



Article

A Novel Hybrid GOA-XGB Model for Estimating Wheat Aboveground Biomass Using UAV-Based Multispectral Vegetation Indices

Yixiu Han ^{1,2}, Rui Tang ¹, Zhenqi Liao ³, Bingnian Zhai ^{1,4,*} and Junliang Fan ³

¹ College of Resources and Environment, North West Agriculture and Forestry University, Yangling 712100, China; hanyixiu@nit.edu.cn (Y.H.); tangui0168@nwsuaf.edu.cn (R.T.)

² School of Hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China

³ Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas of Ministry of Education, North West Agriculture and Forestry University, Yangling 712100, China; nwafulzq@nwsuaf.edu.cn (Z.L.); nwwfjl@nwsuaf.edu.cn (J.F.)

⁴ Key Laboratory of Plant Nutrition and the Agri-Environment in Northwest China, Ministry of Agriculture, Yangling 712100, China

* Correspondence: zhaibingnian@nwsuaf.edu.cn

Abstract: The rapid and nondestructive determination of wheat aboveground biomass (AGB) is important for accurate and efficient agricultural management. In this study, we established a novel hybrid model, known as extreme gradient boosting (XGBoost) optimization using the grasshopper optimization algorithm (GOA-XGB), which could accurately determine an ideal combination of vegetation indices (VIs) for simulating wheat AGB. Five multispectral bands of the unmanned aerial vehicle platform and 56 types of VIs obtained based on the five bands were used to drive the new model. The GOA-XGB model was compared with many state-of-the-art models, for example, multiple linear regression (MLR), multilayer perceptron (MLP), gradient boosting decision tree (GBDT), Gaussian process regression (GPR), random forest (RF), support vector machine (SVM), XGBoost, SVM optimization by particle swarm optimization (PSO), SVM optimization by the whale optimization algorithm (WOA), SVM optimization by the GOA (GOA-SVM), XGBoost optimization by PSO, XGBoost optimization by the WOA. The results demonstrated that MLR and GOA-MLR models had poor prediction accuracy for AGB, and the accuracy did not significantly improve when input factors were more than three. Among single-factor-driven machine learning (ML) models, the GPR model had the highest accuracy, followed by the XGBoost model. When the input combinations of multispectral bands and VIs were used, the GOA-XGB model (having 37 input factors) had the highest accuracy, with RMSE = 0.232 kg m⁻², R² = 0.847, MAE = 0.178 kg m⁻², and NRMSE = 0.127. When the XGBoost feature selection was used to reduce the input factors to 16, the model accuracy improved further to RMSE = 0.226 kg m⁻², R² = 0.855, MAE = 0.172 kg m⁻², and NRMSE = 0.123. Based on the developed model, the average AGB of the plot was 1.49 ± 0.34 kg.

Keywords: metaheuristic optimization algorithm; feature selection; XGBoost; Gaussian process regression; aboveground biomass mapping



Citation: Han, Y.; Tang, R.; Liao, Z.; Zhai, B.; Fan, J. A Novel Hybrid GOA-XGB Model for Estimating Wheat Aboveground Biomass Using UAV-Based Multispectral Vegetation Indices. *Remote Sens.* **2022**, *14*, 3506. <https://doi.org/10.3390/rs14143506>

Academic Editors: Thomas Alexandridis, Christos Chalkias, Konstantinos X. Soulis, Dionissios Kalivas and Emmanouil Psomiadis

Received: 21 June 2022

Accepted: 13 July 2022

Published: 21 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wheat is the primary grain crop in China, with a 227,000-km² planting area and yield of 131.7 Mt in 2020; however, extensive wheat management has resulted in chronically low yields. The current wheat production cannot fully meet the demands of 1.4 billion people, and China has to import 8 Mt of grain on an annual basis. Therefore, it is essential to improve the management level of wheat farmland. Aboveground biomass (AGB) is an important indicator for monitoring agricultural ecosystems. It is not only closely related to crop growth monitoring, yield per unit area, and yield formation, but also the primary content of global climate change, carbon cycle, material flow, and energy exchange. For starters,

crop biomass monitoring is an active research topic. In terms of the relationship between dry matter production and wheat yield, a significant positive correlation between biological and economic yields has been observed for a normal growing wheat population [1]. AGB is a critical component of dominant, dynamic terrestrial eco-systems, accounting for ~30% of the total terrestrial area ecosystem carbon pool [2]. Furthermore, AGB estimates are critical for international carbon inventory negotiations and most carbon trading schemes [2].

The accurate estimation and dynamic monitoring of biomass is an essential basis for the efficient usage of farmland resources. However, traditional biomass estimation methods require considerable human, material, and financial resources and often damage crops during the measurement process. Moreover, the operations are limited to a smaller scale; therefore, it is difficult to estimate the AGB of large area crops. With the development of spectral technology and its advantages, such as near-real-time observation ability, crop AGB estimation methods based on remote sensing are increasingly attracting research attention. There are many studies on estimating wheat biomass based on satellite remote sensing, e.g., Zhou et al. [3] fused enhanced vegetation indices (VIs) from MODIS and Landsat-8 to obtain VI products with a high spatial and temporal resolution, to be assimilated in the crop model to estimate wheat AGB. They reported that good accuracy could be obtained using synthetic data, with $R^2 = 0.76$ and $RMSE = 0.176 \text{ kg m}^{-2}$. Jin et al. [4] reported that the three-band water index from Sentinel-2A data correlated highly with wheat biomass, with $R^2 = 0.76$ and $RMSE = 2.84 \text{ t ha}^{-1}$. Furthermore, using 15 multiple hyperspectral VIs as input to train convolutional neural network models can obtain higher accuracy than a single VI model. Dong et al. [5] used Landsat 8 and Sentinel-2 data to estimate the biomass of six crops in Manitoba, Canada. They reported that high accuracy could be obtained with $R^2 = 0.81$ and $RMSE = 0.135 \text{ kg m}^{-2}$.

Unmanned aerial vehicle (UAV) remote sensing technology is currently used extensively in crop AGB monitoring, because of its advantages of flexible application, simple operation, and high spatiotemporal resolution images. The UAV platform can carry various sensors, such as multispectral, hyperspectral, and LiDAR. Zhang et al. [6] developed an AGB estimation model of winter wheat based on UAV digital images using four methods, partial least squares, backpropagation neural network, support vector machine (SVM), and random forest (RF), along with various VIs; the results demonstrated that the model constructed using partial least squares was the best. Yue et al. [7] used UAV texture information, VIs, and their combination to estimate wheat AGB, producing good results with R^2 ranging from 0.59 to 0.78 and $RMSE$ ranging from 1.22 to 1.59 t ha^{-1} . Jia et al. [8] integrated synergy interval partial least squares with a successive projection algorithm (SIPLS-SPA) to identify the optimal spectral characteristics of wheat biomass. They reported that eight wavelengths (706, 724, 734, 806, 808, 810, 812, and 816 nm) were the most sensitive input variables to AGB. The SIPLS-SPA biomass model achieved high performance with $RMSE = 0.059 \text{ kg m}^{-2}$ and relative $RMSE = 38.55\%$ in the validation period.

