



## Research Article

<https://doi.org/10.1631/jzus.B2600177>

# Ensemble learning for wheat yield prediction using multi-source and multi-temporal data

Runzhou CAO<sup>1\*</sup>, Xuyang SUO<sup>1\*</sup>, Xuping FENG<sup>1</sup>, Yong HE<sup>2✉</sup>

<sup>1</sup>College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

<sup>2</sup>Hainan Institute of Zhejiang University, Sanya 572025, China

**Abstract:** Measuring crop yields for hundreds of wheat germplasm accessions is time-consuming and labor-intensive in plant breeding. Integrating drone-based remote sensing with machine learning techniques has recently emerged as an effective solution for streamlining this process. Despite these advances, selecting an optimal predictive model remains challenging because machine learning algorithms vary in their principles and characteristics. In this study, support vector regression, random forest, extreme gradient boosting, and gradient boosting decision tree served as base learners and were integrated through blending ensemble learning. Multiple blending ensemble models were developed by varying the proportions of training and hold-out sets, and a weighted average was subsequently applied to the resulting models. The proposed model achieved an  $R^2$  of 0.889 and an RMSE of 0.470 t/ha, demonstrating its superior performance, its ability to accurately predict yields across multiple wheat varieties, and its effectiveness in alleviating model-selection challenges. Furthermore, the heading stage was identified as the most suitable growth period for accurate multi-variety wheat yield prediction.

**Key words:** Wheat; Machine learning; Yield prediction; Remote sensing

## 1 Introduction

The global population is expected to reach 9.7 billion by 2050; to meet future food demands, crop yields will therefore need to increase by 25%–70% (Hunter et al., 2017). Cultivating and selecting high-yielding wheat varieties is essential to ensure global food security and improve agricultural productivity, as wheat is a major global staple crop. However, traditional yield measurement methods rely on manual harvesting and evaluation, making the process slow, labor-intensive, and inefficient. To address these limitations, machine learning-based prediction models utilizing multi-source and multi-temporal data have emerged as an efficient and effective solution (Li et al., 2022).

Remote sensing technology, which uses electromagnetic waves to collect information about Earth's surface, has become an indispensable tool in farming (Esmacili et al., 2023; Jalayer et al., 2023). Over the past few decades, satellite-based remote sensing has been widely used to track crop development because it can collect

✉ Yong HE, yhe@zju.edu.cn

\* The two authors contributed equally to this work

Yong HE, <https://orcid.org/0000-0001-6752-1757>

Runzhou CAO, <https://orcid.org/0009-0004-1798-3856>

data efficiently in a non-destructive manner (Demir and Dursun, 2024; Dursun et al., 2025). This approach is particularly effective for predicting yields and monitoring biomass (Huang et al., 2019; Nejad et al., 2023). For instance, Wang et al. (2020a) developed a deep learning model that combines a long short-term memory network and a convolutional neural network with satellite and meteorological data to forecast wheat yields in China, achieving an  $R^2$  of 0.77 and RMSE of 0.721 t/ha. Despite these successes, satellite remote sensing is constrained by challenges such as cloud cover and limited resolution, which can reduce estimation accuracy. In contrast, Unmanned Aerial Vehicle (UAV) remote sensing provides higher-resolution imagery and greater flexibility to adjust flight paths to circumvent cloud interference, making it particularly suitable for detailed field-scale assessments. Yang et al. (2022) used multi-temporal UAV data to achieve high-accuracy corn yield predictions ( $R^2 = 0.93$  and RMSE = 0.77 t/ha). Similarly, Demir et al. (2024) utilized UAV and soil data to predict organic rose yields, also achieving an  $R^2$  of 0.93. These high-resolution UAV images significantly improve prediction-model accuracy, providing reliable data for precision agriculture applications.

Multi-temporal and multi-source data fusion integrates information acquired at different times and from various sources, offering a more comprehensive view of crop dynamics and characteristics (Wan et al., 2020; González-Vivar et al., 2026). In crop yield prediction, multi-temporal data track crop development across various growth stages, while multi-source data integrates diverse factors such as multispectral data, soil properties, and fertilization information, which reduces the impact of missing or erroneous data from any single source. In summary, integrating data from multiple time periods and diverse sources improves yield estimation accuracy, making it a widely used and effective method for enhancing prediction outcomes (Han et al., 2020). For instance, Niraimathi and Saravanan (2025) used multi-source remote sensing and environmental data for rice yield estimation and found that fusing all data sources yielded the best model performance, with an  $R^2$  of 0.78.

Machine learning has significantly improved the accuracy and reliability of crop yield prediction in recent years (Van Klompenburg et al., 2020). For instance, Bian et al. (2022) used six machine learning techniques with drone multispectral data to predict field-scale winter wheat yield with good predictive performance. Despite these successes, single machine learning models face generalization and performance challenges (Li et al., 2023). In practice, variations in feature distribution, noise levels, and sensitivity to input data can significantly impact the effectiveness of machine learning techniques across scenarios, often making it difficult to identify a single best model. To overcome this challenge, ensemble learning, an advanced machine learning approach, combines multiple models to leverage their respective strengths (Fan et al., 2025). This method reduces bias and variance through model diversity, improving prediction accuracy and robustness and reducing reliance on a single model. Deng et al. (2025) developed a stacked ensemble model based on gradient boosting tree, random forest, and support vector regression to predict sorghum yield, achieving the best performance among the models evaluated in the study. Similarly, Chu et al. (2025) predicted yield using an ensemble learning approach that integrated multispectral data, RGB imagery, and wheat height features, achieving superior predictive performance to all single models, with an  $R^2$  of 0.711 and an RMSE of 1041 kg/ha. These studies demonstrate the potential of ensemble learning to improve predictive models' generalization. However, in these frameworks, all base learners typically receive the same input features, and differences in feature sensitivity among algorithms are not accounted for. Furthermore, blending approaches used for yield prediction have generally been constructed on a single training/hold-out ratio, leaving partition sensitivity unaddressed.

The present study aims to: (1) identify the optimal growth stage for wheat yield prediction; (2) determine an effective strategy for integrating multi-source data for wheat yield prediction; and (3) develop an ensemble framework combining model-specific feature selection, multiple blending partitions, and performance-based weighted averaging for yield prediction across genetically diverse wheat germplasm accessions.

## 2 Material and methods

### 2.1 Field experimental design

The experiment took place at the Yuhang Agricultural Experimental Base, part of the Zhejiang University Agricultural Experimental Station (30°53'6"N, 119°37'48"E), from November 2022 to May 2023. This site was selected to develop an ensemble model to predict yields for various winter wheat germplasm accessions. The experimental base, situated in Zhejiang Province along China's southeastern coast, experiences a subtropical monsoon climate. Summers are typically warm and moist, while winters are typically dry, cold, and sunny.

The study area comprised 350 small plots, each planted with a winter wheat accession. Each plot measured 1.5 m x 1.25 m, with a planting density of 80 grains per square meter. Fertilization and irrigation were standardized according to local practices, applying 375 kg of compound fertilizer (N: P: K = 16: 16: 16) per hectare before sowing. Throughout the wheat growing season, data were collected 10 times at different stages, including tillering, jointing, booting, and other growth stages. To simplify the description, the data collection stages were labeled with numbers presented in the following text (Supplementary Table 1). Leaf area index (LAI), relative chlorophyll content, and drone-based multispectral data were collected. All sampling and drone flights were performed between 10:00 and 16:00 local time under cloudless, clear conditions. The following describes the data sources utilized in this study.

