

RESEARCH

Open Access



Improving the prediction performance of leaf water content by coupling multi-source data with machine learning in rice (*Oryza sativa* L.)

Xuenan Zhang^{1†}, Haocong Xu^{1†}, Yehong She¹, Chao Hu¹, Tiezhong Zhu¹, Lele Wang¹, Liqun Wu^{1,2*}, Cuicui You^{1,3}, Jian Ke^{1,3}, Qiangqiang Zhang¹ and Haibing He^{1,3,4*}

Abstract

Background Leaf water content (LWC) significantly affects rice growth and development. Real-time monitoring of rice leaf water status is essential to obtain high yield and water use efficiency of rice plants with precise irrigation regimes in rice fields. Hyperspectral remote sensing technology is widely used in monitoring crop water status because of its rapid, nondestructive, and real-time characteristics. Recently, multi-source data have been attempted to integrate into a monitored model of crop water status based on spectral indices. However, there are fewer studies using spectral index model coupled with multi-source data for monitoring LWC in rice plants. Therefore, 2-year field experiments were conducted with three irrigation regimes using four rice cultivars in this study. The multi-source data, including canopy ecological factors and physiological parameters, were incorporated into the vegetation index to accurately predict LWC in rice plants.

Results The results presented that the model accuracy of rice LWC estimation after combining data from multiple sources improved by 6–44% compared to the accuracy of a single spectral index normalized difference index (ND). Additionally, the optimal prediction accuracy of rice LWC was produced using a machine algorithm of gradient boosted decision tree (GBDT) based on the combination of ND_(1287,1673) and crop water stress index (CWSI) ($R^2 = 0.86$, RMSE = 0.01).

Conclusions The machine learning estimation model constructed based on multi-source data fully utilizes the spectral information and considers the environmental changes in the crop canopy after introducing multi-source data parameters, thus improving the performance of spectral technology for monitoring rice LWC. The findings may be helpful to the water status diagnosis and accurate irrigation management of rice plants.

Keywords Rice (*Oryza sativa* L.), Leaf water content, Hyperspectral remote sensing, Machine learning

[†]Xuenan Zhang and Haocong Xu contributed equally to this study.

*Correspondence:

Liqun Wu

wlq-001@163.com

Haibing He

hhb_agr@ahau.edu.cn

Full list of author information is available at the end of the article



Introduction

Rice is an important staple food worldwide. Rice is the largest water-consuming crop in the central rice production regions in the world, and water scarcity is bound to threaten rice production [1]. Additionally, the water consumption in paddy fields would increase significantly due to global climate change in the future. The unfavorable factors must aggravate the water crisis in rice fields [2, 3]. Leaf water content (LWC) is an important evaluation index for crop water demand status [4]. Monitoring LWC effectively achieves precise irrigation, improves water utilization, and alleviates the water crisis.

Hyperspectral remote sensing technology has recently been widely used in agricultural production due to its rapid, nondestructive, and real-time monitoring of crop physiological and biochemical characteristics [5–7]. Early LWC spectral modeling studies improved prediction accuracy by constructing various types of vegetation indices from sensitive spectral bands, such as the moisture stress index (MSI) and water index (WI) [8, 9]. Previous spectral monitoring studies have initially focused on vegetation indices derived from spectral bands. Researchers construct different vegetation indices by spectrally sensitive bands to improve the prediction accuracy of LWC [10, 11]. Vegetation indices are frequently screened by selecting spectral information and have a good correlation with LWC at multiple points over many years [12, 13], but the established models usually ignore the effects of the growing environment and growth characteristics on LWC of the crops. Climate characteristics and physiological status of the plants are important factors affecting LWC [14, 15]. The neglected information could be the primary reason for the poor generalizability of existing vegetation index models in practical applications. Therefore, integrating the actors into the spectral monitoring models of LWC would be more high precision due to considering the potential effects of the growing environment and growth characteristics on LWC [16, 17]. Qin et al. [18] have recently improved crop nitrogen content prediction accuracy using a spectral model that combines image feature parameters and fluorescence parameters. These results supported our inferences that introducing multi-source data may be important means to improve the monitoring accuracy of LWC spectral modeling in rice plants.

Choosing modeling methods are one of key steps in building monitoring models. Vegetation index monitoring models have been usually established using linear and nonlinear functions [19, 20]. These traditional monitoring models do not represent the complex relationship between various indicators. Machine

learning algorithms with sophisticated functionality and the ability to handle complex relationships between predictors and target variables can be a good solution to this problem [21, 22]. The accuracy and stability of models have significantly improved with the rapid development of machine learning algorithms, and these are widely adopted when establishing rice nitrogen nutrition monitoring models [23–25]. Moreover, machine learning algorithms allow the use of different classes of sample data as input variables, allowing multiple sources of data (physiological and ecological indicators and spectral information) to be effectively coupled and can effectively discriminate between differences in the contributions of the input variables, allowing the model's parameters can be fully utilized. Therefore, we speculated that the monitoring capacities of vegetation index models integrating with multi-source data would also have significant advantages using machine learning algorithms when predicting the LWC of rice plants based on the results previous studies [16, 26].

Three irrigation regimes and four varieties with different drought tolerance capacities were established LWC difference populations in field experiments. This study's objectives are to (1) select spectrally sensitive bands with high correlation with LWC at multiple growth stages; (2) build a new vegetation index model of LWC integrating multi-source data parameters after selecting key physiological and ecological indicators; and (3) use a machine learning algorithm to optimize the coupling model and select the optimal algorithm to monitor the LWC.

Materials and methods

Site and treatment description

Three irrigation regimes, including two water-saving irrigation regimes named mild alternate dry and wet irrigation (MADW), severe alternate dry and wet irrigation (SADW) and traditional irrigation regime (CK), were conducted with four rice cultivars in Anhui province of China in 2021 and 2022 (Fig. 1). The planting and sampling plan details and soil characteristics are presented in Table 1. The climate conditions are displayed in Fig. 2. The irrigation regimes were conducted when rice seedlings were planted in plots. The irrigation criterion referenced the classical definition for the MADW and SADW treatments [27]. The supplementary irrigation with 2–3 cm water layer was carried out when the soil water potential at 20 cm soil layer reached -15 KPa and -30 KPa in the MADW and SADW treatments during whole growth stages of rice plants, respectively, across cultivars, years, and locations. The soil water potential of each plot was monitored by a tensiometer (Watermark, Irrrometer Company Riverside, CA, USA). Moreover,

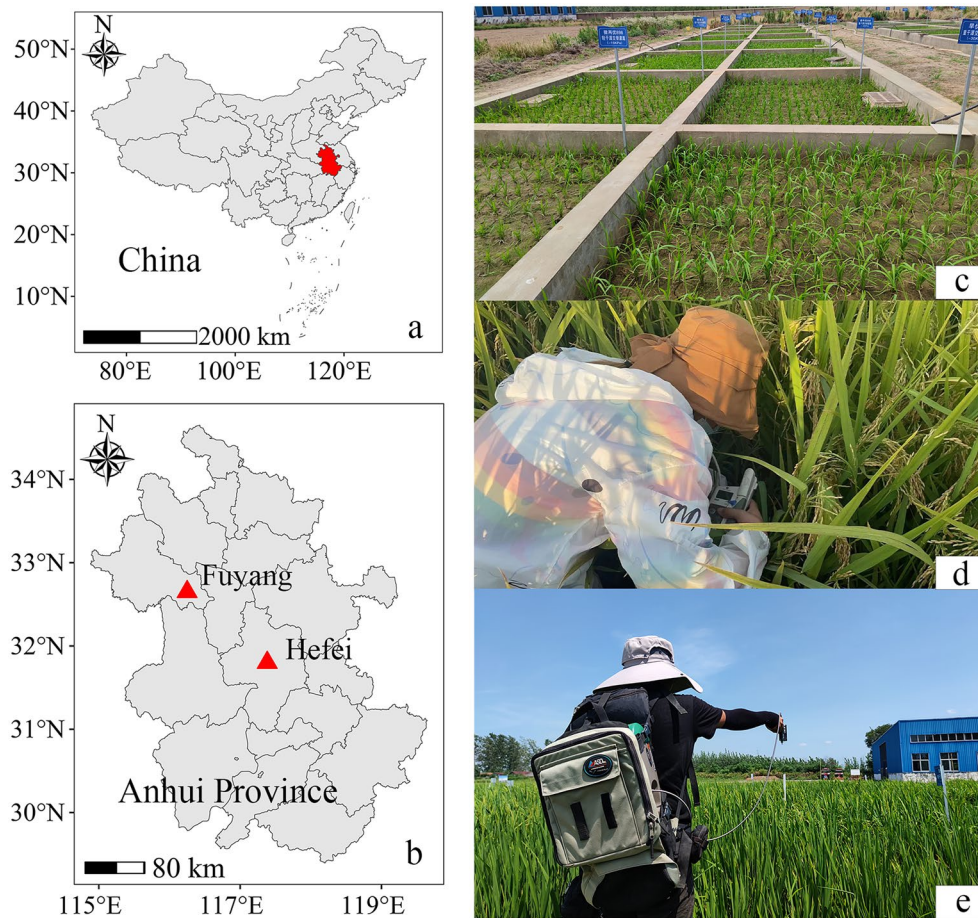


Fig. 1 Sketch of field experiments (c) in Anhui province (b), China (a) using ASD Field Spec 4 (e) and SPAD (d) devices

2–3 cm water later was always maintained for the CK treatment during whole growth stages. Generally, 9–14, 5–8, and 18–24 times irrigated practices were applied for the MADW, SADW, and CK treatments, respectively, across cultivars, years, and locations.

The nitrogen fertilizer was divided into three applications for each treatment: 40% as a basal fertilizer, 30% at the early tiller stage, and 30% at the panicle differentiation stage. The nitrogen fertilizer was applied at the rate of 240 kg ha⁻¹. Phosphorus and potassium fertilizers (P₂O₅ 75 kg ha⁻¹ and K₂O 225 kg ha⁻¹) were also applied as basal fertilizers. Plot sizes for each experiment were 12 m² and 40 m² in Hefei and Fuyang, respectively. All experimental rows were spaced 30.0 cm apart, and plants in a row were spaced 13.3 cm apart.