From the abovementioned studies, machine learning (ML) has become an essential link between spectral information and AGB; however, there are still two major challenges in ML AGB simulation. The first is feature screening. Usually, a single VI will stop increasing after the crop ground coverage reaches a certain value, which differs from the change rule of AGB. Multiple messages alleviate this limitation. However, multiple bands can combine hundreds of VIs; thus, selecting suitable VIs as input factors to drive the ML model is a challenge. The tree-based model can evaluate features by calculating the contribution degrees of different features and delete the relations having redundant features by pruning operation to avoid model overfitting [9]. Liu et al. [10] suggested using a combination of bands and VIs to simulate rice leaf-area indices. They employed two tree-based models to obtain the optimal combination of parameters, and the accuracy of LAI simulation was 20% higher than that of only VI combination. Geng et al. [9] compared four ML models—RF, SVM, artificial neural network (ANN), and extreme gradient boosting (XGBoost)—in corn biomass estimation based on MODIS data. The results indicated that the XGBoost model

was the best ($R^2 = 0.78$, $RMSE = 2.86 \text{ t ha}^{-1}$, and $MAE = 1.86 \text{ t ha}^{-1}$) and could reduce the number of features from 27 to 9.

Moreover, the accuracy and efficiency of the ML models depended largely on the internal model parameters. To improve the accuracy of the simulation, model parameters must be determined in real-time. ML model calibration methods are primarily based on grid-search and gradient descent algorithms. However, these algorithms are complicated and prone to local convergence. Biological heuristic algorithms are extremely accurate and efficient and can provide global optimal solutions. The combination of biological heuristic algorithms and ML models can reduce the optimal solution to optimization problems and improve the calculation speed and performance of the model. Wang et al. [11] designed a hybrid model based on an SVM and binary coded ant colony optimization algorithm to classify remote sensing images. Dong et al. [12] evaluated the performance of four bio-inspired algorithms to optimize the KNEA model for predicting monthly reference evapotranspiration. They reported gray wolf optimizer algorithm outperformed other algorithms in different climate zones of China. As an effective ML model, XGBoost can accurately predict biomass. However, there are still certain limitations on parameter optimization: establishing an accurate relationship between variables and objectives and selecting the most appropriate input parameters. In this study, a state-of-the-art heuristic algorithm, the grasshopper optimization algorithm (GOA), was coupled with the XGBoost model to solve the limitation of parameter optimization and feature selection.

Because of the positive relationship between leaf biomass and VIs, a general model of leaf biomass across all growth stages should be realized. Except for Li [13], there were limitations to the AGB model across all growth stages. Furthermore, more than one hundred different VIs have been used to develop inverse AGB, and scholars are frequently confused about which VIs can best describe the dynamics of biomass during growth stages. It is necessary to devise a method for quickly selecting suitable combinations from a large number of VIs. Therefore, the main objectives of this study are as follows: (1) to develop a new hybrid model by coupling the GOA with XGBoost (GOA-XGB) and evaluate the accuracy of the new model in estimating AGB based on 56 types of multispectral VI data; (2) to compare the performances of six standalone ML models and a hybrid model coupling the GOA with SVM (GOA-SVM); and (3) to recommend the optimal ML model and parameter input combination for estimating wheat AGB.

2. Materials and Methods

2.1. Study Region

Winter wheat field experiments were performed during the growing seasons of 2020–2021 at the Gaoqiao Farm Experiment Station ($34^{\circ}25'39''\text{N}$, $108^{\circ}00'03''\text{E}$, geographic WGS84), Shaanxi, China (Figure 1). The soil texture type was classified as fine loam with organic content of 15.3–19.1 g/kg, nitrate–nitrogen ($\text{NO}_3\text{-N}$) content of 8.21–41.71 mg/kg, ammonium nitrogen ($\text{NH}_3\text{-N}$) content of 0.45–8.2 mg/kg, available potassium content of 60.58–109.61 mg/kg, and available phosphorus content of 3.14–21.18 mg/kg at the top 30-cm soil layer. This experimental station was characterized by a warm temperate and semi-humid continental monsoon climate. The minimum and maximum temperatures were 10°C and 40°C , respectively.

2.2. Data Collection

Six cameras were installed at the quadrotor UAV model DJI-Phantom 4 Multispectral (DJI-P4M, SZ DJI Technology Co., Ltd., Shenzhen, China) containing five sensors of different bands and an RGB visible light sensor. Please see Table 1 for details of the UAV. The drone observation dates (and Zadoks scale [14]) were 23 March (ZS31), 8 April (ZS36), 14 April (ZS44), 29 April (ZS56), 7 May (ZS65), 17 May (ZS75), and 27 May (ZS80). The wheat was harvested on June 11 (ZS90), and the images of different dates were synthesized and spliced into one image using Pix4D, processed by ENVI analysis software (ENVI 5.0, Exelis VIS, Boulder, CO, USA), and then classified into 56 types of VIs (Table 2).

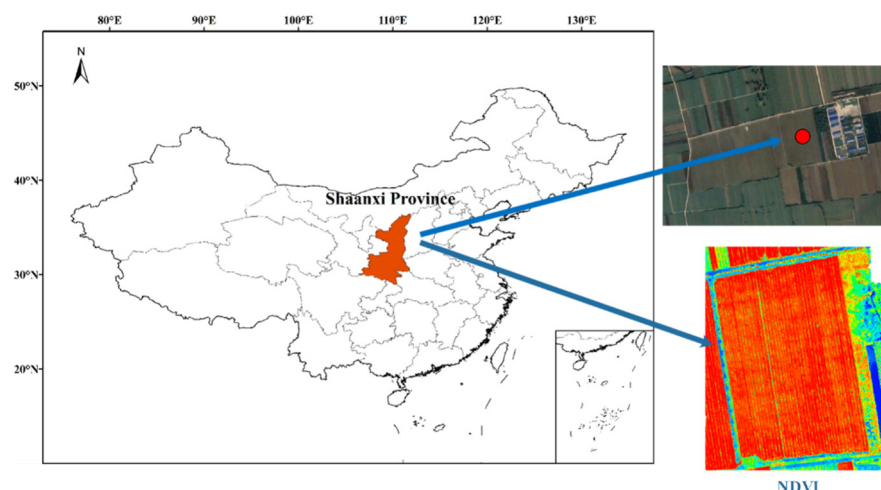


Figure 1. Location of the study region.

Wheat AGB was observed seven times on the same date UAV data were collected. In the test area, 32 sample plots were randomly selected to record wheat density. The area of each plot was 1×1 m, respectively. On the UAV observation day, ten wheat strains were randomly selected, harvested from the aboveground part of the stem, inactivated in an oven at 75°C , and then dried to constant weight at 105°C . The VIs and AGB data were split into two parts: 75% for training and validation of the ML model and 25% for model testing. Moreover, 75% of the data were randomly divided into five folds, four of which were used to train the model and the fifth to validate the model. This product was repeated five times and all folds were used to train the model four times and validate the model one time. The statistical value of AGB can be found in Table 3.