This study used two key physiological parameters (LAI and relative chlorophyll content), along with five-band multispectral data and 14 vegetation indices from UAV remote sensing imagery, to forecast yield across 350 winter wheat varieties. LAI, which measures total leaf area per unit of land, reflects wheat growth status and density. Measurements were taken using the LAI-2000C plant canopy analyzer (LI-COR, USA). A higher LAI typically indicates dense vegetation and robust growth. Relative chlorophyll content (SPAD values), assessed using a SPAD-502 PLUS chlorophyll meter (Konica Minolta, Japan), serves as an indicator of plant health by measuring chlorophyll concentration. This parameter reflects the plants' photosynthetic efficiency and nutritional status, with higher chlorophyll content generally indicating better health and photosynthetic capacity. The multispectral data provided information on photosynthetic activity, water status, and other canopy characteristics. Vegetation indices were derived from the multispectral data to provide additional information about plant growth.

### 2.2 UAV-based data collection

Multispectral imaging was performed using a DJI Phantom 4 Multispectral (DJI P4M) drone equipped with a FC6360 multispectral camera. This camera has six 1/2.9-inch CMOS sensors: one RGB sensor for true-color imaging and five monochrome sensors for multispectral imaging. It captures data across five spectral bands: blue, green, red, red edge, and near-infrared. For each experiment, the flight mission was automatically executed as programmed in DJI GS Pro, DJI's flight-planning software. The drone maintained an average altitude of 15 meters, with a ground sampling distance of 0.8 cm/pixel. Calibration panel images were acquired before and after each flight for radiometric calibration.

Each mission acquired approximately 2,400 spectral images, each with a resolution of  $1600 \times 1300$  pixels. The images were processed into multispectral orthomosaics using DJI Terra. Radiometric correction was performed by applying the known five-band spectral reflectance values of the radiometric calibration panel. Subsequently, ArcGIS Pro was used to create masks with specific IDs and shapes, allowing the extraction of spectral data from designated regions of interest and calculation of vegetation indices (Fig. 1).

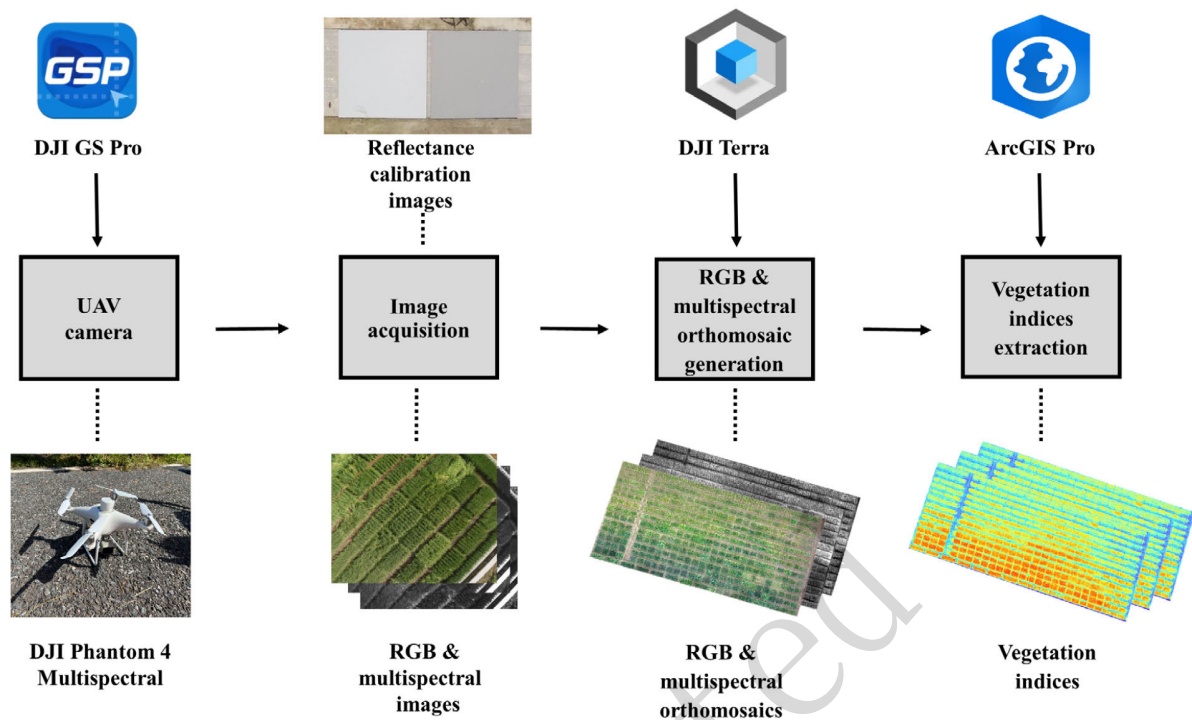


**Fig. 1 Processing workflow for UAV-based images. UAV: Unmanned Aerial Vehicle.**

### 2.3 Crop physiological parameter collection

SPAD values and LAI were measured for each experimental plot after every flight mission. The SPAD-502 PLUS meter was used to measure the relative chlorophyll content of wheat leaves. The LAI of the wheat canopy was assessed using an LAI-2000C plant canopy analyzer. Each plot was measured three times. The sampling points were randomly selected, and their locations were recorded. At maturity, the crop in each plot was harvested manually. The harvested seeds were thoroughly dried and then weighed. The yield of each wheat variety was calculated by dividing the seed weight by the harvested area.

### 2.4 Vegetation indices

To predict wheat yield, the study used 14 vegetation indices calculated from the red, green, blue, red-edge, and near-infrared bands acquired by the UAV multispectral sensors. The definitions and calculation formulas for these vegetation indices are provided in Supplementary Table 2. Vegetation indices have been extensively utilized in grain yield prediction and have demonstrated good predictive performance (Panda et al., 2010; Johnson et al., 2016; Da Silva et al., 2020).

### 2.5 Ensemble model development

This study proposed an ensemble learning strategy that combines blending ensemble learning with weighted averaging to improve prediction performance (Fig. 2). The procedure was as follows: (1) The dataset was split, with 20% allocated to the test set (Du et al., 2022; Wu et al., 2026). The remaining data were divided into training and hold-out sets using five different training-to-hold-out ratios (5:5, 6:4, 7:3, 8:2, and 9:1) (Sun et al., 2023; Hasan et al., 2024). (2) For each base model, instead of using all available features, an appropriate feature subset with better predictive performance was selected. (3) For each data split, the base learners were trained on the training set, and the corresponding hold-out set was used to train the meta-learner, resulting in five blending ensemble models. (4) The predictive performance of each model on its corresponding hold-out

set was evaluated using  $R^2$ . Based on their relative performance, the five blending models were assigned weights of 20, 5, 5, 5, and 1, respectively. The five models were then integrated through weighted averaging of their predictions to construct the final ensemble model, the predictive performance of which was then evaluated on the independent test set. The base models included support vector regression (SVR), gradient boosting decision tree (GBDT), extreme gradient boosting (XGBoost), and random forest (RF).