Measurements

Hyperspectral and vegetation index

Canopy reflectance spectra of rice plants were collected for the booting, flowering, initial grain filling, and middle grain filling stages using spectral scanning equipment

(ASD Field Spec 4, Boulder, CO, USA). The band amplitude of the device ranged from 350 to 2500 nm. The measurements were conducted under clear and cloudless sky conditions between 10:00 and 14:00 [16]. After measuring 10 times for each sample, the average value was calculated, and the reference plate was used to correct the instrument every 15 min. The representative hills were monitored in each plot for all treatments at each observed time.

Numerous vegetation indices have been applied to monitoring crop water content. The normalized difference vegetation index (NDVI) has been widely utilized in monitoring plant water using remote sensing methods due to its simple construction and effective improvement of spectral monitoring accuracy with multi-band analysis compared with other vegetation indices. Thus, this study used canopy spectral data to construct the normalized difference index (ND) and five traditional vegetation indices. The calculation equations were illustrated in Table 2.

Table 1 Basic information of the three experiments

| Experiment | Year | Cultivar | Soil water potential | Sampling stage | Sowing/Harvesting date | Soil characteristics |
|---------------------------------------|------|--|-------------------------------|---|-----------------------------|--|
| Exp.1 (Hefei, 31° 48' N, 117° 23' E) | 2021 | Hanyou 73 (HY-73) Huanghuazhan (HHZ) Huaidao 5 (HD-5) | 0 KPa − 15 KPa − 30 KPa | Booting Flowering Initial grain filling Middle grain filling | July 5th / November 20th | Type: sandy loam soil Organic matter: 14.81 g kg ^{−1} Total N: 0.97 g kg ^{−1} Available P: 21.06 mg kg ^{−1} Available K: 96.29 mg kg ^{−1} |
| Exp.2 (Hefei, 31° 48' N, 117° 23' E) | 2022 | Hanyou 73 (HY-73) Huanghuazhan (HHZ) Huiliangyou 898 (HLY-898) | 0 KPa − 15 KPa − 30 KPa | Booting Flowering Initial grain filling Middle grain filling | May 23th/ September 26th | Type: sandy loam soil Organic matter: 15.39 g kg ^{−1} Total N: 1.00 g kg ^{−1} Available P: 22.67 mg kg ^{−1} Available K: 98.10 mg kg ^{−1} |
| Exp.3 (Fuyang, 32° 39' N, 116° 15' E) | 2022 | Hanyou 73 (HY-73) Huanghuazhan (HHZ) | 0 KPa | Booting Flowering Initial grain filling Middle grain filling | May 26th/ September 29th | Type: sandy loam soil Organic matter: 17.77 g kg ^{−1} Total N: 0.95 g kg ^{−1} Available P: 14.18 mg kg ^{−1} Available K: 65.23 mg kg ^{−1} |

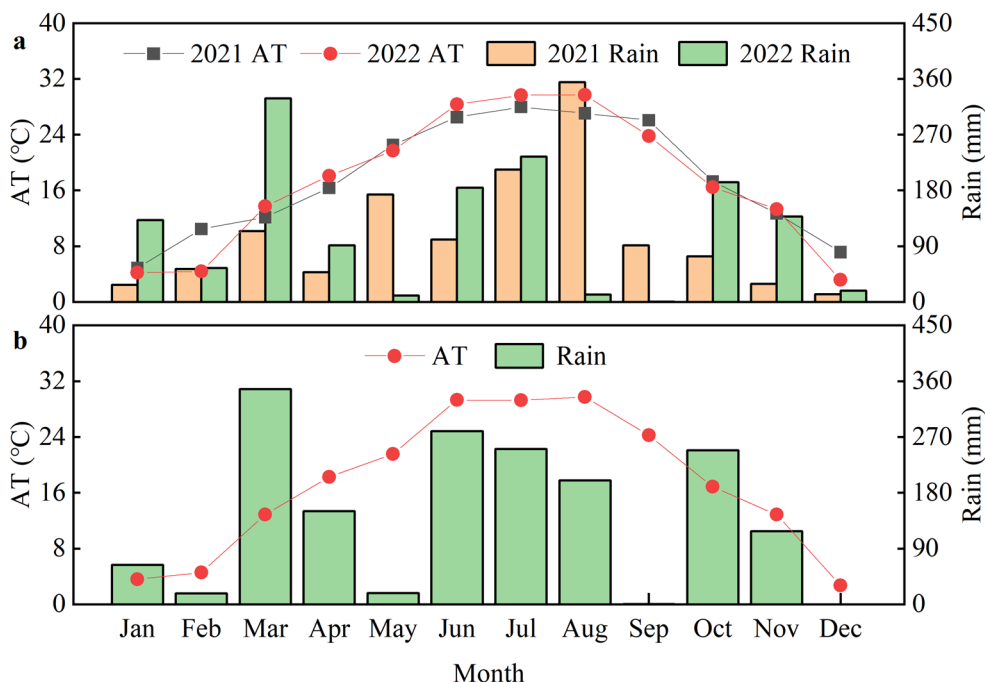


Fig. 2 Climatic conditions at Wanzhong Comprehensive Experimental Station during the 2021 and 2022 seasons (a). Climatic conditions at Yingshang Agricultural Green Development Experiment Station during 2022 (b). AT: monthly average temperature; Rain: monthly accumulated rainfall

Table 2 List of vegetation indices used in this study

| Vegetation index | Algorithm | Reference |
|---|--|------------|
| Normalized difference index (ND) | $ND = (R_{\lambda_1} - R_{\lambda_2}) / (R_{\lambda_1} + R_{\lambda_2})$ | This study |
| Normalized difference vegetation index (NDVI) | $NDVI = (R_{895} - R_{675}) / (R_{895} + R_{675})$ | [28] |
| Normalized difference infrared index (NDII) | $NDII = (R_{819} - R_{1600}) / (R_{819} + R_{1600})$ | [29] |
| Normalized difference water index (NDWI) | $NDWI = (R_{860} - R_{1240}) / (R_{860} + R_{1240})$ | [30] |
| Water index (WI) | $WI = R_{970} / R_{900}$ | [9] |
| Moisture stress index (MSI) | $MSI = R_{1600} / R_{820}$ | [8] |

The λ_1 and λ_2 are arbitrary bands in the 350–2500 nm bands, and R_{λ_1} and R_{λ_2} are the spectral reflectance of the bands

SPAD value and chlorophyll fluorescence parameters

SPAD value of rice leaves was monitored using a chlorophyll meter (Konica Minolta Company, SPAD-520 plus, Japan). The average value of the top three fully expanded leaves was defined as the SPAD value of the whole plant. The five independent plants were measured in each plot across cultivars, water treatments, and observed periods.

The leaves used to measure SPAD value were chosen, and chlorophyll fluorescence parameters of rice leaves were determined using a portable pulse-modulated chlorophyll fluorometer (WALZ Company, PAM-2500, Germany). These leaves were used to determine the minimum fluorescence level (F_0) and the maximum fluorescence level (F_m) of dark-adapted leaves at dawn for all cultivars and treatments. The steady-state F_0 was measured using far-red light with illumination less than $1 \text{ mol m}^{-2} \text{ s}^{-1}$. Then, a 0.80-s saturating pulse with $8000 \text{ mol m}^{-2} \text{ s}^{-1}$ PAR was supplied to determine the F_m . Moreover, the steady-state fluorescence yield (F_s) was recorded at the forenoon with clear and cloudless sky conditions. Then, a 0.80-s saturating pulse with $8000 \text{ mol m}^{-2} \text{ s}^{-1}$ PAR was stimulated to obtain the maximal fluorescence level (F_m'). Finally, the maximum photochemical efficiency (F_v/F_m) and actual photochemical efficiency ($Y(II)$) were calculated:

$$F_v/F_m = (F_m - F_0)/F_m \quad (1)$$

$$Y(II) = (F_m' - F_s)/F_m' \quad (2)$$

Leaf water content (LWC)

The fresh leaf mass (FW) was determined by weighing the top three fully expanded leaves immediately. Five hills were sampled in each plot for four cultivars and three water treatments during each observed period. Fresh leaves were weighed and dried at $80 \text{ }^\circ\text{C}$ to constant weight at each sampling. The weight was marked as dry mass (DW). The LWC was calculated as follows:

$$LWC = (FW - DW)/FW \quad (3)$$

Crop water stress index (CWSI)

CWSI was calculated as follows [31, 32]:

$$CWSI = \frac{(T_c - T_a) - T_{\min}}{T_{\max} - T_{\min}} \quad (4)$$

$$T_{\min} = A + B \times VPD \quad (5)$$

$$T_{\max} = A + B \times VPG \quad (6)$$

$$VPD = 0.61 \times e^{\frac{17.27T_a}{T_a+237.3}} \times \left(1 - \frac{RH}{100}\right) \quad (7)$$

where T_c is the crop canopy temperature ($^\circ\text{C}$), and T_a is the air temperature ($^\circ\text{C}$). T_{\min} is the lower limit of the canopy and air temperature difference ($^\circ\text{C}$), T_{\max} is the upper limit of the canopy and air temperature difference ($^\circ\text{C}$), and VPD is the air saturation water vapor pressure deficit (KPa). VPG is the difference between the air saturation water vapor pressure and VPD when the air temperature is T_a and $(T_a + A)$, respectively. A and B are linear regression coefficients. RH is the relative humidity of air (%).

In this study, a HOBO UX100-003 temperature and humidity recorder (Onset, USA) was placed in each plot to automatically record rice canopy temperature and humidity every 1 h interval throughout the day.