Table 1. Basic information on UAV remote sensing observation equipment.

UAV	Description	Sensor	Description
Name	DJI-Phantom 4 pro	Type	FC6360
Flight altitude above ground level	50 m	Bands	Blue ($450 \text{ nm} \pm 16 \text{ nm}$), Green ($560 \text{ nm} \pm 16 \text{ nm}$), Red ($650 \text{ nm} \pm 16 \text{ nm}$), Red edge ($730 \text{ nm} \pm 16 \text{ nm}$), NIR ($840 \text{ nm} \pm 26 \text{ nm}$)
Flight speed	4 m/s	Number of images	11,200
Satellite systems	GPS	Shutter speed	1.2
Forward overlap	80%	ISO sensibility	ISO-200
Side overlap	80%	Image dimension	1600×1300
Field of view	90°	Resolution	5.3 cm/pixel
Shooting interval	2 s	Image format	JPEG, TIFF

Table 2. VIs used in this study.

No.	VI	Formula	Reference
1	Green Ratio Vegetation Index (GRVI)	$GRVI = NIR^1 / G$	Buschmann and Nagel (1993) [15]
2	Green Difference Vegetation Index (GDVI)	$GDVI = NIR - G$	Tucker (1979) [16]
3	Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = (NIR - G) / (NIR + G)$	Gitelson and Kaufman (1996) [17]
4	Green Wide Dynamic Range Vegetation Index (GWDRVI)	$GWDRVI = (0.12 \times NIR - G) / (0.12 \times NIR + G)$	Cao et al. (2013) [18]
5	Green Chlorophyll Index (CIg)	$CIg = NIR / G - 1$	Gitelson (2005) [19]
6	Modified Green Simple Ratio (MSR_G)	$MSR_G = (NIR / G - 1) / \sqrt{NIR / G + 1}$	Cao et al. (2013) [18]
7	Green Soil-Adjusted Vegetation Index (GSAVI)	$GSAVI = 1.5 \times (NIR - G) / (NIR + G + 0.5)$	Sripada et al. (2016) [20]
8	Green Re-normalized Different Vegetation Index (GRDVI)	$GRDVI = (NIR - G) / \sqrt{NIR + G}$	Cao et al. (2013) [18]
9	Normalized Green Index (NGI)	$NGI = G / (NIR + G + RE)$	Sripada et al. (2016) [20]
10	Normalized Red Edge Index (NREI)	$NREI = RE / (NIR + G + RE)$	Cao et al. (2013) [18]
11	Normalized Red Index (NRI)	$NRI = R / (NIR + R + RE)$	Lu et al. (2014) [21]
12	Normalized NIR Index (NNIR)	$NNIR = NIR / (NIR + R + RE)$	Sripada et al. (2016) [20]
13	Modified Double Difference Index (MDD)	$MDD = (NIR - RE) / (RE - G)$	Lu et al. (2014) [21]
14	Modified Normalized Difference Index (MNDI)	$MNDI = (NIR - RE) / (NIR - G)$	Cao et al. (2013) [18]
15	Modified Enhanced Vegetation Index (MEVI)	$MEVI = 2.5 \times (NIR - RE) / (NIR + 6 \times RE - 7.5 \times G + 1)$	Cao et al. (2013) [18]
16	Modified Normalized Difference Red Edge (MNDRE)	$MNDRE = (NIR - RE - 2 \times G) / (NIR + RE - 2 \times G)$	Cao et al. (2013) [18]
17	Modified Chlorophyll Absorption In Reflectance Index 1 (MCARI1)	$MCARI1 = ((NIR - RE) - 0.2 \times (NIR - R)) \times (NIR / RE)$	Haboudane et al. (2004) [22]
18	Modified Chlorophyll Absorption In Reflectance Index 2 (MCARI2)	$MCARI2 = 1.5 \times (2.5 \times (NIR - R) - 1.3 \times (NIR - RE)) / \sqrt{(2 \times NIR + 1)^2 - (6 \times NIR - 5 \times \sqrt{R} - 0.5)}$	Haboudane et al. (2004) [22]
19	Normalized Difference Vegetation Index (NDVI)	$NDVI = (NIR - R) / (NIR + R)$	Rouse et al. (1974) [23]
20	Ratio Vegetation Index (RVI)	$RVI = NIR / R$	Jordan et al. (1969) [24]
21	Difference Vegetation Index (DVI)	$DVI = NIR - R$	Tucker (1979) [16]
22	Renormalized Difference Vegetation Index (RDVI)	$RDVI = (NIR - R) / \sqrt{NIR + R}$	Roujean and Breon (1995) [25]
23	Wide Dynamic Range Vegetation Index (WDRVI)	$WDRVI = (0.12 \times NIR - R) / (0.12 \times NIR + R)$	Gitelson et al. (2004) [19]
24	Soil-Adjusted Vegetation Index (SAVI)	$SAVI = 1.5 \times (NIR - R) / (NIR + R + 0.5)$	Huete et al. (1988) [26]
25	Transformed Normalized Vegetation Index (TNDVI)	$TNDVI = \sqrt{(NIR - R) / (NIR + R) + 0.5}$	Sandham (1997) [27]
26	Modified Simple Ratio (MSR)	$MSR = (NIR / R - 1) / \sqrt{NIR / R + 1}$	Chen (1996) [28]
27	Optimal Vegetation Index (Vlopt)	$Vlopt = 1.45 \times (NIR^2 + 1) / (R + 0.45)$	Reyniers et al. (2006) [29]
28	MERIS Terrestrial Chlorophyll Index (MTCI)	$MTCI = (NIR - RE) / (RE - R)$	Dash and Curran (2004) [30]
29	Nonlinear Index (NLI)	$NLI = (NIR^2 - R) / (NIR^2 + R)$	Goel and Qin (1994) [31]
30	Modified Nonlinear Index (MNLI)	$MNLI = 1.5 \times (NIR^2 - R) / (NIR^2 + R + 0.5)$	Gong et al. (2003) [32]
31	NDVI $\times$ RVI	$NDVI_RVI = (NIR^2 - R) / (NIR + R^2)$	Gong et al. (2003) [32]
32	SAVI $\times$ SR	$SAVI_SR = (NIR^2 - R) / (NIR + R + 0.5) \times R$	Gong et al. (2003) [32]
33	Normalized Difference Red Edge (NDRE)	$NDRE = (NIR - RE) / (NIR + RE)$	Barnes et al. (2000) [33]
34	Red Edge Ratio Vegetation Index (RERVI)	$RERVI = (NIR / RE)$	Gitelson et al. (1996) [17]

Table 2. Cont.