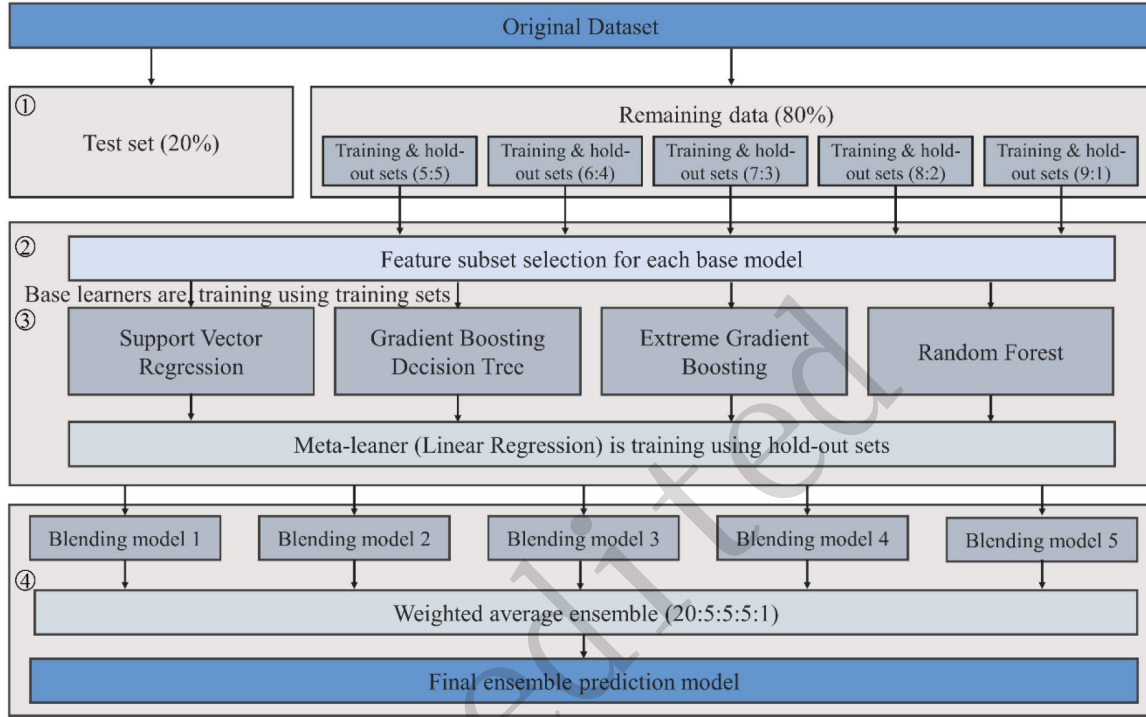


**Fig. 2 Workflow of the winter wheat yield prediction framework.**

In this study, the coefficient of determination ( $R^2$ ) and root mean square error (RMSE) were used to evaluate model performance. The corresponding equations are as follows:

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

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

Note:  $y_i$  represents the measured wheat yield for sample  $i$ ,  $\hat{y}_i$  represents the predicted wheat yield for sample  $i$ ,  $\bar{y}$  represents the mean measured wheat yield of all samples, and  $n$  is the total number of samples.

### 3 Results and discussion

### 3.1 Correlation analysis

#### 3.1.1 Correlation between multispectral information and wheat yield

In analyzing the correlation between spectral bands and wheat yield across growth stages (Fig. 3a), it was observed that the blue band consistently showed a negative correlation with yield throughout all stages. This negative association was strongest during stages 4–8, reaching the strongest negative correlation at stage 7, with a correlation coefficient of -0.72. Similarly, the green band was negatively correlated with yield across all growth stages, with stages 4–8 showing strong correlations and the strongest negative correlation (-0.86) at stage 7. The red band also showed a predominantly negative correlation, particularly strong between stages 4 and 8, with the strongest negative correlation (-0.82) at stage 5 and a marked weakening during stages 9 and 10. The red-edge band generally showed a negative correlation, with fluctuations in correlation strength, and showed strong correlations at stages 4, 7, and 8, with the strongest negative correlation of -0.78 occurring at stage 7, followed by a notable weakening during stages 9 and 10. In contrast, the near-infrared band was positively correlated with yield in most stages, especially from stages 5 to 8, with the strongest positive correlation (0.68) at stage 6.

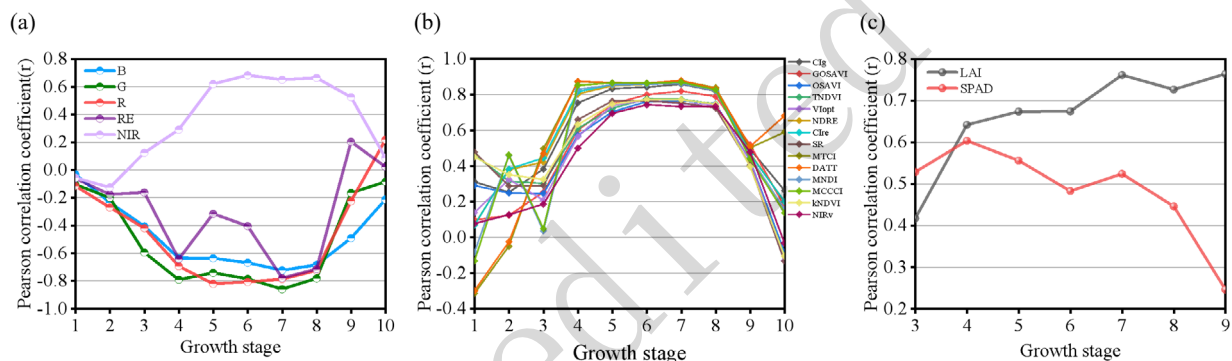


**Fig. 3 Correlations: (a) Multispectral data and measured yield of winter wheat, (b) Vegetation indices and measured yield of winter wheat, (c) LAI, SPAD values, and measured yield of winter wheat. The vegetation indices used are listed in Supplementary Table 2. B: blue band; G: green band; R: red band. RE: red edge band; NIR: near-infrared band.**

Near-infrared reflectance was positively correlated with wheat grain yield across most growth stages. This relationship may reflect near-infrared reflectance sensitivity to internal leaf structure, canopy leaf area, and multiple scattering within the canopy (Wan et al., 2024). Wheat genotypes with well-developed canopies, larger leaf areas, and greater biomass generally exhibit higher near-infrared reflectance. These canopy characteristics enhance light interception, dry matter accumulation, and grain yield formation (Marshall et al., 2022). In contrast, reflectance in the blue, green, and red bands was generally negatively correlated with grain yield. This may be because wheat genotypes with larger leaf areas, higher chlorophyll content, and greater canopy cover typically absorb more visible radiation and therefore exhibit lower reflectance in these bands (Zhang et al., 2021). Reflectance in the red-edge band was also generally negatively correlated with wheat grain yield. This relationship may reflect stronger and broader red-light absorption with increasing canopy chlorophyll content, shifting the red-edge position toward longer wavelengths and potentially reducing reflectance at a fixed red-edge wavelength (Fei et al., 2022).

#### 3.1.2 Correlation between vegetation indices and wheat yield

Analysis of correlations between wheat vegetation indices and yield at different growth stages (Fig. 3b) showed that from stages 1 to 5, correlations between most vegetation indices and yield increased overall and were strongest at stage 5. For example, the correlation coefficient of CIg reached 0.83 at stage 5, while that of

MNDI was 0.86. From stages 6 to 8, most vegetation indices remained highly correlated with yield, with slight fluctuations. For instance, DATT correlation coefficients at stages 6, 7, and 8 were 0.87, 0.86, and 0.88, respectively. Similarly, NDRE consistently maintained a high level of correlation during these stages, reaching a maximum of 0.86. At stages 9 and 10, the correlations between most vegetation indices and yield began to decline, with some indices even showing negative correlations. For example, the correlation coefficient of VI<sub>opt</sub> at stage 10 was -0.13. Overall, most vegetation indices were strongly positively correlated with yield during the mid-growth stages, while during the early and late stages, correlations were relatively low or even negative.