Leaf area index (LAI), above-ground biomass (Biomass), and Grain yield

Five representative hills, which plants in the hills had uniform growth capacities, were randomly selected from each plot at each sampling time across all cultivars and irrigation treatments. The hills with excessively vigorous or weak plants were not sampled to minimize sampling error. All green leaves per hill were scanned using a portable leaf area meter (CI-203, CID Inc., USA). The averaged leaf area in a each plot was then calculated from the five hills and marked with D (cm^2). Finally, the scanned green leaf, remaining yellow

leaves, stem, and panicle organs were dried at 80 °C to constant weight to assess the biomass of each plot.

$$\text{LAI} = \frac{D \times \rho}{10000} \quad (8)$$

where ρ is the planting density per square meter (hill m^{-2}).

Plants from a 2 m^2 area were harvested at maturity to calculate the actual grain yield with 13.5% moisture content in each plot across cultivars and water treatments.

Regression models

Some common and classic machine learning algorithms were preliminary assessed to select elite algorithms with high monitoring abilities to LWC of rice plants. Finally, the four machine learning algorithms named decision tree regression (DT), random forest regression (RF), K-nearest neighbor regression (KNN), and gradient boosting decision tree regression (GBDT) were adopted in this study. The four selected methods were successfully applied on estimating various ecological parameters such as water content of wheat plants and soil moisture of wheat fields [16, 33]. In addition, the scikit-learn packages of the four algorithms were from Python 3.8 software (<https://scikit-learn.org/stable/index.html>).

Multiple linear regression (MLR)

MLR is the most basic and commonly used method for combining two or more independent variables that jointly predict or estimate the dependent variable [34]. The y is the dependent variable. The x_1, x_2, \dots, x_k are the independent variables. The multiple linear regression was calculated as follows:

$$y = b_0 + b_1x_1 + \dots + b_kx_k + e \quad (9)$$

where b_0 is the constant term, e is the error term, and b_1, b_2, \dots, b_k are the regression coefficients. When x_1, x_2, \dots , and x_k are fixed, $b_{1(k)}$ represents the effect of increase or decrease in $x_{1(k)}$ on y for each unit or named the partial regression coefficient of $x_{1(k)}$ on y .

Decision tree regression (DT)

DT is a way to infer classification rules as a decision tree from a set of unordered and irregular data, using a top-down recursive approach to compare attribute values of nodes inside the decision tree [35]. Each internal node is a splitting problem in a decision tree. A test for some instance attribute is specified, and it splits the samples arriving at that node according to a particular attribute. Each subsequent branch of the node corresponds to one of the possible values of that attribute. The prediction

results are the average values of the output variables in the samples contained in the leaf nodes of the regression tree. In this study, the maximum depth of the tree was set as 10, and the number of trees was 100.

Random forest regression (RF)

RF resulted from random sampling from sample observations and feature variables of the modeled data among many decision trees; each sampling result is a tree [36]. Meanwhile, each tree generated rules and judgment values that match its properties. Finally, the forest algorithm integrated the rules and judgment values of all decision trees to achieve random forest regression.

K-nearest neighbor regression (KNN)

KNN was predicted by computing the spatial similarity relationship between the k nearest neighbors and the predictor. The algorithm was frequently used for classification problems in the early stage and gradually applied to parameter estimation [37]. The primary principle of the KNN algorithm is that a prediction sample has K nearest neighbors in the feature space. Then, the class of the prediction sample was usually determined by the majority class of the K nearest neighbors. The K data set value was chosen appropriately according to the samples. The model was simplified, and useful information was lost if the data set was too large. Oppositely, the model would be over-fitted if the data set is too small. The K data set was defined as 3 in this study.

Gradient boosting decision tree regression (GBDT)

GBDT is an improved algorithm based on the booting algorithm [38]. The booting algorithm assigned the same weight to each training sample in the initial training and then increased the weights of the error points after each training session to generate multiple base learners. Finally, these base learners were combined, and the model was formed using weighting or voting approaches. The difference between gradient boosting tree regression and classification algorithms was that the input training data was residual. The previous prediction was incorporated into the residual to find the training data for the current round instead of the gradient of the loss function.

Data analysis and model verification

The Pearson correlation coefficient (PCC) helps to measure the linear correlation between variables [39]. In this study, PCC was simply regarded as supplementary means to eliminate unimportant indicators with low correlation coefficient and to obtain eigenvalues with high correlation coefficient. This meant that indicators closely related to LWC were used to build models to improve performance.

After removing invalid samples from the two-year experiments from 2021 to 2022, 91 sample data for each growth stage were obtained. In multiple linear regression (MLR), 2022 data were used for modeling (55), whereas 2021 data were used for validation (36). The measured data were randomly divided into training (70%) and testing (30%) sets in the machine learning regression algorithm. The regression model's accuracy was evaluated using the determination coefficient (R^2) and root mean square error (RMSE). The overall model was evaluated in the graph, including linear regression and a 1:1 dash-line, to determine the relationship between the predicted and measured values. The calculation equations are presented in (9)–(10):

$$R^2 = \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{10}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}} \tag{11}$$

where x_i is the measured value of LWC, \bar{x} is the measured mean value of LWC, \hat{x}_i is the predicted value of the model, and n is the sample size. The larger the R^2 value, the better the accuracy of the model. The RMSE reflects the degree of dispersion and deviation between the model's predicted and true values. The smaller the value, the better the prediction of the model.

The workflow of the LWC prediction procedures is illustrated in Fig. 3. Data preprocessing included the following processes: (1) normal distribution test and scatter plots were performed for the original data such as vegetation index parameters, leaf water content, and physiological and ecological indexes (Additional file 1: Fig. S1–S4). (2) Outliers detection of all parameters were shown in Additional file 1: Table S1. (3) The multicollinearity test between ND and SPAD, between ND and Fv/Fm, between ND and CWSI at whole observed stages were also adopted by variance inflation factor and tolerance values (Additional file 1: Table S2).

Results

The classical and new vegetation index for monitoring LWC

The sensitive bands of LWC were analyzed and tested using the normalized difference vegetation index formula at the booting, flowering, initial grain filling, and middle grain filling stages to improve the accuracy of the hyperspectral monitoring model for LWC. The classical vegetation indices, such as NDII and MSI, had a significant positive correlation with LWC and high predictive abilities (Table 3). The vegetation index ND constructed by the screened sensitive bands (1287 and 1673 nm) also had a significant positive correlation with LWC (Fig. 4). The R^2 in the ND_(1287,1673) model was higher than that in NDII and MSI vegetation indices at each growth stage. These results indicated that vegetation index ND was an optimal predicted model at each growth stage of rice

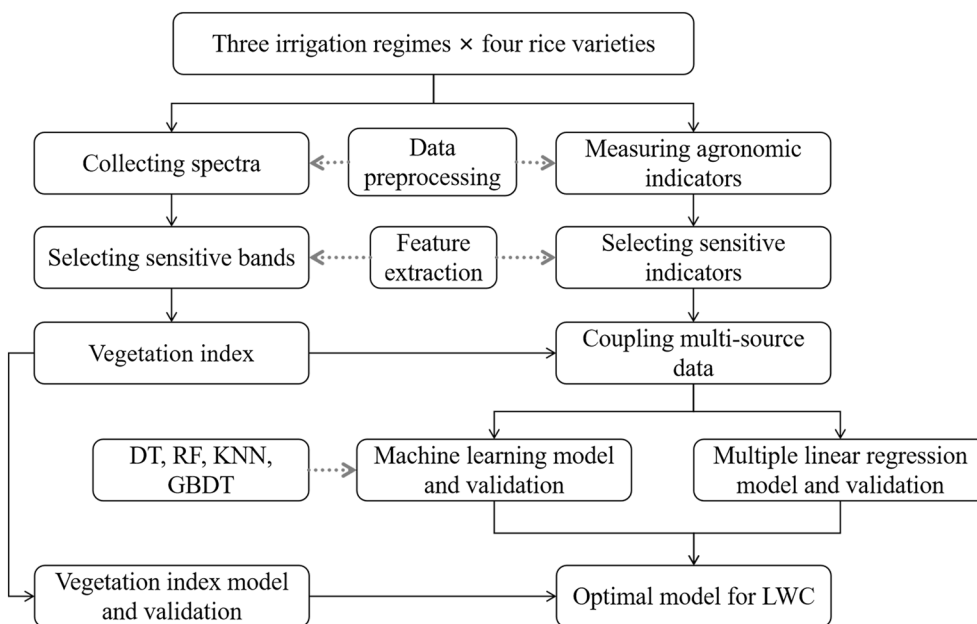


Fig. 3 Flow chart of this study

Table 3 Coefficient of determination (R^2) of predicting LWC by different vegetation index at different growth stages in rice plants

| Growth stage | ND _(1287,1673) | NDVI | NDII | NDWI | WI | MSI |
|-----------------------|---------------------------|--------|--------|--------|--------|--------|
| Booting | 0.48** | 0.01 | 0.26** | 0.00 | 0.01 | 0.22** |
| Flowering | 0.64** | 0.02 | 0.48** | 0.06 | 0.27** | 0.47** |
| Initial grain filling | 0.57** | 0.24** | 0.30** | 0.03 | 0.09* | 0.30** |
| Middle grain filling | 0.53** | 0.01 | 0.34** | 0.32** | 0.34** | 0.34** |

* and ** indicate significant correlation at 5% and 1% probability level, respectively. ND_(1287,1673) is the normalized difference index; NDVI is the normalized difference vegetation index; NDII represents the normalized difference infrared index; NDWI is the normalized difference water index; WI shows the water index; MSI represents the moisture stress index