No.	VI	Formula	Reference
35	Red Edge Difference Vegetation Index (REDVI)	$REDVI = (NIR - RE)$	Cao et al. (2013) [18]
36	Red Edge Re-normalized Different Vegetation Index (RERDVI) Red Edge Wide Dynamic Range Vegetation Index (REWDRVI) Red Edge Soil-Adjusted Vegetation Index (RESAVI)	$REWDRVI = (0.12 \times NIR - R)/(0.12 \times NIR + R)$	Cao et al. (2013) [18]
37	Red Edge Optimal Soil-Adjusted Vegetation Index (REOSAVI)	$REOSAVI = 1.5 \times (NIR - RE)/(NIR + RE + 0.5)$	Cao et al. (2013) [18]
38	Optimized Red Edge Vegetation Index (REVI _{opt})	$REVI_{opt} = 100 (\log(NIR) - \log(RE))$	Jasper et al. (2009) [34]
39	Red Edge Chlorophyll Index (CI _{re})	$CI_{re} = NIR/RE - 1$	Gitelson et al. (2003) [35]
40	Modified Red Edge Simple Ratio (MSR _{RE})	$MSR_{RE} = (NIR/RE - 1)/\sqrt{NIR/RE + 1}$	Lu et al. (2014) [21]
41	Red Edge Normalized Difference Vegetation Index (RENDVI)	$RENDVI = (NIR - RE)/(NIR + RE)$	Elsayed et al. (2015) [36]
42	Red Edge Simple Ratio (RESR)	$RESR = RE/R$	Erdle et al. (2011) [37]
43	Modified Red Edge Difference Vegetation Index (MREDVI) MERIS Terrestrial Chlorophyll Index (MTCI)	$MREDVI = RE - R$	Cao et al. (2013) [18]
44	DATT Index (DATT)	$DATT = (NIR - RE)/(NIR - R)$	Datt (1999) [38]
45	Normalized Near-Infrared Index (NNIRI)	$NNIRI = NIR/(NIR + RE + R)$	Lu et al. (2014) [21]
46	Normalized Red Edge Index (NREI)	$NREI = RE/(NIR + RE + R)$	Lu et al. (2014) [21]
47	Normalized Red Index (NRI)	$NRI = R/(NIR + RE + R)$	Lu et al. (2014) [21]
48	Modified Double Difference Index (MDD)	$MDD_R = NIR - R$	Lu et al. (2014) [21]
49	Modified Red Edge Simple Ratio (MRESR)	$MRESR = (NIR - R)/(RE - R)$	Lu et al. (2014) [21]
50	Modified Normalized Difference Index (MNDI)	$MNDI = (NIR - RE)/(NIR + RE - 2 \times R)$	Lu et al. (2014) [21]
51	Modified Enhanced Vegetation Index (MEVI)	$MEVI_R = 2.5 \times (NIR - RE)/(NIR + 6 \times RE - 7.5 \times R + 1)$	Lu et al. (2014) [21]
52	Modified Normalized Difference Red Edge (MNDRE2)	$MNDRE2 = (NIR - RE + 2 \times R)/(NIR + RE - 2 \times R)$	Lu et al. (2014) [21]
53	Red Edge Transformed Vegetation Index (RETVI)	$RETVI = 0.5 \times (120 \times (NIR - R) - 200 \times (RE - R))$	Lu et al. (2014) [21]
54	Modified Chlorophyll Absorption In Reflectance Index 3 (MCARI3)	$MCARI3 = ((NIR - RE) - 0.2 \times (NIR - R)) \times (NIR/RE)$	Haboudane et al. (2004) [22]
55	Modified Chlorophyll Absorption In Reflectance Index 4 (MCARI4)	$MCARI4 = (1.5 \times (2.5 \times (NIR - G) - 1.3 \times (NIR - RE)))/(\sqrt{(2 \times NIR + 1)^2 - (6 \times NIR - 5 \times \sqrt{G})} - 0.5)$	Haboudane et al. (2004) [22]
56	Modified Red Edge Transformed Vegetation Index (MRETVI) Modified Canopy Chlorophyll Content Index (MCCCI)	$MRETVI = 1.2 \times (1.2 \times (NIR - R) - 2.5 \times (RE - R))$	Lu et al. (2014) [21]

Note 1: G, R, RE, NIR indicates green, red, red edge, and near-infrared band reflectance, respectively.

Table 3. AGB statistical characteristics during training and testing periods.

Wheat AGB	Max	Min	Mean	Median	Std.	CV	Skewness
Training	2.31	0.25	1.16	1.15	1.10	0.49	0.21
Testing	2.11	0.14	1.10	1.04	1.00	0.54	0.28

2.3. Artificial Intelligence Methods

2.3.1. Multilayer Perceptron (MLP)

MLP is a type of forward feedback ANN, which is called a deep feedforward network. MLP has the property of mapping and can be considered a directed graph from the input vector to the output vector, where the data points between input and output vectors are nodes. The node layer is formed in homogenous vector data points for data transmission. During network data transmission from an input node to a lower layer, other nodes are neurons with activation functions. The neural network training algorithm is the most crucial part of the model application. The backpropagation algorithm is adopted in MLP training, and supervised training can overcome the weakness of unidentifiable data. A feedforward neural network can contain three types of nodes: input, implicit, and output nodes [39]. In this study, a three-layer network with an input layer, a hidden layer, and an output layer was used. The number of nodes in the input layer equals the number of input features. The trial and error method was used to debug the number of neurons in the hidden layer, and the number of nodes in the output layer was 1.

2.3.2. Gaussian Process Regression (GPR)

GPR is a nonparametric supervised model that uses Gaussian process priors to perform regression analysis on data. The model hypothesis of GPR includes noise (regression residual) and Gaussian process prior, and its solution is obtained as per Bayesian inference. Without constraining the kernel form, GPR is theoretically a universal approximation of arbitrary continuous functions in compact space [40]. Moreover, GPR can provide a posteriori of predicted results, which has an analytic form when the likelihood is a normal distribution. Therefore, GPR is a probabilistic model with generality and analyzability. We did not tune the parameters of GPR because it is a parameter-free algorithm.

2.3.3. Support Vector Machine (SVM)

The SVM model was established by Vapnik [41]. General models usually adopt empirical risk minimization theory, i.e., minimum cumulative error as the goal. To reduce the overfitting problem, the structural risk minimization theory was used to completely discard some points as noise. SVM models can be used for classification, pattern recognition, and regression analysis. For the regression analysis, the original problem is transformed into a convex quadratic programming problem so that the model has a unique global optimal solution. At the core of SVM are kernel functions, which can implicitly transform original low-dimensional input datasets into higher-dimensional feature spaces. The kernel function based on the radial basis function was used in this study because it has better applicability in prediction than other kernel functions, such as linear, polynomial, and S-shaped functions. Please refer to the literature for more information on the SVM model. A grid-search method was used to tune the parameters of SVM, which are the regularization coefficient and the width of the kernel function. The range of these two parameters was from 0.01 to 10,000; the step was 10 times the previous value.