The correlations between most vegetation indices and wheat grain yield generally increased from stages 1 to 5, remained relatively stable from stages 5 to 8, and declined during stages 9 and 10. This may reflect the sensitivity of vegetation indices to canopy chlorophyll status and structural characteristics (Liu et al., 2022). During the early growth stages, the crop canopy was not fully closed, and vegetation cover and LAI were relatively low. Consequently, remote-sensing signals were susceptible to interference from soil background and plant shadows, limiting vegetation indices' ability to characterize crop growth. As the crop developed and the canopy expanded, the soil background's contribution to the spectral signal gradually decreased, enhancing vegetation indices' ability to capture canopy physiological and structural characteristics (Almeida-Ñauñay et al., 2023). From jointing to flowering (stages 5–8), LAI and canopy closure were relatively high, and functional leaves maintained high chlorophyll content. Canopy spectral characteristics could therefore effectively reflect crop biomass. This period also encompasses critical processes associated with spike formation, floral development, pollination, and the determination of potential grain number. Canopy status during these stages was therefore closely associated with final yield potential (Sabir et al., 2023). During stages 9 and 10, progressive leaf senescence, chlorophyll degradation, and changes in canopy structure caused vegetation indices to increasingly reflect canopy senescence and declining leaf function, weakening their correlations with grain yield (Yu et al., 2024).

### 3.1.3 Correlations between LAI, SPAD values, and wheat yield

The analysis of correlations between LAI and SPAD values and wheat yield at different growth stages (Fig. 3c) showed that both LAI and SPAD values were consistently positively correlated with wheat yield throughout all growth stages. The correlation between LAI and wheat yield increased overall across growth stages, starting at 0.42 at the 3rd stage and reaching 0.76 at the 9th stage. In contrast, the correlation between SPAD values and wheat yield displayed an overall decreasing trend, with the highest correlation of 0.60 at the 4th stage and the lowest correlation of 0.25 at the 9th stage.

LAI and SPAD values were positively correlated with wheat grain yield across all observation stages, indicating that a larger canopy leaf area and higher leaf chlorophyll content are generally associated with greater wheat yield potential (Han et al., 2021).

## 3.2 Comparison of wheat yield prediction across growth stages

To identify the most suitable growth stage for wheat yield prediction, this study evaluated the four machine learning algorithms using data collected separately during each of the 10 growth stages (Table 1). For each growth stage, the same multimodal feature set, comprising multispectral bands, vegetation indices, LAI, and SPAD values, was used as model input. Thus, the feature configuration remained constant, while the growth stage varied. The comparison revealed that all models performed best when using data from the 7th growth stage (heading period). Notably, the SVR model achieved the highest prediction accuracy, with an  $R^2$  of 0.846 and an RMSE of 0.554 t/ha. Compared with the second-best growth stage (stage 5),  $R^2$  increased by 2.9%, while the RMSE decreased from 0.596 to 0.554 t/ha. The RF model performed best with an  $R^2$  of 0.819 and an RMSE of 0.601 t/ha, representing a 2.0% increase in  $R^2$  and a reduction in RMSE from 0.627 to 0.601 t/ha compared with the second-best growth stage (stage 4). Similarly, the

XGBoost model performed best with an  $R^2$  of 0.838 and an RMSE of 0.568 t/ha, corresponding to a 5.7% improvement in  $R^2$  and a decrease in RMSE from 0.641 to 0.568 t/ha relative to the second-best growth stage (stage 5). The GBDT model achieved optimal prediction accuracy, with an  $R^2$  of 0.819 and an RMSE of 0.600 t/ha, representing a 0.9 % increase in  $R^2$  and a reduction in RMSE from 0.612 to 0.600 t/ha compared with the second-best growth stage (stage 4).

**Table 1 Evaluation of yield estimation model accuracy based on single growth stage data of wheat**

| Growth stage | Models     | SVR          | RF           | XGBoost      | GBDT         |
|--------------|------------|--------------|--------------|--------------|--------------|
| 1            | $R^2$      | 0.505        | 0.353        | 0.220        | 0.325        |
|              | RMSE(t/ha) | 0.992        | 1.134        | 1.245        | 1.159        |
| 2            | $R^2$      | 0.564        | 0.531        | 0.399        | 0.456        |
|              | RMSE(t/ha) | 0.932        | 0.966        | 1.093        | 1.040        |
| 3            | $R^2$      | 0.733        | 0.672        | 0.635        | 0.658        |
|              | RMSE(t/ha) | 0.729        | 0.808        | 0.852        | 0.825        |
| 4            | $R^2$      | 0.800        | 0.803        | 0.786        | 0.812        |
|              | RMSE(t/ha) | 0.631        | 0.627        | 0.652        | 0.612        |
| 5            | $R^2$      | 0.822        | 0.779        | 0.793        | 0.795        |
|              | RMSE(t/ha) | 0.596        | 0.662        | 0.641        | 0.639        |
| 6            | $R^2$      | 0.819        | 0.788        | 0.745        | 0.775        |
|              | RMSE(t/ha) | 0.600        | 0.649        | 0.713        | 0.670        |
| 7            | $R^2$      | <b>0.846</b> | <b>0.819</b> | <b>0.838</b> | <b>0.819</b> |
|              | RMSE(t/ha) | <b>0.554</b> | <b>0.601</b> | <b>0.568</b> | <b>0.600</b> |
| 8            | $R^2$      | 0.817        | 0.779        | 0.714        | 0.770        |
|              | RMSE(t/ha) | 0.604        | 0.663        | 0.754        | 0.677        |
| 9            | $R^2$      | 0.726        | 0.712        | 0.664        | 0.686        |
|              | RMSE(t/ha) | 0.738        | 0.756        | 0.818        | 0.791        |
| 10           | $R^2$      | 0.667        | 0.611        | 0.588        | 0.663        |
|              | RMSE(t/ha) | 0.813        | 0.879        | 0.905        | 0.819        |

Bold values indicate the best performance achieved by each algorithm. SVR: support vector regression; RF: random forest; XGBoost: extreme gradient boosting; GBDT: gradient boosting decision tree;  $R^2$ : coefficient of determination; RMSE: root mean square error.

In this study, models based on data from the wheat heading stage produced more accurate predictions than models using data from other growth periods. This result is consistent with earlier studies emphasizing the key role of the heading stage in predicting wheat yield(Chandel et al., 2019; Tanabe et al., 2023). During the heading stage (Stage 7), wheat enters a critical reproductive phase, during which canopy development is nearly complete and key yield components, such as spike and spikelet numbers, are largely determined (Lou et al., 2024). Consequently, spectral and physiological measurements collected at this stage can more accurately reflect crop growth status and yield potential, resulting in superior prediction

performance. Data from the stem elongation (Stage 4) and jointing (Stage 5) stages also demonstrated reasonable predictive ability, primarily because plants are actively growing at these stages, with increasing leaf area and biomass, allowing multispectral data, vegetation indices, LAI and SPAD values to capture some yield-related features. However, canopy structure and physiological status are still changing rapidly, so their predictive power is lower than at the heading stage (Demir, 2024b).

### 3.3 Comparison of wheat yield prediction using different feature sets

The effects of different input feature sets on model performance were evaluated using data collected throughout the entire wheat growth period. Four feature groups—multispectral bands, vegetation indices, LAI, and SPAD values—were first evaluated separately, followed by an evaluation of their combinations to identify the optimal feature set for each machine learning algorithm. Unlike the growth-stage analysis, the temporal coverage remained consistent, while the input feature set varied.