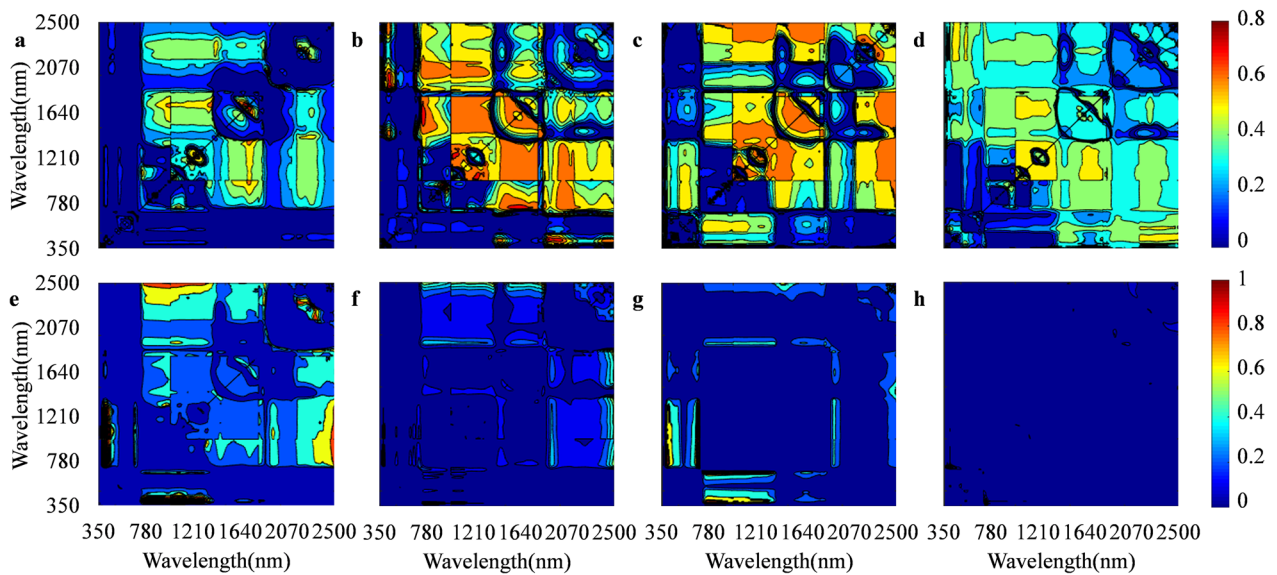


Fig. 4 Image map of the coefficient of determination (R^2) and coefficient of determination (RMSE) for screening sensitive band combinations of two wavebands at different growth stages of rice plants. **a-d** and **e-h** represent R^2 and RMSE of the booting, flowering, initial grain filling, and middle grain filling stages, respectively

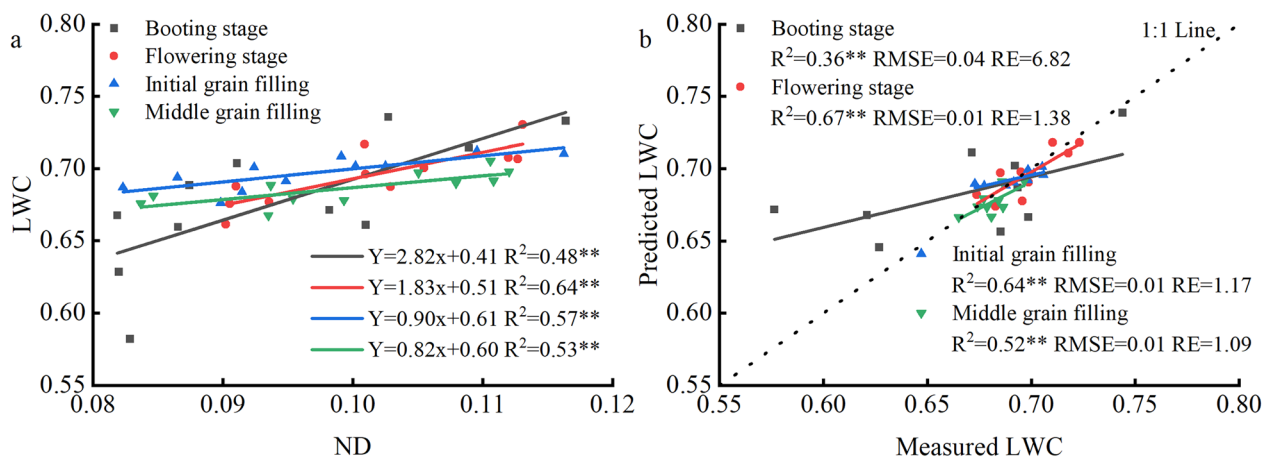


Fig. 5 Model construction (**a**) and validation (**b**) of normalized difference index (ND) for monitoring LWC at different growth stages of rice plants. Each symbol is the average value from 5 (**a**) and 4 (**b**) repeated measurements. Dashed lines represent 1:1 lines; * and ** indicate significant correlation at 5% and 1% probability level, respectively

plants. The R^2 of the linear equation between LWC (y) and $ND_{(1287,1673)}$ (x) was 0.48, 0.64, 0.57, and 0.53, and the corresponding prediction R^2 was 0.36, 0.67, 0.64, and 0.52 at the booting, flowering, initial grain filling, and middle grain filling stages, respectively (Fig. 5).

The relationship between physiological-ecological indicators and LWC

The physiological-ecological indicators had obvious effects on LWC at each growth stage and yield (Fig. 6). The LWC had the highest correlation with grain yield at the booting, flowering, and initial grain filling stages. The LWC is the most important parameter that regulates grain yield, especially for different irrigation treatments. CWSI had the highest correlation coefficient with LWC (-0.60 to -0.85) at each observed stage, followed by the Fv/Fm (0.56–0.71) and SPAD (0.53–0.67). Therefore, CWSI, Fv/Fm , and SPAD can be considered reliable multi-source data to improve the detection accuracy of LWC using the vegetation index model.

In this study, the slope (a) and intercept (b) of linear functions between LWC and $ND_{(1287,1673)}$ at different growth stages (Fig. 5) were systematically analyzed using multi-source data, such as SPAD, Fv/Fm , and CWSI (Fig. 7). Both a and b have significantly quadratic curve relations with CWSI, SPAD, and Fv/Fm . For a parameter, the determination coefficient (R^2) of the curves reached the highest for the CWSI ($R^2=0.92$), followed by the SPAD and Fv/Fm . In contrast, Fv/Fm had the greatest R^2 with parameter b ($R^2=0.99$), while CWSI had the lowest R^2 with parameter b ($R^2=0.37$). The results illustrate that SPAD, Fv/Fm , and CWSI can be regarded as useful multi-source data to improve the prediction capacities of LWC.

Improving the prediction performance of LWC by coupling multi-source data

There were no multicollinearity between ND and SPAD, between ND and Fv/Fm , between ND and CWSI based on the tolerance and variance inflation factor values (Additional file 1: Table S2). Also, Durbin Watson

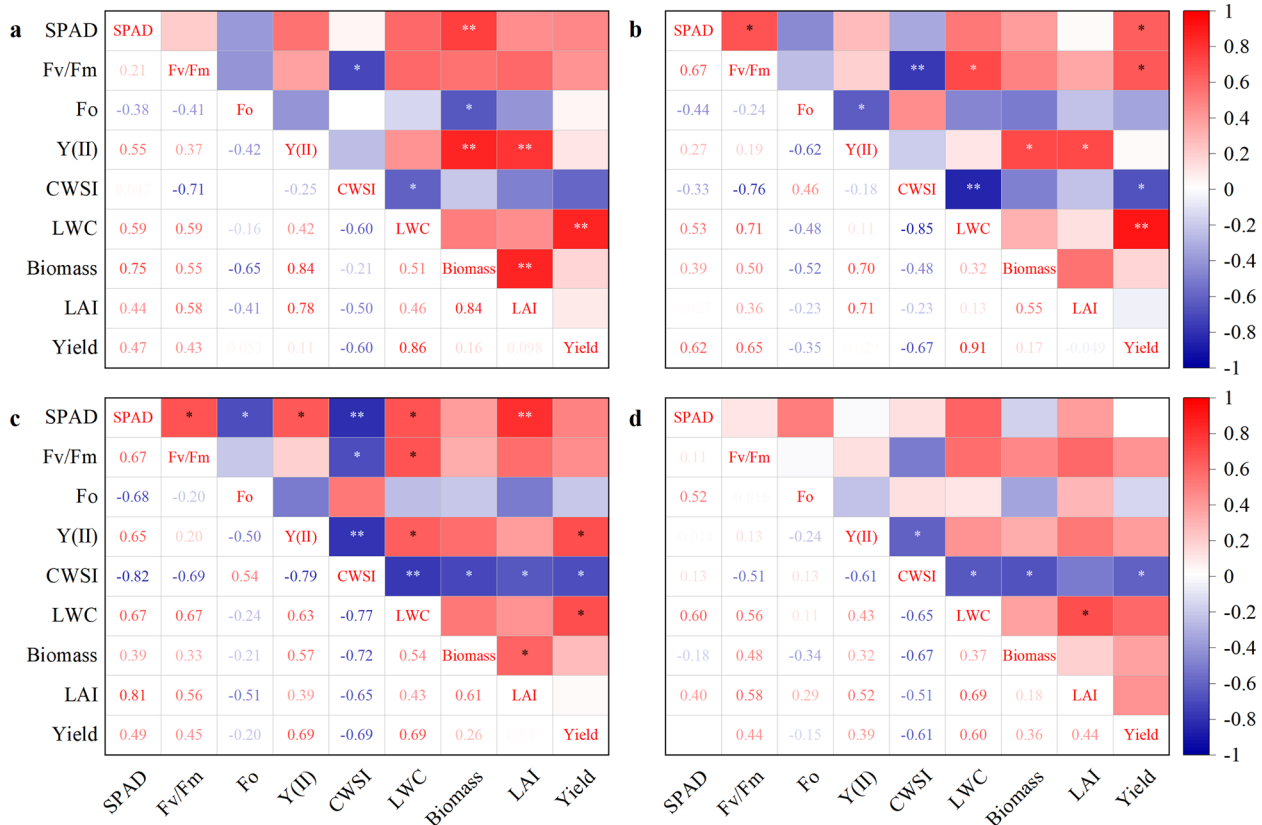


Fig. 6 Pearson correlation coefficient (PCC) among LWC, biomass, LAI, CWSI, SPAD, Fo, Fv/Fm, Y(II), and yield at different growth stages. **a–d** Represents the PCC of the booting, flowering, initial grain filling, and middle grain filling stages, respectively. * and ** indicate significant correlation at 5% and 1% probability level, respectively. SPAD: chlorophyll content; Fv/Fm: maximum photochemical efficiency; Fo: minimal fluorescence; Y(II): actual photochemical efficiency; CWSI: crop water stress index; LWC: leaf water content; Biomass: above-ground biomass; LAI: leaf area index