2.3.4. Random Forest (RF)

The RF model, proposed by Breiman [42], integrates a set of decision tree ensemble learning methods with control variances using the “Bagging” idea. Compared with a single decision tree, this method integrates a set of CARTs to improve the model generalization ability and determine a more accurate and stable prediction pattern [42]. RF trains each tree using only a subset of the dataset, known as bootstrap samples. All decision trees

predict results based on the voting mechanism. The remaining samples that do not belong to the bootstrap set, namely, out-of-bag samples, are used for internal cross-validation to improve accuracy and assist data importance assessment. A quantitative measure of the contribution of each input auxiliary data to the predictive program is known as RF variable importance. Compared with other ML algorithms, RF is insensitive to noise and overtraining. A grid-search method was also used. Two parameters were tuned in this study, namely, the number of rounds ranged from 100 to 1000 with step 100 and the max tree depth ranged from 2 to 30 with step 5.

2.3.5. Gradient Boosting Decision Tree (GBDT)

The GBDT model proposed by Friedman [43] has been extensively used for classification regression problems. It is based on the boosting strategy and uses the classification and regression tree (CART) as a weak classifier. In the GBDT model, a weak learner measure observes errors in each node and uses test functions to segment nodes. Weak learners require to be successively set up, and the residual of the previous weak classifier will be learned by the next weak classifier as the input state. The GBDT model uses the gradient descent strategy to accelerate the convergence of weak classifiers. Note that additional details are reported by Friedman [40]. This study tuned three parameters: the number of trees [100, 1000], the maximum tree depth [2, 30], and the learning rate [0.001, 0.3]. The grid search technique was used.

2.3.6. XGBoost

XGBoost is a novel tree-based ensemble algorithm based on GBDT, which has many improvements. First, XGBoost uses a presorting mechanism to find the optimal split point. Feature columns are sorted and stored in blocks [44]. All features are presorted according to feature values and are stored in the first traversal. When traversing the segmentation points again, the best segmentation points on features are found with the cost of $O(\text{data number})$. The presorted data will be stored in memory as a block structure, which can be used repeatedly in subsequent iterations to significantly reduce the amount of computation [44]. Therefore, XGBoost supports parallel computing but not characteristic parallelism. Second, XGBoost optimizes the loss function into a second-order Taylor expansion form using both the first and second derivatives to accelerate the optimization speed. Third, the complexity of the tree model is added into the regular term to participate in the loss function calculation to avoid the overfitting problem. This study tuned three parameters: the number of trees [100, 1000], the maximum tree depth [2, 30], and the learning rate [0.001, 0.3]. Moreover, the grid search technique was employed.

2.3.7. Grasshopper Optimization Algorithm (GOA)

In this algorithm, the exploration process is realized by imitating grasshoppers' sudden movement behaviors when searching for a food source, and the mining process is realized by imitating the grasshoppers' moving to a local food source and consuming food [45]. These behaviors of grasshoppers are performed naturally. The algorithm simulates the two behaviors of grasshoppers through a location update formula. The next location of grasshoppers is updated according to the current location, global target value, and locations of all other grasshoppers. The updated simulation mathematical model of grasshopper position in the D -dimensional space is as follows:

$$X_i = c_w \left(\sum_{\substack{j=1 \\ j \neq i}}^N c_n \frac{\varepsilon - \eta}{2} s(d_{ij}) \frac{x_j - x_i}{d_{ij}} \right) + \bar{T}_d \quad (1)$$

where X_i is the i -th grasshopper position vector; c_w is similar to the inertial weight in the particle swarm algorithm, which balances the exploration and mining processes of the entire group around the target; ε and η are the upper and lower bounds of the search scope, respectively; s is the attraction function; $d_{ij} = |x_j - x_i|$ is the distance between the i -th and j -th grasshoppers; x_i and x_j are the location vectors of the i -th and j -th grasshoppers, respectively. $\bar{T}_d$ is the optimal target value currently searched; c_n is the parameter, which is the decreasing coefficient of the contraction comfort, repulsive, and attraction zones; the calculation formula is as follows:

$$c_w = c_n = c_{max} - t \frac{c_{max} - c_{min}}{L} \quad (2)$$

where c_{max} is the maximum value, t is the current number of iterations, c_{min} is the minimum value, and L is the maximum number of iterations.

$$s(r) = ue^{\frac{-r}{\tau}} - e^{-r} \quad (3)$$

where $s(r)$ is the attraction function, u is the strength of attraction, r is the distance variable, and τ is the range of attraction.

The literature explains how the functions of social forces (attraction or repulsion) change with the distance between grasshoppers. When a grasshopper is 2.0 units away from another grasshopper, there is neither an attraction nor repulsion—the comfort zone. When the value of the function $s(r)$ is greater than 0, the grasshoppers are attracted to each other: the attraction zone. When the value of the function $s(r)$ is less than 0, grasshoppers are in a state of exclusion from each other: the repulsion zone. Equation (4) is the distance between the i -th and j -th grasshoppers:

$$d_{ij} = |x_j - x_i| \quad (4)$$

Equation (5) is the unit vector from the i -th to j -th grasshopper:

$$\hat{d}_{ij} = \frac{x_j - x_i}{d_{ij}} \quad (5)$$

Through the interaction of grasshoppers, the algorithm gradually approaches the food source and eventually consumes the food with the linear decreasing algorithm of parameters to search the global optimal value. In this study, GOA was used to optimize the parameters of the SVM and XGBoost models (Figure 2). The parameters of the SVM model are the regularization coefficient [0.01, 100] and the width of the Gaussian kernel function [0.01, 100]. The parameters of the XGBoost model are the number of trees [100, 800], learning rate [0.01, 0.3], proportion of training data [0.51, 1], and weight of child nodes [1, 25].

2.3.8. Particle Swarm Optimization (PSO) Algorithm

Kennedy and Eberhart proposed the PSO algorithm in 1995 [46]. The similarity between PSO and GA is in the initialization of the population; both randomly generate initial solutions, but in PSO, a random velocity and position are set for each potential solution, called particles, and the particles fly in the problem space to find the optimal solution. Particles have a simple behavior: they simulate the “success” of those around them as well as the “success” of the particle itself. The emergence of simple behavior groups from these simple individuals allows for an optimal solution search in high-dimensional spaces. The original PSO is frequently plagued by the premature convergence problem. The fitness-distance-ratio method [47] can resolve this problem, and the new equation is obtained as follows:

$$v_{i,d}^{k+1} = w \cdot v_{i,d}^k + c_1 \cdot r_1 \cdot (p_{i,d}^k - x_{i,d}^k) + c_2 \cdot r_2 \cdot (p_{g,d}^k - x_{i,d}^k) + c_3 \cdot r_3 \cdot (p_{n,d}^k - x_{i,d}^k) \quad (6)$$

$$x_{i,d}^{k+1} = x_{i,d}^k + v_{i,d}^k \quad (7)$$

where x_i is the location of the i -th particle; v_i is the corresponding velocity; p_i is the i -th particle's best position; p_g is the global best position; p_n is the nearby best position. w is the inertia weight; r_1 , r_2 , and r_3 are the random numbers; c_1 , c_2 , and c_3 are the acceleration constants, and these values are equal to 2 in this study. The range of x_i was the same as the GOA.