#### 3.3.1 Individual feature groups

Yield prediction models were developed based on datasets collected across the entire wheat growth period, with each model using one of the following datasets: multispectral information, vegetation indices, LAI, or SPAD values (Fig. 4). The vegetation indices used included all 14 indices extracted in this experiment. The study found that, for the same algorithm, yield prediction accuracy was highest when applying vegetation indices, followed by multispectral information, LAI, and SPAD values. The  $R^2$  values for the models based on vegetation indices all exceeded 0.850, and the RMSE values all fell below 0.600 t/ha. The GBDT method produced the best results, achieving an  $R^2$  of 0.863 and an RMSE of 0.522 t/ha. In contrast, models based on SPAD values had  $R^2$  values below 0.600, making them unsuitable for accurate wheat yield estimation. Compared with models based on multispectral information, LAI, and SPAD values, vegetation-index-based models achieved higher average  $R^2$  values, with improvements of 5.0%, 16.4%, and 88.4%, respectively, while reducing the average RMSE by 0.071, 0.193, and 0.528 t/ha, respectively.

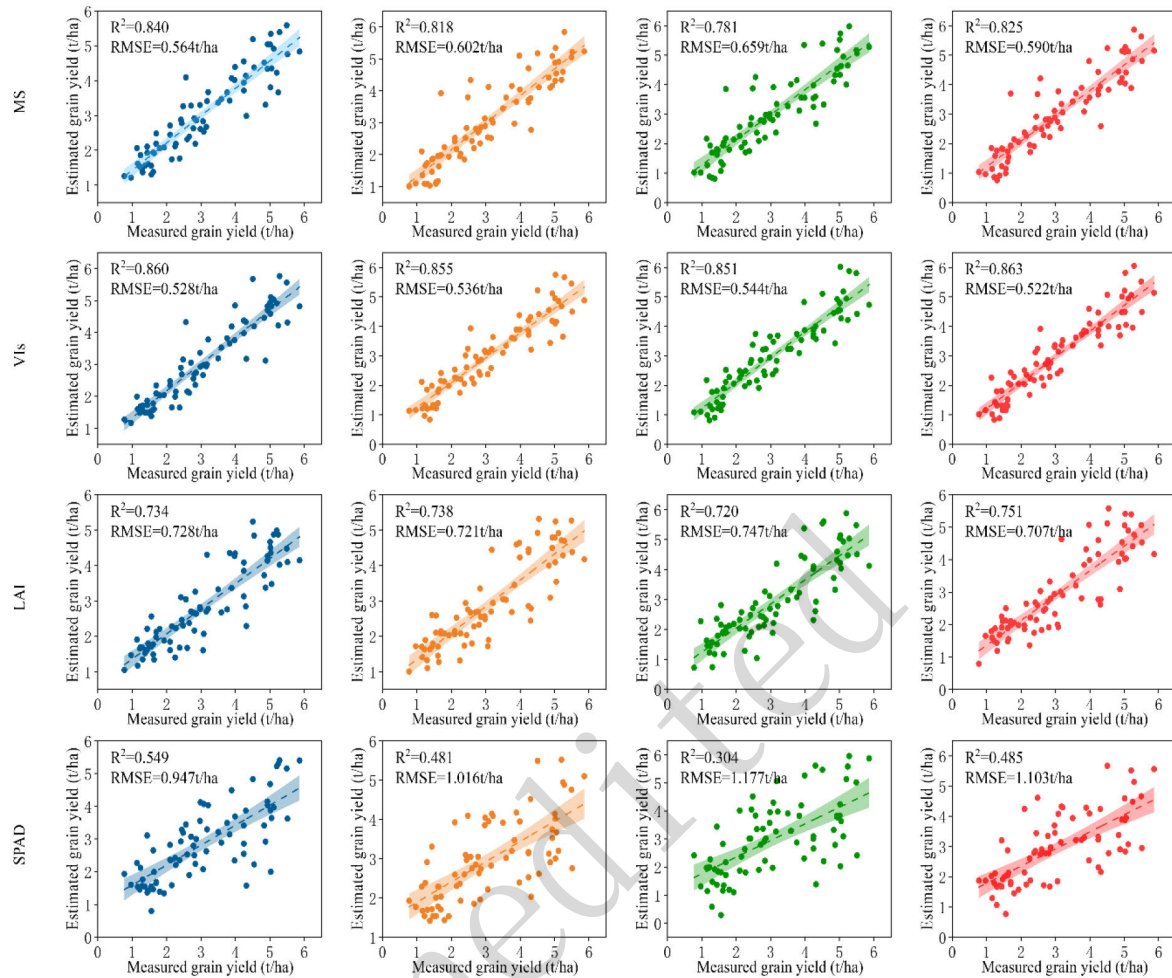


**Fig. 4 Evaluation of yield prediction models based on individual feature groups. Dotted lines represent the fitted linear regression lines between measured and estimated grain yield, and shaded areas represent the 95% confidence intervals of the corresponding fitted regression lines. MS: multispectral data; VIs: vegetation indices; LAI: leaf area index; SPAD: Soil Plant Analysis Development value, an indicator of relative chlorophyll content.**

The results demonstrate that vegetation indices can effectively predict wheat yield, corroborating previous research results (Kaya and Polat, 2023). Multispectral information and LAI also possessed prediction capabilities. However, using SPAD values alone is ineffective, as they failed to provide accurate predictions. The superior performance of vegetation indices across all models may be attributed to their ability to integrate information from multiple spectral bands and effectively characterize key crop traits associated with yield formation, including canopy coverage, leaf area, chlorophyll content, and nitrogen status (Kaya et al., 2026). Compared with directly combining reflectance values from individual spectral bands, vegetation indices provide biologically meaningful features that more comprehensively reflect yield-related crop growth conditions. In contrast, the LAI primarily reflects canopy structural characteristics and does not directly represent physiological traits such as chlorophyll content or nitrogen status. Meanwhile, SPAD values primarily represent leaf chlorophyll content and provide relatively limited information regarding canopy structure and overall crop growth (Monostori et al., 2016). In addition, SPAD measurements may be influenced by leaf characteristics and environmental conditions. Therefore,

LAI and SPAD each capture only specific aspects of crop growth, potentially limiting their predictive power when used individually for yield prediction. Consequently, vegetation indices exhibited the highest prediction accuracy, followed by multispectral data, LAI, and SPAD values.

### 3.3.2 Combined feature groups

This study developed several yield prediction models using multimodal data collected throughout the entire wheat growth period (Table 2). For the SVR algorithm, combining multispectral information and vegetation indices achieved the best performance, with an  $R^2$  of 0.857 and an RMSE of 0.534 t/ha. Similarly, the XGBoost model performed best when using multispectral information and vegetation indices, yielding an  $R^2$  of 0.861 and an RMSE of 0.526 t/ha. For the RF algorithm, the optimal model incorporated vegetation indices and SPAD values, achieving an  $R^2$  of 0.859 and an RMSE of 0.530 t/ha. Likewise, the GBDT model achieved its highest prediction accuracy using vegetation indices and SPAD values, with an  $R^2$  of 0.869 and an RMSE of 0.510 t/ha. Compared with the corresponding models trained using the complete multimodal dataset, these optimal models improved  $R^2$  by 8.1%, 4.0%, 1.9%, and 3.8%, respectively. These findings suggest that incorporating all available data sources does not necessarily lead to superior prediction performance. Instead, appropriate feature selection appears to be more important for maximizing model accuracy, as redundant informative variables may increase model complexity and reduce predictive effectiveness.