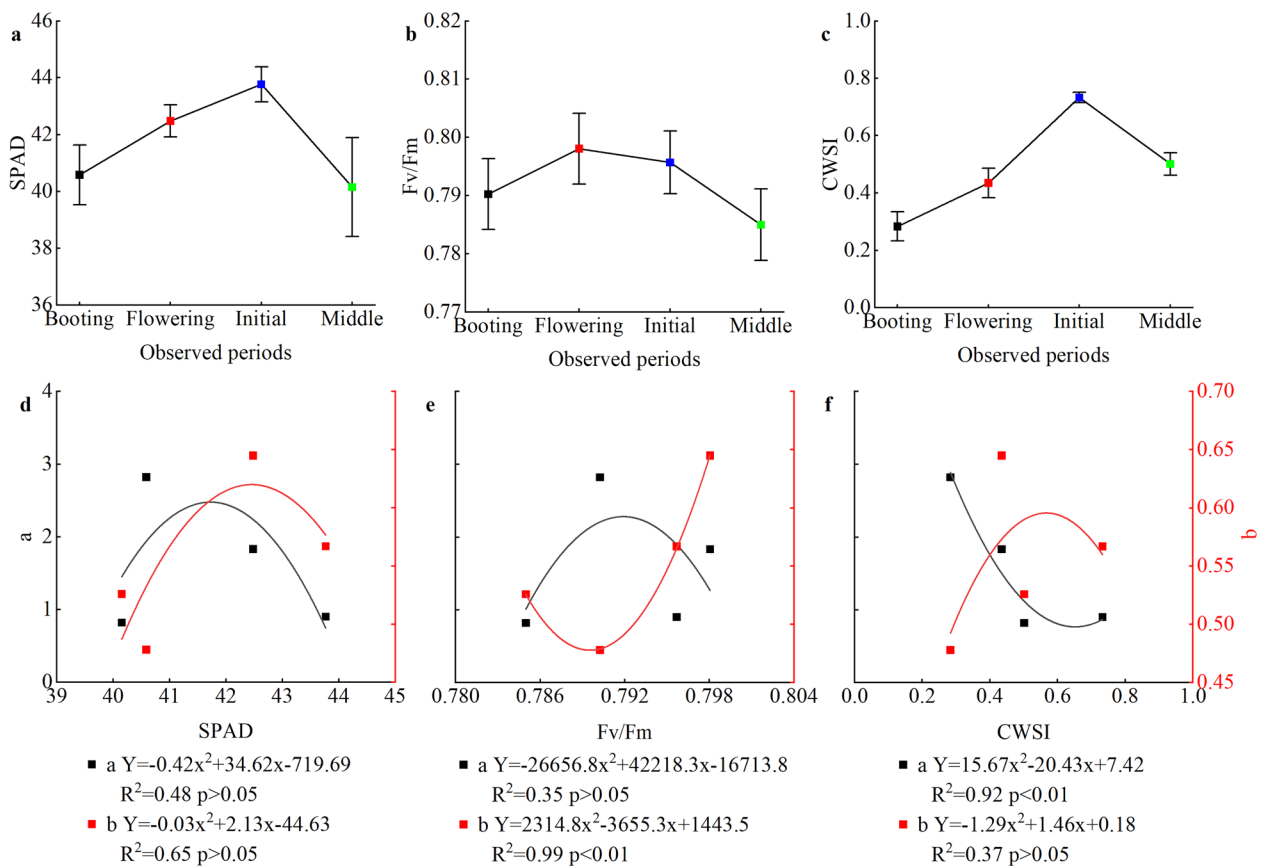


Fig. 7 The SPAD value (a), Fv/Fm (b), and CWSI (c) at different growth stages merged cultivars, water treatments, and years. The quadratic function relations of both a (black line) and b (red line) parameters derived from the ND_(1287,1673) model in Fig. 4a with SPAD d, Fv/Fm (e), and CWSI (f). Booting: booting stage; Flowering: flowering stage; Initial: initial grain filling; Middle: middle grain filling; $p < 0.05$ and $p < 0.01$ indicate significant correlation at 5% and 1% probability level, respectively

Table 4 The ND_(1287,1673) models predicting LWC after integrating with multi-source data

| Growth stage | Model | Multiple linear regression equation | Model Precision | Prediction Precision | RMSE |
|-----------------------|----------|-------------------------------------|-----------------|----------------------|------|
| Booting | ND+SPAD | LWC = 1.91ND + 0.02SPAD - 0.19 | 0.60* | 0.53* | 0.04 |
| Booting | ND+Fv/Fm | LWC = 2.75ND + 1.99Fv/Fm - 1.16 | 0.61* | 0.62** | 0.03 |
| Booting | ND+CWSI | LWC = 1.22ND - 0.34CWSI + 0.66 | 0.69** | 0.65** | 0.03 |
| Flowering | ND+SPAD | LWC = 1.48ND + 0.01SPAD + 0.19 | 0.71** | 0.60* | 0.02 |
| Flowering | ND+Fv/Fm | LWC = 2.10ND + 0.37Fv/Fm + 0.18 | 0.74** | 0.64* | 0.02 |
| Flowering | ND+CWSI | LWC = 1.49ND - 0.08CWSI + 0.57 | 0.75** | 0.69** | 0.01 |
| Initial grain filling | ND+SPAD | LWC = 1.15ND + 0.01SPAD + 0.16 | 0.67** | 0.64** | 0.01 |
| Initial grain filling | ND+Fv/Fm | LWC = 1.58ND + 0.88Fv/Fm - 0.16 | 0.68** | 0.66** | 0.01 |
| Initial grain filling | ND+CWSI | LWC = 1.11ND - 0.20CWSI + 0.73 | 0.61* | 0.55* | 0.01 |
| Middle grain filling | ND+SPAD | LWC = 0.60ND + 0.002SPAD + 0.56 | 0.56* | 0.49* | 0.01 |
| Middle grain filling | ND+Fv/Fm | LWC = 0.70ND + 1.19Fv/Fm - 0.31 | 0.59* | 0.56* | 0.01 |
| Middle grain filling | ND+CWSI | LWC = 0.59ND - 0.10CWSI + 0.67 | 0.57* | 0.44 | 0.01 |

LWC: leaf water content; ND_(1287,1673): normalized difference index; SPAD: chlorophyll content; Fv/Fm: maximum photochemical efficiency; CWSI: crop water stress index; * and ** indicate significant correlation at 5% and 1% probability level, respectively

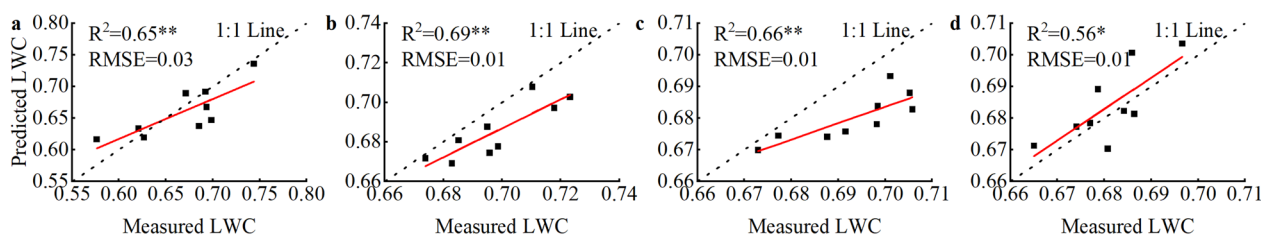


Fig. 8 The 1:1 validation of the ND_(1287,1673) and multi-source data coupling model for monitoring LWC at different growth stages. **a, b** represents ND + CWSI model validation of the booting and flowering stages, and **c, d** represents ND + Fv/Fm model validation of the initial and middle grain filling stages. Each symbol is the average value of 4 measurements. * and ** indicate significant correlation at 5% and 1% probability level, respectively

Table 5 The R² and RMSE of simulating and verifying data sets for the four machine learning algorithms

| Algorithm | Simulating data set | | Verification data set | |
|-----------|-----------------------------|-------------------|-----------------------------|-------------------|
| | R _s ² | RMSE _s | R _v ² | RMSE _v |
| DT | 0.80 | 0.01 | 0.77 | 0.02 |
| RF | 0.85 | 0.01 | 0.78 | 0.02 |
| KNN | 0.80 | 0.01 | 0.75 | 0.02 |
| GBDT | 0.93 | 0.01 | 0.86 | 0.01 |

R_s²: determination coefficient of simulating data set; RMSE_s: root mean square error of simulating data set; R_v²: determination coefficient of verification data set; RMSE_v: root mean square error of verification data set

test indicated that the established multivariate linear regressed models in this study were basically meet the requirements of linear model (Additional file 1: Table S2). The indicators CWSI, Fv/Fm, and SPAD were introduced into the ND_(1287,1673) model to improve the prediction model accuracy of LWC further (Table 4). The results demonstrated that the ND_(1287,1673) model integrating with multi-source data enhanced predictive power than the conventional ND_(1287,1673) model (Fig. 5 and Table 4). The coupled models of ND_(1287,1673) and CWSI had the best monitoring capacity to LWC at the booting and flowering stages (Fig. 8). The R² improved from 0.48–0.64 in conventional ND_(1287,1673) models to 0.69–0.75 in the

coupled models. Prediction R² also improved from 0.36–0.67 in conventional ND_(1287,1673) models to 0.65–0.69 in the coupled models at the booting and flowering stages. Moreover, the coupling models of ND_(1287,1673) and Fv/Fm were optimal at the initial and middle grain filling. The R² and the corresponding prediction R² improved in the coupling models than in conventional ND_(1287,1673) models at the initial and middle grain filling.