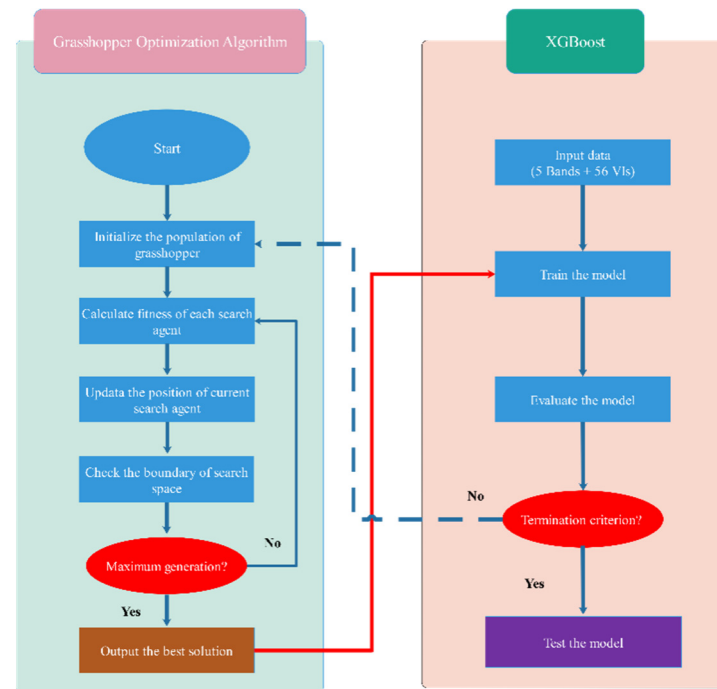


Figure 2. Flowchart of the GOA-XGB model.

2.3.9. Whale Optimization Algorithm (WOA)

Mirjalili et al. [48] developed the WOA algorithm, which simulates humpback whale prey behavior. The whales hunt for food using a technique known as bubble-net hunting, in which bubbles are created by encircling or passing through the “9-shaped path.” When hunting, the humpback whale dives into the water more than 10 m deep and creates a spiral-shaped bubble that surrounds its prey. The bubble formed by the whale eventually contraction-rises to the surface, trapping prey in the swarm. The optimization process in the WOA algorithm begins by randomly initializing the whale population. The whales then use the wraparound method or the bubble-net hunting method to find (optimal) locations of their prey. Whales use two mechanisms to locate and attack their prey. First, the prey is surrounded, and the second involves making a net of bubbles. The WOA can be calculated as follows:

(1) Searching and encircling prey:

$$X_i^{t+1} = X_{rand} - A|C \cdot X_{rand} - X_i^t| \quad (8)$$

where A and C are coefficients and can be computed as follows:

$$A = 2ar - a \quad (9)$$

$$C = 2r \quad (10)$$

where a is the linearly decreasing coefficient from 2 to 0, r is the random number range $[0, 1]$. If $A > 1$, then prey behavior can be achieved around the best location of the whale:

$$X_i^{t+1} = g_{best} - A|C \cdot g_{best} - X_i^t| \quad (11)$$

(2) Spirally updating location:

In this section, some whales will randomly prey around the best location of the whale and others will prey obey the “9” path:

$$X_i^{t+1} = \begin{cases} g_{best} - A|C \cdot g_{best} - X_i^t| & p < 0.5 \\ |C \cdot g_{best} - X_i^t| \cdot \exp(bl) \cdot \cos(2\pi l) + g_{best} & p \geq 0.5 \end{cases} \quad (12)$$

where p is the random number ranging between 0 and 1, l varies from 0 to 1, b is a constant to describe the spiral shape and is set by 1 in this study.

2.3.10. Tune the Parameters of the Hybrid Machine Learning Models

In this study, the population of PSO, WOA, and GOA was set to 50, with a total of 200 iterations. The parameter range of different hybrid machine learning models was the same as that of their corresponding standalone models. To achieve these models, except for LSTM, the R language (R package 4.4) was used.

2.4. Statistical Indicators

We employed five commonly used statistical indicators to evaluate the performances of the model from different dimensions. The calculation formulas are as follows:

- Determination Coefficient (R^2)

$$R^2 = \frac{\left[\sum_{i=1}^n (O_{i,m} - \bar{O}_{i,m})(O_{i,e} - \bar{O}_{i,e}) \right]^2}{\sum_{i=1}^n (O_{i,m} - \bar{O}_{i,m})^2 \sum_{i=1}^n (O_{i,e} - \bar{O}_{i,e})^2} \quad (13)$$

- Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_{i,m} - O_{i,e})^2} \quad (14)$$

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_{i,m} - O_{i,e}| \quad (15)$$

- Normalized Root Mean Square Error (NRMSE)

$$NRMSE = RMSE / \bar{O}_{i,m} \quad (16)$$

- Percent of Bias (PBIAS)

$$PBIAS = \frac{O_{i,m} - O_{i,e}}{O_{i,m}} \times 100\% \quad (17)$$

where n is the total number of data; O_e and O_m are the estimated AGB values by the ML models and measured AGB values, respectively; $\bar{O}_m$ is the mean AGB value.

R^2 is between 0 and 1; a value closer to 1 indicates better regression fitting between the estimated and measured AGB; hence, the better model performance. Meanwhile, RMSE or MAE being closer to 0 indicates a better model performance. RMSE is generally useful when model errors follow a normal distribution, whereas MAE is suitable for models with a uniform error distribution. To evaluate the model performance with NRMSE, four levels of indication are used, which are $NRMSE \leq 10\%$, $20\% \geq NRMSE > 10\%$,

$30\% \geq NRMSE > 20\%$, and $NRMSE > 40\%$, corresponding to perfect, good, fair, and poor model performances, respectively.

3. Results

3.1. Linear Regression (LR) Model

The LR model showed that when there was only one input factor (Band G), the agreement between the simulated and observed AGB values was poor; $R^2 = 0.383$ (Table 4). When the number of input factors was 2 and 3, the multiple LR (MLR) accuracies slightly improved compared with those of the single-factor LR model, but the RMSE and MAE decreased by approximately 7%. When the number of input factors was 4, the model accuracy was no longer improved compared with the top three MLR models.

Table 4. Stepwise MLR models and GOA-MLR.