**Table 2 Evaluation of yield estimation models based on combined feature groups**

| Features                             | Models     | SVR          | RF           | XGBoost      | GBDT         |
|--------------------------------------|------------|--------------|--------------|--------------|--------------|
| MS bands, VIs(derived from MS bands) | $R^2$      | <b>0.857</b> | 0.852        | <b>0.861</b> | 0.850        |
|                                      | RMSE(t/ha) | <b>0.534</b> | 0.543        | <b>0.526</b> | 0.547        |
| MS bands, LAI                        | $R^2$      | 0.762        | 0.818        | 0.813        | 0.819        |
|                                      | RMSE(t/ha) | 0.687        | 0.601        | 0.610        | 0.600        |
| MS bands, SPAD                       | $R^2$      | 0.507        | 0.828        | 0.777        | 0.816        |
|                                      | RMSE(t/ha) | 0.990        | 0.586        | 0.667        | 0.606        |
| VIs, LAI                             | $R^2$      | 0.856        | 0.848        | 0.800        | 0.828        |
|                                      | RMSE(t/ha) | 0.535        | 0.550        | 0.630        | 0.585        |
| VIs, SPAD                            | $R^2$      | 0.782        | <b>0.859</b> | 0.850        | <b>0.869</b> |
|                                      | RMSE(t/ha) | 0.659        | <b>0.530</b> | 0.546        | <b>0.510</b> |
| LAI, SPAD                            | $R^2$      | 0.732        | 0.798        | 0.737        | 0.764        |
|                                      | RMSE(t/ha) | 0.731        | 0.634        | 0.723        | 0.684        |
| VIs, LAI, SPAD                       | $R^2$      | 0.793        | 0.854        | 0.824        | 0.849        |
|                                      | RMSE(t/ha) | 0.642        | 0.540        | 0.591        | 0.549        |
| MS bands, LAI, SPAD                  | $R^2$      | 0.685        | 0.828        | 0.804        | 0.826        |
|                                      | RMSE(t/ha) | 0.792        | 0.586        | 0.625        | 0.589        |
| MS bands, VIs, SPAD                  | $R^2$      | 0.780        | 0.851        | 0.854        | 0.855        |
|                                      | RMSE(t/ha) | 0.662        | 0.544        | 0.539        | 0.537        |

|                             |            |       |       |       |       |
|-----------------------------|------------|-------|-------|-------|-------|
| MS bands, VIs,<br>LAI       | $R^2$      | 0.854 | 0.840 | 0.822 | 0.841 |
|                             | RMSE(t/ha) | 0.539 | 0.564 | 0.595 | 0.562 |
| MS bands, VIs, LAI,<br>SPAD | $R^2$      | 0.793 | 0.843 | 0.828 | 0.837 |
|                             | RMSE(t/ha) | 0.642 | 0.558 | 0.584 | 0.570 |

The results indicate that optimal feature combinations varied across models, which may reflect the models' distinct algorithmic characteristics and their varying sensitivities to different types of input data (Wang et al., 2020b; Ruan et al., 2022). Furthermore, the findings demonstrate that simply increasing the number of data sources does not necessarily improve model performance. When all multimodal data were included, model accuracy generally declined, suggesting that some variables may provide redundant or weakly informative information, increasing model complexity and reducing generalization. Therefore, appropriate feature selection and data combination appear more important than indiscriminately incorporating all available data sources to improve yield prediction accuracy. Notably, each model's best-performing feature combination included the vegetation-index feature group, comprising the 14 selected vegetation indices. This further highlights the important role of vegetation indices in wheat yield prediction (Nguyen and De La Torre, 2010). By integrating information from multiple spectral bands, vegetation indices can effectively characterize crop canopy structure, chlorophyll content, and nitrogen status, all of which are closely associated with yield formation. Consequently, vegetation indices contributed substantially to prediction performance across the different models.

### 3.4 Yield prediction models based on a self-developed ensemble learning algorithm

In this study, the ensemble learning model combined different base models, each trained on a model-specific feature subset to leverage their respective strengths. Each base learner's input feature set based on the results presented in Sections 3.2 and 3.3, where model performance was compared across different growth stages and feature combinations. The SVR algorithm utilized vegetation indices from the entire growth period, while both the RF algorithm and the GBDT algorithm incorporated vegetation indices and SPAD values. The XGBoost algorithm used all data from the heading stage as its feature subset to increase information diversity. By selecting suitable feature subsets for different base models, this strategy allowed each model to better leverage its predictive strengths, thereby improving the ensemble model's overall performance. Ultimately, the proposed ensemble learning model achieved an  $R^2$  of 0.889 and an RMSE of 0.470 t/ha (Fig. 5a). Compared with the best-performing single model—that is, the GBDT model developed using vegetation indices and SPAD values from the entire wheat growth period ( $R^2 = 0.869$ , RMSE = 0.510 t/ha)—the ensemble model improved  $R^2$  by 2.3% and reduced RMSE by 7.8%. Furthermore, the control ensemble model without model-specific feature selection achieved an  $R^2$  of 0.859 and an RMSE of 0.529 t/ha (Fig. 5b). Compared with this control model, the proposed framework increased  $R^2$  from 0.859 to 0.889 and reduced RMSE from 0.529 to 0.470 t/ha. These results suggest that model-specific feature selection contributed substantially to improvements in the ensemble model's predictive performance. To verify the robustness of these results, the entire modeling procedure was repeated 10 times with different random seeds. The proposed ensemble model achieved an average  $R^2$  of  $0.878 \pm 0.018$  and an RMSE of  $0.490 \pm 0.030$  t/ha across the 10 repetitions (Supplementary Fig. 1), outperforming every model built using the feature subsets presented in Sections 3.2 and 3.3.

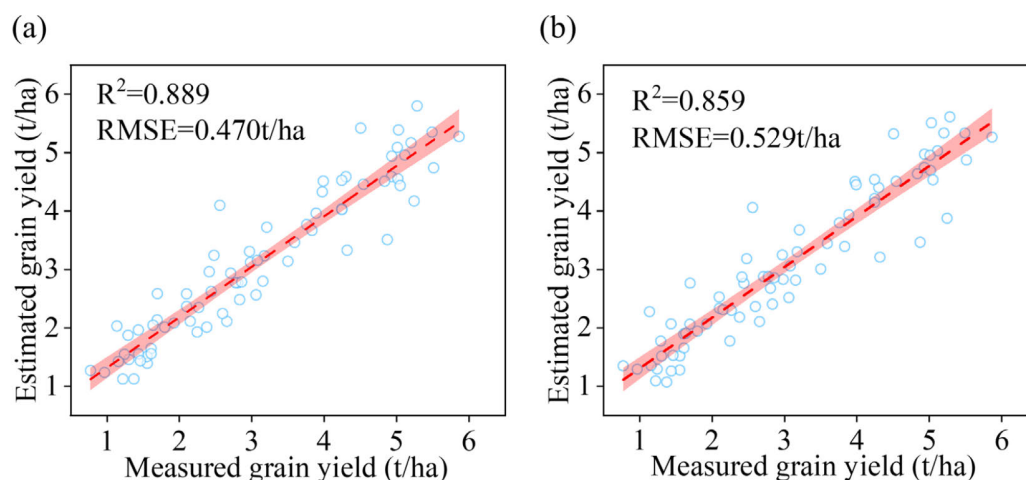


**Fig. 5. Performance comparison of ensemble learning models for wheat yield prediction: (a) with model-specific feature selection and (b) without model-specific feature selection.**

The proposed ensemble model's superior performance over the best individual model may be attributed to the ability of blending to integrate the complementary strengths of diverse base models through a meta-learner, reducing overfitting risk and improving robustness (Nti et al., 2020). The distinct feature sensitivities of individual algorithms—SVR performing best with vegetation indices alone, whereas GBDT requires vegetation indices combined with SPAD—underscore the importance of model-specific feature selection. Notably, the control ensemble without feature selection was outperformed by the best single model, suggesting that redundant features can dilute base learner complementarity (Demir, 2024a). In addition, the multi-partition weighted averaging strategy further enhanced fusion robustness by reducing sensitivity to specific data splits. Together, these design elements explain the proposed framework's superior performance.