Machine learning algorithm optimization model

The data from the flowering stage, most sensitive to water treatments, were used as an example to assess the effects of the machine learning algorithm on ND_(1287,1673) combined with CWSI model precision (Table 5; Fig. 9). The R² were 0.77, 0.78, 0.75, and 0.86 based on DT, RF, KNN, and GBDT models, respectively. Additionally, the RMSE was 0.01 and 0.02 for the simulating and verifying data sets in the machine learning algorithm, respectively.

The R² and RMSE of simulation and verification sets of machine learning algorithm were compared with multiple linear regression (MLR) model (Fig. 10). R² of DT, RF, KNN and GBDT simulation sets were 1.16, 1.23, 1.16 and 1.35 times higher than that of MLR, respectively (Fig. 10). However, the RMSE of all simulated models were almost the same, which the RMSE in simulation sets were no advantage in machine learning algorithms compared with MLR. In addition, R² of DT, RF, KNN

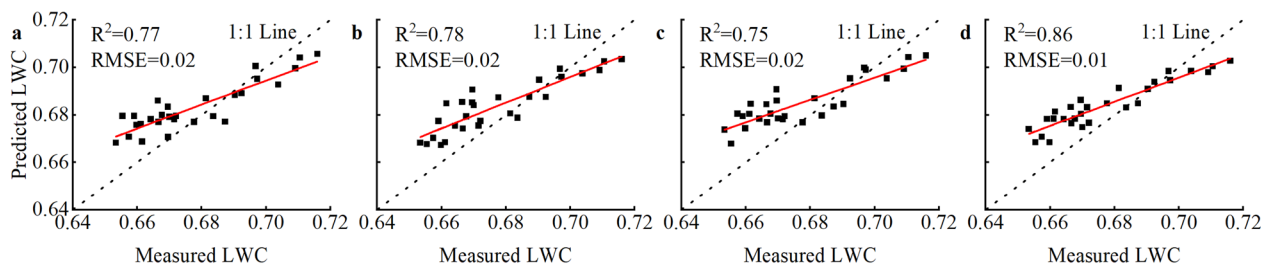


Fig. 9 The 1:1 validation of the machine learning algorithms with ND + CWSI model for monitoring LWC at the flowering stage. **a** Decision tree regression (DT); **b** Random forest regression (RF); **c** K-nearest neighbor regression (KNN); **d** Gradient boosting decision tree regression (GBDT)

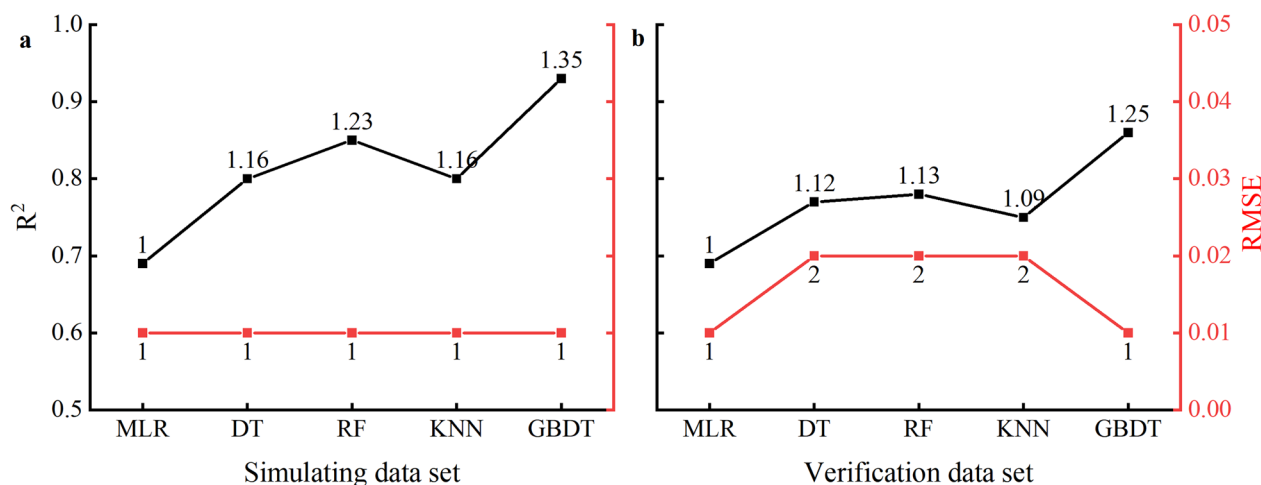


Fig. 10 The R^2 (black line) and RMSE (red line) of simulating (a) and verifying (b) data sets for the four machine learning algorithms and multiple linear regression. The numbers closed to the dots represent the changed folds in machine learning algorithms when compared with the normalized multiple linear regression, respectively

and GBDT verification sets were 1.12, 1.13, 1.09 and 1.25 times higher than that of MLR, respectively. The RMSE in verification sets of DT, RF and KNN was about 2 times higher than that of MLR, but the RMSE of GBDT was at the same level when compared with MLR. Particularly, GBDT synergistically improved the R^2 of simulating and verifying models with low RMSR level compared with MLR.

Discussion

Relationship between multi-source data and LWC

A suitable vegetation index is important to construct an LWC spectral monitoring model in smart agriculture. The first important step is to screen the sensitive band. However, there are some differences on the selected water-sensitive bands for different crops. Thomas et al. [40] discovered a significant correlation between the relative water content of cotton leaves and the reflectance values in the near-infrared (NIR) bands (1450 and 1930 nm). Yang et al. [41] suggest that the sensitive band range of LWC in wheat plants is located in the visible (400–780 nm) and NIR bands (1400–2500 nm). This study selected different spectral indices from previous studies to establish relationships with rice LWC. However, the correlation was not optimal. The ND constructed by the NIR band at 1287 and 1673 nm was instead optimal (Table 3). This study's results were inconsistent with those of previous studies. One of the important reasons could be that vegetation index models for monitoring LWC in our study considered the multiple growth stages of rice plants, rather than focusing on a specific growth stage in previous studies. Additionally,

the spectral data obtained from different ecological sites, climates, and varieties. In this study, the varieties were selected for their commonality bands, considering the universality of the varieties from which the four rice varieties with a large span of validation years and large differences in drought resistance were selected.

Recently, multi-source data related to environmental factors coupled with the spectral monitoring model of LWC has been successfully established in wheat plants to reduce the errors caused by environmental and physiological factors during the spectral monitoring process [16]. This study used the Pearson correlation coefficient method to analyze the physiological and ecological indicators related to LWC. Our results displayed that CWSI, Fv/Fm and SPAD would be considered as optimized multi-source data because the parameters had a higher correlation with LWC compared to other physiological and agronomic indices during whole observed periods (Fig. 6). These findings were also consistent with the basic physiological laws of plants that the three parameters would be changed with the change of LWC. First, the transpiration rate quickly reduces with declining LWC in rice plants [42, 43]. A low transpiration rate increases canopy temperature for a short time and then fleetly regulates CWSI in the canopy of plants [44, 45]. Second, H_2O is insufficient to maintain its physiological activities when LWC is low in rice plants; the phenomenon quickly downregulates photosynthetic performance, such as declining Fv/Fm [46–48]. Third, SPAD biosynthesis would be hindered, and decomposition could be accelerated if LWC remains low for a long time, resulting in the yellowing of leaves [49]. Finally, the fluctuant sensitive parameters lead to

significant changes in population characteristics, such as leaf area, biomass, and yield, presenting different degrees of decline [50]. Therefore, we deduce that CWSI, Fv/Fm , and SPAD should be introduced into the spectral monitoring model to improve the R^2 for monitoring LWC in whole growth stages of rice plants.

Coupling multi-source data for monitoring rice LWC

The canopy structure and growth environment of rice have been changing throughout the whole growth period, resulting in different spectral reflectance data, making it difficult to accurately build a water prediction model for the whole growth period with a single vegetation index model [51]. A good relationship between physiological and ecological indicators and spectral information is the key to rapid and non-destructive monitoring of coupled multi-source data. The relationships between CWSI, Fv/Fm , and SPAD and the slope (a) and intercept (b) of the conventional linear $ND_{(1287,1673)}$ models were systematically analyzed at each observed period to achieve accurate prediction of LWC (Fig. 7). The parameter a had the closest relationship with CWSI with R^2 of 0.92. However, the parameter b had the closest relationship with Fv/Fm with R^2 of 0.99. These results indicate that the conventional linear $ND_{(1287,1673)}$ models could be largely influenced by physiological and ecological factors when LWC was monitored using spectroscopic equipment. Therefore, slope and intercept could be defined as ecological and physiological factors in conventional linear $ND_{(1287,1673)}$ models, especially for monitoring LWC in rice plants. This evidence confirms the necessity of incorporating physiological parameters into the conventional $ND_{(1287,1673)}$ model to improve LWC monitoring accuracy in rice plants.

Currently, multi-source data are widely used in hyperspectral monitoring. However, most studies have utilized a single type of multi-source numbers to construct models independently or in combination, and few studies have discussed the effect of coupled multi-source data on the monitoring performance of rice LWC models. This study uses a commonly used multiple linear regression algorithm to estimate rice LWC based on coupled multi-source data and compare its monitoring performance. The results revealed that the monitoring capacities in $ND_{(1287,1673)}$ models coupled with physiological and ecological factors were significantly increased than conventional $ND_{(1287,1673)}$ models across all observed stages (Tables 3 and 4). The coupled model of $ND + CWSI$ and $ND + Fv/Fm$ had the best monitoring capacities at the booting to flowering and initial to middle grain filling stages, respectively. The different coupled parameters at different growth stages may

related to the differences in rice growth characteristics. Previous studies have demonstrated that the canopy structure is unstable and that leaf area and biomass are in a state of constant growth during the booting and flowering stages in rice plants [52]. Irrigation regimes directly regulate canopy growth and development [53]. Therefore, the canopy micro-ecological factors, such as photosynthetically active radiation, temperature, and humidity, constantly fluctuate [54], especially for different water treatments. Finally, CWSI is a sensitive index at this stage because CWSI is mainly regulated by the micro-ecological factor [31, 32]. However, rice plants are in the senescence stage after flowering. Leaf photosynthetic capacity significantly declines during the grain-filling stage compared to pre-flowering in rice plants [43, 55]. Additionally, photosynthetic performances are susceptible to different irrigation regimes during senescence processes [56]. Fv/Fm is an important index evaluating photosynthetic performance. Therefore, photosynthetic performance could be a better multi-source parameter to predict LWC during the rice-filling stage.