ID	Regression Equation	RMSE	R ²	MAE	NRMSE	PBIAS
MLR						
LR	$0.286 + 0.152 \times G$	0.851	0.383	0.693	0.399	0
MLR1	$0.760 + 0.088 \times G - 2.04 \times Cig$	0.842	0.396	0.687	0.395	0
MLR2	$14.957 + 0.3550G + 31.28 \times Cig - 2.312 \times MSR_G$	0.796	0.461	0.626	0.373	0
MLR3	$-5.306 - 0.304 \times G - 21.125 \times Cig + 0.151 \times MSR_G + 0.022 \times GNDVI$	0.796	0.461	0.626	0.373	0
GOA-MLR						
GOA-LR	$0.286 + 0.152 \times G$	0.851	0.383	0.693	0.399	0
GOA-MLR1	$-1.326 - 1.1606 \times MSR_G + 1.506 \times NGI$	0.718	0.560	0.564	0.337	0
GOA-MLR2	$0.891 - 1.205Cig + 1.028 \times GRDVI + 0.027 \times MCARI3$	0.67	0.587	0.553	0.327	0
GOA-MLR3	$2.591 - 1.017GNDVI + 0.023 \times Cig + 0.07835047 \times GNDVI + 0.001 \times NGI$	0.701	0.582	0.558	0.328	0

When the number of input factors was 1, the GOA-LR model was identical to the corresponding LR model, showing that the optimization algorithm could not reflect its advantages when a single factor was used. When the number of input factors was 2, the GOA-MLR model (GOA-MLR1) had different input factors from the corresponding MLR model, and its accuracy was significantly better than that of the MLR1 model; RMSE and MAE decreased by 14.7% and 17.9%, respectively. When the number of input factors was 3, the GOA-MLR2 model was also superior to the MLR2 model; RMSE and MAE decreased by 15.8% and 11.6%, respectively. In addition, the accuracy of the GOA-MLR2 model slightly improved compared with the GOA-MLR1 model. However, the GOA-MLR2 model had no advantage over the GOA-MLR1 and GOA-MLR3 models, indicating that the MLR model could no longer describe the nonlinear relationship between VIs and AGB when the number of VIs was more than three.

3.2. ML Models with Single VI as Input

To investigate the VI that is most associated with AGB, 5 bands and 56 types of VIs were used to drive different ML models. The results are illustrated in Figure 3 and Table 5. Near-infrared (NIR) band ranked first in all six ML models. However, the accuracy of different ML models varied. When NIR was used as the input, the GPR model had the highest accuracy, with $RMSE = 0.318$, $R^2 = 0.725$, $MAE = 0.267$, and $NRMSE = 0.173$, and the deviation was small. The XGBoost model ranked second, followed by the SVM, MLP, GBDT, and RF models. Six ML models of VIs, which ranked second in accuracy, gave different answers, among which GBDT, RF, SVM, and XGB models showed Cig, whereas GPR and MLP models recommended MSR_G and GRVI, respectively. In terms of accuracy, the XGBoost model was the highest, which had an error slightly worse than the XGBoost and GPR models with NIR as input. In addition, the GPR model fed by MSR_G also had high accuracy and was superior to the GBDT, RF, and SVM models with Cig as input. Except for MLP, the third-ranked VI-driven ML models had the same accuracy as the

second ones, indicating that, for the ML models, different VIs have numerical differences, but their effects are equivalent.

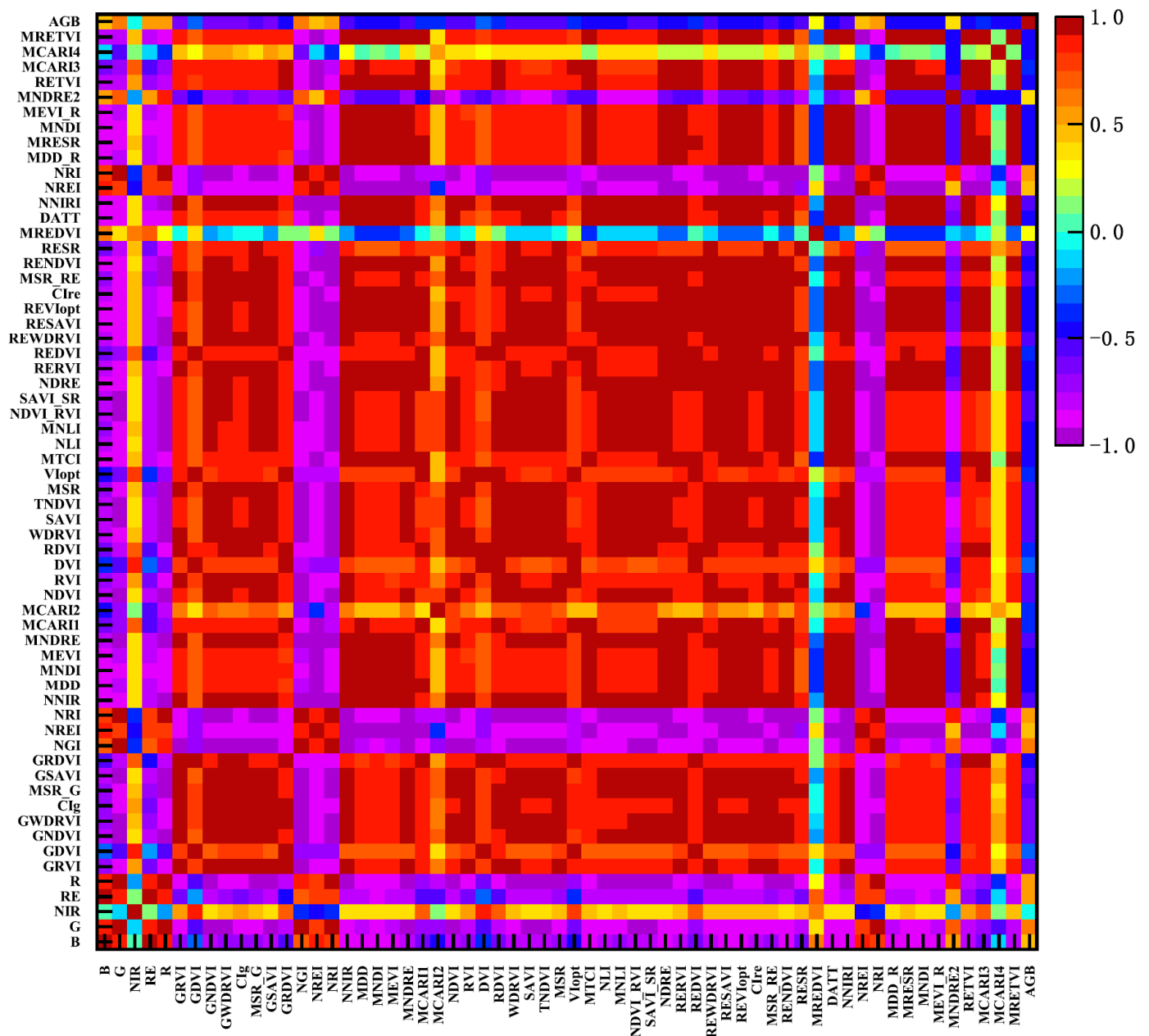


Figure 3. Relationship of bands, Vis, and AGB.

3.3. ML Models with All Features as Input

To preliminarily evaluate the performance of eight different ML models, all features, i.e., 5 bands and 56 types of VIs, were used as input to drive the models (Table 6). The results showed that all ML models yielded good accuracies with NRMSE less than 0.2. In addition, there were no obvious overestimation and underestimation problems in all models ($PBIAS > 0.1$). However, different models varied slightly in accuracy; that is, the GOA-XGB1 model was superior to other models, with $RMSE = 0.232 \text{ kg m}^{-2}$, $R^2 = 0.847$ (Figure 4), and $MAE = 0.178 \text{ kg m}^{-2}$. The XGBoost1 model was slightly better than GOA in the consistency of simulated and predicted values, but in accuracy, it was slightly worse than the GOA-XGB1 model; RMSE and MAE increased by 6% and 5%, respectively (Figure 4). The RF1 model ranked third, with 13.7% and 9.5% higher RMSE and MAE than the GOA-XGB1 model, respectively. The GBDT model was comparable to the RF model.