### 3.5 Limitations and future perspectives

Although the ensemble learning models demonstrated promising performance in wheat yield estimation, several limitations warrant consideration. First, crop yield is affected by a range of factors, including farm management practices, weather conditions, and soil properties (Ravikumar et al., 2025). Understanding and incorporating these factors into prediction models are crucial for improving model performance. Second, collecting more diverse datasets is important for improving model accuracy and generalizability. Future studies should therefore focus on expanding dataset diversity to further enhance model performance. Finally, timely yield prediction is crucial, as early and accurate forecasts can significantly improve the flexibility and economic benefits of agricultural operations. Therefore, future studies should focus on developing more effective forecasting models that provide accurate, timely yield predictions.

## 4 Conclusions

This study developed an ensemble learning framework that integrates model-specific feature selection, multiple blending partitions, and performance-based weighted averaging for yield prediction across genetically diverse wheat germplasm accessions. The ensemble model outperformed each individual model, achieving an  $R^2$  of 0.889 and an RMSE of 0.470 t/ha. This approach reduces reliance on selecting a single prediction model and improves wheat yield prediction accuracy, showing potential for practical application across diverse wheat germplasm accessions. Additionally, the findings suggested that the heading stage may be the most suitable period for predicting wheat yield based on single-stage data, possibly because canopy structure is relatively well developed and plant physiological status is closely associated with subsequent yield formation at this stage.

### Data availability statement

The datasets used during the current study are available from the corresponding author on reasonable request.

### Acknowledgments

This research was supported by the National Key Research and Development Program of China(2021YFD2000102).

### Author contributions

Runzhou CAO: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing–original draft. Xuyang SUO: Data curation, Writing–review & editing. Xuping FENG: Resources, Supervision, Project administration, Writing–review & editing. Yong HE: Conceptualization, Supervision, Project administration, Writing–review & editing.

### Compliance with ethics guidelines

Runzhou CAO, Xuyang SUO, Xuping FENG, and Yong HE declare that they have no conflicts of interest. This article does not contain any studies with human or animal subjects performed by any of the authors.

### Declaration on the use of generative AI tools

During the preparation of this work, the authors used GPT to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

### References

- Almeida-Ñauñay AF, Tarquis AM, López-Herrera J, et al., 2023. Optimization of soil background removal to improve the prediction of wheat traits with uav imagery. *Computers and Electronics in Agriculture*, 205 <https://doi.org/10.1016/j.compag.2022.107559>
- Bian C, Shi H, Wu S, et al., 2022. Prediction of field-scale wheat yield using machine learning method and multi-spectral uav data. *Remote Sensing*, 14(6) <https://doi.org/10.3390/rs14061474>
- Chandel NS, Tiwari PS, Singh KP, et al., 2019. Yield prediction in wheat (*triticum aestivum* L.) using spectral reflectance indices. *Current Science*, 116(2) <https://doi.org/10.18520/cs/v116/i2/272-278>
- Chu C, Hu Y, Wu W, et al., 2025. Enhanced wheat yield estimation with uav multi-source remote sensing and optimized ensemble learning. *Smart Agricultural Technology*, 12 <https://doi.org/10.1016/j.atech.2025.101486>
- Da Silva EE, Rojo Baio FH, Ribeiro Teodoro LP, et al., 2020. Uav-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. *Remote Sensing Applications: Society and Environment*, 18 <https://doi.org/10.1016/j.rsase.2020.100318>
- Demir S, 2024a. Determination of suitable agricultural areas and current land use in isparta province, türkiye, through a linear combination technique and geographic information systems. *Environment, Development and Sustainability*, 26(5):13455–13493. <https://doi.org/10.1007/s10668-023-04359-7>
- Demir S, 2024b. Prediction of canopy cover for agricultural land classification in land parcel identification system (Ipis) data using planet-scope multispectral images: A case study of gelendost district. *Black Sea Journal of Agriculture*, 7(4):407–417. <https://doi.org/10.47115/bsagriculture.1490400>
- Demir S, Dedeoğlu M, Başayığit L, 2024. Yield prediction models of organic oil rose farming with agricultural unmanned aerial vehicles (uavs) images and machine learning algorithms. *Remote Sensing Applications: Society and Environment*, 33 <https://doi.org/10.1016/j.rsase.2023.101131>
- Demir S, Dursun İ, 2024. Assessment of pre- and post-fire erosion using the rusle equation in a watershed affected by the forest fire on google earth engine: The study of manavgat river basin. *Natural Hazards*, 120(3):2499–

2527. <https://doi.org/10.1007/s11069-023-06291-5>
- Deng L, Li Y, Liu X, et al., 2025. Sorghum yield prediction using uav multispectral imaging and stacking ensemble learning in arid regions. *Frontiers in Plant Science*, 16 <https://doi.org/10.3389/fpls.2025.1636015>
- Du Z, Yang L, Zhang D, et al., 2022. Corn variable-rate seeding decision based on gradient boosting decision tree model. *Computers and Electronics in Agriculture*, 198 <https://doi.org/10.1016/j.compag.2022.107025>
- Dursun İ, Demir S, Başayığit L, et al., 2025. Post-fire erosion dynamics in the dim river basin: A remote sensing and google earth engine approach. *Fire Ecology*, 21(1) <https://doi.org/10.1186/s42408-025-00363-6>
- Esmaili M, Abbasi-Moghadam D, Sharifi A, et al., 2023. Hyperspectral image band selection based on cnn embedded ga (cnnega). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1927–1950. <https://doi.org/10.1109/jstars.2023.3242310>
- Fan Z, Yu Z, Yang K, et al., 2025. Diverse models, united goal: A comprehensive survey of ensemble learning. *CAAI Transactions on Intelligence Technology*, 10(4):959–982. <https://doi.org/10.1049/cit2.70030>
- Fei S, Hassan MA, Xiao Y, et al., 2022. Application of multi-layer neural network and hyperspectral reflectance in genome-wide association study for grain yield in bread wheat. *Field Crops Research*, 289 <https://doi.org/10.1016/j.fcr.2022.108730>
- González-Vivar J, Demir S, Rodrigo-Comino J, 2026. Improving yield prediction and decision-making in mediterranean vineyards using machine learning on multisource remote sensing and soil data. *Air, Soil and Water Research*, 19 <https://doi.org/10.1177/11786221261432379>
- Han J, Zhang Z, Cao J, et al., 2020. Prediction of winter wheat yield based on multi-source data and machine learning in china. *Remote Sensing*, 12(2) <https://doi.org/10.3390/rs12020236>
- Han X, Wei Z, Chen H, et al., 2021. Inversion of winter wheat growth parameters and yield under different water treatments based on uav multispectral remote sensing. *Frontiers in Plant Science*, 12 <https://doi.org/10.3389/fpls.2021.609876>
- Hasan M, Abedin MZ, Hajek P, et al., 2024. A blending ensemble learning model for crude oil price forecasting. *Annals of Operations Research*, 353(2):485–515. <https://doi.org/10.1007/s10479-023-05810-8>
- Huang J, Ma H, Sedano F, et al., 2019. Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled wofost–prosail model. *European Journal of Agronomy*, 102:1–13. <https://doi.org/10.1016/j.eja.2018.10.008>
- Hunter MC, Smith RG, Schipanski ME, et al., 2017. Agriculture in 2050: Recalibrating targets for sustainable intensification. *BioScience*, 67(4):386–391. <https://doi.org/10.1093/biosci/bix010>
- Jalayer S, Sharifi A, Abbasi-Moghadam D, et al., 2023. Assessment of spatiotemporal characteristic of droughts using in situ and remote sensing-based drought indices. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1483–1502. <https://doi.org/10.1109/jstars.2023.3237380>
- Johnson MD, Hsieh WW, Cannon AJ, et al., 2016. Crop yield forecasting on the canadian prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and Forest Meteorology*, 218-219:74–84. <https://doi.org/10.1016/j.agrformet.2015.11.003>
- Kaya F, Demir S, Dedeoğlu M, et al., 2026. Multi-crop yield estimation and spatial analysis of agro-climatic indices based on high-resolution climate simulations in türkiye's lakes region, a typical mediterranean biogeography. *Agronomy*, 16(3) <https://doi.org/10.3390/agronomy16030321>
- Kaya Y, Polat N, 2023. A linear approach for wheat yield prediction by using different spectral vegetation indices. *International Journal of Engineering and Geosciences*, 8(1):52–62. <https://doi.org/10.26833/ijeg.1035037>
- Li L, Wang B, Feng P, et al., 2022. Developing machine learning models with multi-source environmental data to predict wheat yield in china. *Computers and Electronics in Agriculture*, 194 <https://doi.org/10.1016/j.compag.2022.106790>
- Li Q-C, Xu S-W, Zhuang J-Y, et al., 2023. Ensemble learning prediction of soybean yields in china based on meteorological data. *Journal of Integrative Agriculture*, 22(6):1909–1927. <https://doi.org/10.1016/j.jia.2023.02.011>
- Liu J, Zhu Y, Tao X, et al., 2022. Rapid prediction of winter wheat yield and nitrogen use efficiency using consumer-