Comparison of machine learning algorithms

Machine learning algorithms have recently been widely used in model monitoring research and have become popular tools in precision agricultural production research. In this study, four machine learning algorithms, DT, RF, KNN, and GBDT were used to perform operations based on coupled models using rice flowering data as an example (Table 5). The results indicated that the R^2 for the simulation and validation sets of DT, RF, KNN and GBDT were 0.80–0.93 and 0.75–0.86, respectively, which the R^2 of machine learning algorithms were 1.16–1.35 times higher for the simulation sets and 1.09–1.25 times higher for the validation sets compared with MLR (Fig. 10). Among them, GBDT has the highest R^2 with low and stable RMSE. Therefore, the GBDT model would be considered as the best machine learning algorithm for monitoring LWC in this study's simulating and verification data sets. Previous research supports our results that the GBDT algorithm has higher prediction accuracy for leaf area in maize plants among the aforementioned algorithms [57]. GBDT can improve prediction accuracy by constructing a weak learner to correct the original model error via residuals and resultant iterations when the sample size is small [58]. The same weights are assigned for each input factor for the other algorithms; the algorithm would fail to determine the primary contributing factor in the input factor if the training data are insufficient [57]. This may be the primary reason the GBDT algorithm has significant advantages to improve LWC monitoring capabilities in

this study because our sample size could not be abundant enough to support the data requirements of other algorithms. In summary, machine learning positively improves the monitoring abilities of LWC in rice plants.

Future perspectives

Factors such as fertility period and rice LWC changes lead to parameter differences between relevant indicators extracted from various data sources. Current crop monitoring models are mostly single-factor statistical models that are difficult to consider crop growth and development, yield formation and its interactions with climate and soil environment, and lack universality and dynamics. Therefore, coupling spectral remote sensing information with multivariate data can construct a spectral monitoring model with high accuracy and stable reliability of crop growth, moisture content, and other indicators, providing an effective solution to the spectral monitoring problem. In this study, only four rice cultivars were considered after the booting stage, and the early growth of rice has yet to be monitored. Establishing a standardized water stress detection and diagnosis system based on water critical thresholds at various growth stages of rice will require the accumulation of data from multi-year, multi-point, continuous trials based on different rice variety types. This can accurately monitor the LWC to ensure high rice yield, improving water resource utilization efficiency and providing a basis for implementing precision agriculture.

Conclusion

In this study, the LWC of rice with different cultivars, years, and water treatments was monitored based on multi-source data (physiological and ecological indicators and spectral information) with machine learning, explored the performance of single and combined multi-source data in LWC monitoring, and utilized different physiological indicators to establish two monitoring models for the growth differences of rice in the initial and middle stages. The results displayed that ND + CWSI had better monitoring performance in the early stage (booting to flowering stage), while ND + Fv/Fm was better in the late stage (initial to middle grain filling stage). Additionally, this study's newly constructed vegetation index ND_(1287,1673) also has good monitoring capacities for LWC in rice. Meanwhile, the machine learning algorithm (GBDT) further improves the monitoring performance of the model. In summary, this study confirms that using multi-source data and machine learning can improve the performance of hyperspectral prediction of rice LWC.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-024-01168-5>.

Additional file 1: Fig. S1. The normal distribution characteristics of all parameters, linear distribution scatter plots and the corresponding correlation coefficient between LWC and physiological and ecological parameters including SPAD, Fv/Fm, Fo, Y(II), CWSI, LWC, Biomass, and LAI at booting stage and Yield at maturity. **Fig. S2.** The normal distribution characteristics of all parameters, linear distribution scatter plots and the corresponding correlation coefficient between LWC and physiological and ecological parameters including SPAD, Fv/Fm, Fo, Y(II), CWSI, LWC, Biomass, and LAI at flowering stage and Yield at maturity. **Fig. S3.** The normal distribution characteristics of all parameters, linear distribution scatter plots and the corresponding correlation coefficient between LWC and physiological and ecological parameters including SPAD, Fv/Fm, Fo, Y(II), CWSI, LWC, Biomass, and LAI at initial grain filling stage and Yield at maturity. **Fig. S4.** The normal distribution characteristics of all parameters, linear distribution scatter plots and the corresponding correlation coefficient between LWC and physiological and ecological parameters including SPAD, Fv/Fm, Fo, Y(II), CWSI, LWC, Biomass, and LAI at middle grain filling stage and Yield at maturity. **Table S1.** The maximum, minimum, and mean values of the main measured parameters and standard deviation and coefficient of variation of each parameter in this study. **Table S2.** The multicollinearity test between ND and SPAD, between ND and Fv/Fm, between ND and CWSI at different observed periods based on the tolerance and variance inflation factor values and Durbin Watson test of multivariate linear regressed models presented in Table 4 in text.

Acknowledgements

We gratefully acknowledge Pro. Daniel Dias and four anonymous reviewers for their critical comments and suggestions on the manuscript. We also acknowledge Dr. Danyan Chen and Dr. Juan Liao of Anhui agricultural university for their guidances on machine learning algorithms in this study.

Author contributions

Conceived and designed the experiments: HH. Performed the experiments: XZ, HX and YS. Analyzed the data: XZ, TZ, CH, and LWang. Contributed to reagents/materials/analysis tools: CY, JK, and QZ. Wrote the manuscript: XZ and HX. Corrected the manuscript: HH and LWu.

Funding

This work was supported by the National Key Research and Development Program (2022YFD2301404) and the Open Project of National Modern Agricultural Industrial Park in Yingshang County (hx23298).

Availability of data and materials

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

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Agricultural College, Anhui Agricultural University, Hefei 230036, Anhui, People's Republic of China. ²Collaborative Innovation Center for Modern Crop Production Co-Sponsored by Province and Ministry (CIC-MCP), Nanjing 210095, Jiangsu, People's Republic of China. ³Yingshang Agricultural Green Development Promotion Center, Fuyang 236200, Anhui, People's Republic of China. ⁴Germplasm Creation and Application Laboratory of Grain

and Oil Crops in Wanjiang Plain, Enterprise Key Laboratory of Ministry of Agriculture and Rural Affairs, Tongling 244002, China.

