacharias S, Heatwole C, Coakley C (1996) Robust quantitative techniques for validating pesticide transport models. Trans ASAE 39(1):47–54

Zhang X, Izaurralde RC, Manowitz D, West T, Post W, Thomson AM, Bandaru, VP, Nichols J, Williams J (2010) An integrative modeling framework to evaluate the productivity and sustainability of biofuel crop production systems. Gcb Bioenergy, 2(5): 258–277. https://doi.org/10.1111/j.1757-1707.2010.01046.x

Zhu P, Abramoff R, Makowski D, Ciais P (2021) Uncovering the past and future climate drivers of wheat yield shocks in Europe with machine learning. Earth's Future, 9(5), e2020EF001815. https://doi.org/10.1029/2020EF001815