- grade unmanned aerial vehicles multispectral imagery. *Frontiers in Plant Science*, 13 <https://doi.org/10.3389/fpls.2022.1032170>
- Lou Z, Lu X, Li S, 2024. Yield prediction of winter wheat at different growth stages based on machine learning. *Agronomy*, 14(8) <https://doi.org/10.3390/agronomy14081834>
- Marshall M, Belgiu M, Boschetti M, et al., 2022. Field-level crop yield estimation with prisma and sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing*, 187:191–210. <https://doi.org/10.1016/j.isprsjprs.2022.03.008>
- Monostori I, Árendás T, Hoffman B, et al., 2016. Relationship between spad value and grain yield can be affected by cultivar, environment and soil nitrogen content in wheat. *Euphytica*, 211(1):103–112. <https://doi.org/10.1007/s10681-016-1741-z>
- Nejad SMM, Abbasi-Moghadam D, Sharifi A, et al., 2023. Multispectral crop yield prediction using 3d-convolutional neural networks and attention convolutional lstm approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:254–266. <https://doi.org/10.1109/jstars.2022.3223423>
- Nguyen MH, De La Torre F, 2010. Optimal feature selection for support vector machines. *Pattern Recognition*, 43(3):584–591. <https://doi.org/10.1016/j.patcog.2009.09.003>
- Niraimathi J, Saravanan S, 2025. Integration of multi-source remote sensing data and ensemble machine learning for spatial paddy yield estimation in the cauvery delta, india. *Earth Systems and Environment*, 10(3):3609–3632. <https://doi.org/10.1007/s41748-025-00838-1>
- Nti IK, Adekoya AF, Weyori BA, 2020. A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, 7(1) <https://doi.org/10.1186/s40537-020-00299-5>
- Panda SS, Ames DP, Panigrahi S, 2010. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing*, 2(3):673–696. <https://doi.org/10.3390/rs2030673>
- Ravikumar A, Periyasamy V, Devi KMK, et al., 2025. A hybrid serwi ensemble model for crop yield prediction using an inverse rmse weighting strategy. *Scientific Reports*, 15(1) <https://doi.org/10.1038/s41598-025-31987-y>
- Ruan G, Li X, Yuan F, et al., 2022. Improving wheat yield prediction integrating proximal sensing and weather data with machine learning. *Computers and Electronics in Agriculture*, 195 <https://doi.org/10.1016/j.compag.2022.106852>
- Sabir K, Rose T, Wittkop B, et al., 2023. Stage-specific genotype-by-environment interactions determine yield components in wheat. *Nature Plants*, 9(10):1688–1696. <https://doi.org/10.1038/s41477-023-01516-8>
- Sun B, Liang Y-C, Zhou Y, et al., 2023. Ensemble learning based on stacking and blending predicts glass forming ability. *Materials Today Communications*, 37 <https://doi.org/10.1016/j.mtcomm.2023.107385>
- Tanabe R, Matsui T, Tanaka TST, 2023. Winter wheat yield prediction using convolutional neural networks and uav-based multispectral imagery. *Field Crops Research*, 291 <https://doi.org/10.1016/j.fcr.2022.108786>
- Van Klompenburg T, Kassahun A, Catal C, 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177 <https://doi.org/10.1016/j.compag.2020.105709>
- Wan L, Cen H, Zhu J, et al., 2020. Grain yield prediction of rice using multi-temporal uav-based rgb and multispectral images and model transfer – a case study of small farmlands in the south of china. *Agricultural and Forest Meteorology*, 291 <https://doi.org/10.1016/j.agrformet.2020.108096>
- Wan L, Ryu Y, Dechant B, et al., 2024. Improving retrieval of leaf chlorophyll content from sentinel-2 and landsat-7/8 imagery by correcting for canopy structural effects. *Remote Sensing of Environment*, 304 <https://doi.org/10.1016/j.rse.2024.114048>
- Wang X, Huang J, Feng Q, et al., 2020a. Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of china with deep learning approaches. *Remote Sensing*, 12(11) <https://doi.org/10.3390/rs12111744>
- Wang Y, Zhang Z, Feng L, et al., 2020b. Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous united states. *Remote Sensing*, 12(8) <https://doi.org/10.3390/rs12081232>
- Wu S, Han S, Zhang J, et al., 2026. Multi-scale keypoints detection and motion features extraction in dairy cows using

- resnet101-aspp network. *Journal of Integrative Agriculture*, 25(5):2028–2040. <https://doi.org/10.1016/j.jia.2024.07.023>
- Yang B, Zhu W, Rezaei EE, et al., 2022. The optimal phenological phase of maize for yield prediction with high-frequency uav remote sensing. *Remote Sensing*, 14(7) <https://doi.org/10.3390/rs14071559>
- Yu R, Cao X, Liu J, et al., 2024. Using uav-based temporal spectral indices to dissect changes in the stay-green trait in wheat. *Plant Phenomics*, 6 <https://doi.org/10.34133/plantphenomics.0171>
- Zhang Y, Hui J, Qin Q, et al., 2021. Transfer-learning-based approach for leaf chlorophyll content estimation of winter wheat from hyperspectral data. *Remote Sensing of Environment*, 267:112724. <https://doi.org/https://doi.org/10.1016/j.rse.2021.112724>

### Supplementary information

Supplementary Table 1, Supplementary Table 2 and Supplementary Fig. 1

unedited