Received: 8 September 2023 Accepted: 28 February 2024

Published online: 23 March 2024

References

- Thakur AK, Kassam A, Stoop WA, Uphoff N. Modifying rice crop management to ease water constraints with increased productivity, environmental benefits, and climate-resilience. *Agr Ecosyst Environ*. 2016;235:101–4.
- Belder P, Bouman BAM, Cabangon R, Lu GA, Quilang EJP, Li YH, Spiertz JHJ, Tuong TP. Effect of water-saving irrigation on rice yield and water use in typical lowland conditions in Asia. *Agr Water Manage*. 2004;65(3):193–210.
- Ding YM, Wang WG, Song RM, Shao QX, Jiao XY, Xing WQ. Modeling spatial and temporal variability of the impact of climate change on rice irrigation water requirements in the middle and lower reaches of the Yangtze River, China. *Agr Water Manage*. 2017;193:89–101.
- Stimson HC, Breshears DD, Ustin SL, Kefauver SC. Spectral sensing of foliar water conditions in two co-occurring conifer species: *Pinus edulis* and *Juniperus monosperma*. *Remote Sens Environ*. 2005;96(1):108–18.
- Chakrabarti S, Bongiovanni T, Judge J, Zotarelli L, Bayer C. Assimilation of SMOSS soil moisture for quantifying drought impacts on crop yield in agricultural regions. *IEEE J-STARS*. 2014;7(9):3867–79.
- Chen JG, Chen J, Wang QJ, Zhang Y, Ding HF, Huang Z. Retrieval of soil dispersion using hyperspectral remote sensing. *J Indian Soc Remote*. 2016;44(4):563–72.
- Wang JJ, Cui LJ, Gao WX, Shi TZ, Chen YY, Gao Y. Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. *Geoderma*. 2014;216:1–9.
- Mack AR, Ferguson WS. A moisture stress index for wheat by means of a modulated soil moisture budget. *Can J Plant Sci*. 1968;48:535–44.
- Peñuelas J, Pinol J, Ogaya R, Filella I. Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). *Int J Remote Sens*. 1997;18(13):2869–75.
- Danson FM, Steven MD, Malthus TJ, Clark JA. High-spectral resolution data for determining leaf water concentration. *Int J Remote Sens*. 1992;13(3):461–70.
- Liang L, Yang MH, Zang Z. Determination of wheat canopy nitrogen content ratio by hyperspectral technology based on wavelet denoising and support vector regression. *Trans Chin Soc Agric Eng*. 2010;26(12):248–53.
- Shibayama M, Takahashi W, Morinaga S, Akiyama T. Canopy water deficit detection in paddy rice using a high resolution field spectroradiometer. *Remote Sens Environ*. 1993;45(2):117–26.
- Liu XJ, Tian YC, Yao X, Cao WX, Zhu Y. Monitoring leaf water content based on hyperspectra in rice. *Scientia Agricultura Sinica*. 2012;45(03):435–42.
- Kunz K, Hu YC, Schmidhalter U. Carbon isotope discrimination as a key physiological trait to phenotype drought/heat resistance of future climate-resilient German winter wheat compared with relative leaf water content and canopy temperature. *Front Plant Sci*. 2022;13:1043458.
- García-Haro FJ, Campos-Taberner M, Moreno A, Tagesson HT, Camacho F, Martínez B, Sánchez S, Piles M, Camps-Valls G, Yebra M, Gilabert MA. A global canopy water content product from AVHRR/Metop. *ISPRS J Photogramm*. 2020;162:77–93.
- Shi B, Yuan YF, Zhuang TX, Xu X, Schmidhalter U, Ata-UI-Karim ST, Zhao B, Liu XJ, Tian YC, Zhu Y, Cao WX, Cao Q. Improving water status prediction of winter wheat using multi-source data with machine learning. *Eur J Agron*. 2022;139: 126548.
- Zhang XN, Wang LL, Niu MX, Zhan N, Ren HJ, Xu HC, Yang K, Wu LQ, Ke J, You CC, He HB. Estimation of rice leaf water content based on leaf reflectance spectrum and chlorophyll fluorescence. *Acta Agriculturae Zhejiangensis*. 2023;35(6):1265–77.
- Qin SZ, Ding YR, Zhou ZX, Zhou M, Wang HY, Xu F, Yao QS, Lv X, Zhang Z, Zhang LF. Study on the nitrogen content estimation model of cotton leaves based on “image-spectrum-fluorescence” data fusion. *Front Plant Sci*. 2023;14:1117277.
- Rodríguez-Pérez JR, Ordóñez C, González-Fernández AB, Sanz-Ablanedo E, Valenciano JB, Marcelo V. Leaf water content estimation by functional linear regression of field spectroscopy data. *Biosyst Eng*. 2018;165:36–46.
- Wu YP, He L, Wang YY, Liu BC, Wang YH, Guo TC, Feng W. Dynamic model of vegetation indices for biomass and nitrogen accumulation in winter wheat. *Acta Agron Sin*. 2019;45(08):1238–49.
- Cao J, Zhang Z, Wang C, Liu J, Zhang L. Susceptibility assessment of landslides triggered by earthquakes in the Western Sichuan Plateau. *CATENA*. 2019;175:63–76.
- Kamir E, Waldner F, Hochman Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS J Photogramm*. 2020;160:124–35.
- Zhang J, Xu B, Feng HK, Jing X, Wang JJ, Ming SK, Fu YQ, Song XY. Monitoring nitrogen nutrition and grain protein content of rice based on ensemble learning. *Spectrosc Spect Anal*. 2022;42(14):1956–64.
- An GQ, Xing MF, He BB, Liao CH, Huang XD, Shang JL, Kang HQ. Using machine learning for estimating rice chlorophyll content from in situ hyperspectral data. *Remote Sens*. 2020;12(18):3104.
- Li JM, Chen XQ, Yang Q, Shi LS. Deep learning models for estimation of paddy rice leaf nitrogen concentration based on canopy hyperspectral data. *Acta Agron Sin*. 2021;47(07):1342–50.
- Das B, Sahoo RN, Pargal S, Krishna G, Verma R, Viswanathan C, Sehgal VK, Gupta VK. Evaluation of different water absorption bands, indices and multivariate models for water-deficit stress monitoring in rice using visible-near infrared spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc*. 2021;247: 119104.
- Chu G, Chen TT, Wang ZQ, Yang JC, Zhang JH. Morphological and physiological traits of roots and their relationships with water productivity in water-saving and drought-resistant rice. *Field Crop Res*. 2014;162:108–19.
- Serrano L, Ustin SL, Roberts DA, Gamon JA, Peñuelas J. Deriving water content of chaparral vegetation from AVIRIS data. *Remote Sens Environ*. 2000;74(3):570–81.
- Hardisky MS, Klemas V, Smart RM. The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of *Spartina alterniflora* canopies. *Photogramm Eng Remote Sensing*. 1983;48(1):77–84.
- Gao BC. NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens Environ*. 1995;58(3):257–66.
- Idso SB, Jackson RD, Pinter PJ, Reginato RJ, Hatfield JL. Normalizing the stress-degree-day parameter for environmental variability. *Agric Meteorol*. 1981;24:45–55.
- Jackson RD, Idso SB, Reginato RJ, Pinter PJ. Canopy temperature as a crop water stress indicator. *Water Resour Res*. 1981;17(4):1133–8.
- Chen L, Xing MF, He BB, Wang JF, Shang JL, Huang XD, Xu M. Estimating soil moisture over winter wheat fields during growing season using machine-learning methods. *IEEE J-STARS*. 2021;14:3706–18.
- Zhuang TX, Zhang Y, Li D, Schmidhalter U, Ata-UI-Karim ST, Cheng T, Liu XJ, Tian YC, Zhu Y, Cao WX, Cao Q. Coupling continuous wavelet transform with machine learning to improve water status prediction in winter wheat. *Precis Agric*. 2023;24:2171–99.
- Bressan TS, Marcelo KDS, Girelli TJ, Junior FC. Evaluation of machine learning methods for lithology classification using geophysical data. *Comput Geosci*. 2020;139: 104475.
- Zhao QX, Yu SC, Zhao F, Tian LH, Zhao Z. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. *Forest Ecol Manag*. 2019;434:224–34.
- Souza DV, Nievola JC, Santos JX, Wojciechowski J, Gonçalves AL, Corte APD, Sanquetta CR. K-nearest neighbor regression in the estimation of tectona grandis trunk volume in the state of Pará, Brazil. *J Sustain Forest*. 2019;38(8):755–68.
- Friedman JH. Stochastic gradient boosting. *Comput Stat Data An*. 2002;38(4):367–78.
- Taylor JA, Bates TR. A discussion on the significance associated with Pearson's correlation in precision agriculture studies. *Precis Agric*. 2013;14:558–64.
- Thomas JR, Namken LN, Oerther GF, Brown RG. Estimating leaf water content by reflectance measurements. *Agron J*. 1971;63(6):845–7.
- Yang FF, Liu T, Wang QY, Du MZ, Yang TL, Liu DZ, Li SJ, Liu SP. Rapid determination of leaf water content for monitoring waterlogging in winter wheat based on hyperspectral parameters. *J Integr Agr*. 2021;20(10):2613–26.
- He HB, Wang Q, Wang LL, Yang K, Yang R, You CC, Ke J, Wu LQ. Photosynthetic physiological response of water-saving and drought-resistant rice

- to severe drought under wetting-drying alternation irrigation. *Physiol Plantarum*. 2021;173(4):2191–206.
43. Wang LL, Zhang XN, She YH, Hu C, Wang Q, Wu LQ, You CC, Ke J, He HB. Physiological adaptation mechanisms to drought and rewatering in water-saving and drought-resistant rice. *Int J Mol Sci*. 2022;23(22):14043.
 44. García-Tejero IF, Hernandez A, Padilla-Díaz CM, Diaz-Espejo A, Fernandez J. Assessing plant water status in a hedgerow olive orchard from thermography at plant level. *Agr Water Manage*. 2017;188:50–60.
 45. Ekinzog EK, Schlerf M, Kraft M, Werner F, Riedel A, Rock G, Mallick K. Revisiting crop water stress index based on potato field experiments in northern Germany. *Agr Water Manage*. 2022;269: 107664.
 46. Murchie EH, Lawson T. Chlorophyll fluorescence analysis: A guide to good practice and understanding some new applications. *J Exp Bot*. 2013;64(13):3983–98.
 47. Clauw P, Coppens F, Beuf KD, Dhondt S, Daele TV, Maleux K, Storme V, Clement L, Gonzalez N, Inzé D. Leaf responses to mild drought stress in natural variants of *Arabidopsis*. *Plant Physiol*. 2015;167(3):800–16.
 48. Hazrati S, Tahmasebi-Sarvestani Z, Modarres-Sanavy SAM, Mokhtassi-Bidgoli A, Nicola S. Effects of water stress and light intensity on chlorophyll fluorescence parameters and pigments of *Aloe vera* L. *Plant Physiol Bioch*. 2016;106:141–8.
 49. Kumar A, Sengar RS, Pathak RK, Singh AK. Integrated approaches to develop drought-tolerant Rice: demand of era for global food security. *J Plant Growth Regul*. 2023;42:96–120.
 50. Nouna BB, Katerji N, Mastrorilli M. Using the CERES-Maize model in a semi-arid mediterranean environment. Evaluation of model performance. *Eur J Agron*. 2000;13(4):309–22.
 51. Tilling AK, O'Leary GJ, Ferwerda JG, Jones SD, Fitzgerald GJ, Rodriguez D, Belford R. Remote sensing of nitrogen and water stress in wheat. *Field Crop Res*. 2007;104(1–3):77–85.
 52. He HB, Yang K, Xu HC, Yao B, Li GH, Zhang XN, Yang R, You CC, Ke J, Wu LQ. Precision nitrogen management regimes to obtain high yield and good eating quality of medium indica hybrid rice in machine transplanting with bowl-type nursery tray (MTB) based on the critical nitrogen concentration. *Eur J Agron*. 2023;143: 126711.
 53. Raj R, Walker JP, Vinod V, Pingale R, Naik B, Jagarlapudi A. Leaf water content estimation using top-of-canopy airborne hyperspectral data. *Int J Appl Earth Obs*. 2021;102: 102393.
 54. Selvaraj MG, Ishizaki T, Valencia M, Ogawa S, Dedicova B, Ogata T, Yoshiwara K, Maruyama K, Kusano M, Saito K, Takahashi F, Shinozaki K, Nakashima K, Ishitani M. Overexpression of an *Arabidopsis thaliana* galactinol synthase gene improves drought tolerance in transgenic rice and increased grain yield in the field. *Plant Biotechnol J*. 2017;15(11):1465–77.
 55. Turner NC, O'Toole JC, Cruz RT, Namuco OS, Ahmad S. Response of seven diverse rice cultivars to water deficits I. Stress development, canopy temperature, leaf rolling and growth. *Field Crop Res*. 1986;13:257–71.
 56. Wang YW, Hua L, Xu C, Chen GX. Long-term drought resistance in rice (*Oryza sativa* L.) during leaf senescence: a photosynthetic view. *Plant Growth Regul*. 2019;88:253–66.
 57. Zhang HM, Liu W, Han WT, Liu QZ, Song RJ, Hou GH. Inversion of summer maize leaf area index based on gradient boosting decision tree algorithm. *Trans Chin Soc Agric Mach*. 2019;50(05):251–9.
 58. Fridsman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189–232.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.