Compared with the GOA-XGB1 model, RMSE and MAE increased by 16.8% and 10.1%, respectively. The RMSE of the GPR1 model was comparable to the GBDT1 model, but its MAE was significantly lower than that of the GBDT1 model, which was 20% higher than that of the GOA-XGB1 model. The performance of the SVM model was poor; RMSE and MAE were 25% higher than those of the GOA-XGB1 model. Even when GOA was employed to optimize SVM's parameters, the accuracy was not significantly improved. The MLP model was significantly inferior to other models and could only explain 72.2% of the data from the perspective of R^2 ; its RMSE and MAE were 43% higher than those of the best model.

Table 5. Top three VIs for estimating AGB based on six different ML models.

Model/VI	RMSE (kg m ⁻²)	R ²	MAE (kg m ⁻²)	NRMSE	PBIAS (kg m ⁻²)
MLP					
NIR	0.375	0.631	0.293	0.205	0.052
GRVI	0.465	0.443	0.364	0.253	0.072
NREI	0.473	0.411	0.386	0.258	0.062
GPR					
NIR	0.318	0.725	0.267	0.173	0.020
MSR_G	0.348	0.701	0.283	0.190	0.068
GDVI	0.348	0.701	0.283	0.190	0.068
SVM					
NIR	0.349	0.677	0.278	0.190	0.050
CIg	0.377	0.654	0.312	0.205	0.074
MSR_G	0.377	0.669	0.310	0.205	0.085
RF					
NIR	0.391	0.642	0.307	0.213	0.029
CIg	0.413	0.583	0.322	0.225	0.043
GWDRVI	0.413	0.581	0.322	0.225	0.043
GBDT					
NIR	0.382	0.643	0.300	0.208	0.029
CIg	0.401	0.599	0.320	0.219	0.055
GWDRVI	0.402	0.595	0.321	0.219	0.053
XGB					
NIR	0.327	0.708	0.267	0.178	−0.036
CIg	0.339	0.679	0.281	0.185	−0.016
MSR_G	0.339	0.679	0.281	0.185	−0.016

Table 6. Statistics indicators of twelve ML models with all features as input.

Model	RMSE (kg m ⁻²)	R ²	MAE (kg m ⁻²)	NRMSE	PBIAS (kg m ⁻²)
MLP1	0.334	0.722	0.256	0.182	0.064
GPR1	0.276	0.801	0.214	0.151	0.009
SVM1	0.301	0.747	0.223	0.164	−0.015
PSO-SVM1	0.299	0.750	0.220	0.162	0.011
WOA-SVM1	0.291	0.751	0.218	0.162	0.201
GOA-SVM1	0.298	0.752	0.239	0.162	0.015
RF1	0.264	0.815	0.195	0.144	0.035
GBDT1	0.271	0.808	0.196	0.148	0.038
XGBoost1	0.246	0.854	0.187	0.134	−0.042
PSO-XGB1	0.247	0.84	0.192	0.224	0.175
WOA-XGB1	0.240	0.842	0.182	0.218	0.165
GOA-XGB1	0.232	0.847	0.178	0.127	−0.004

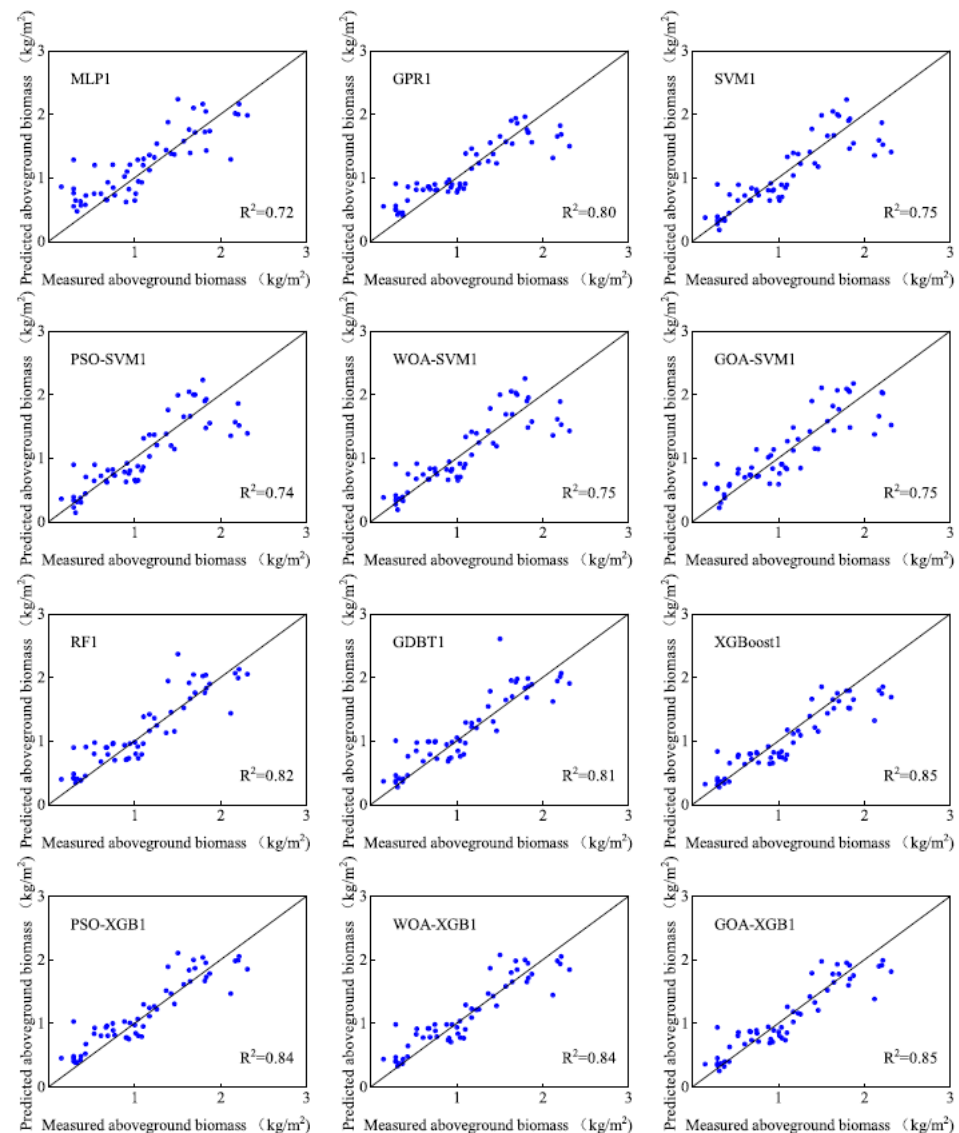


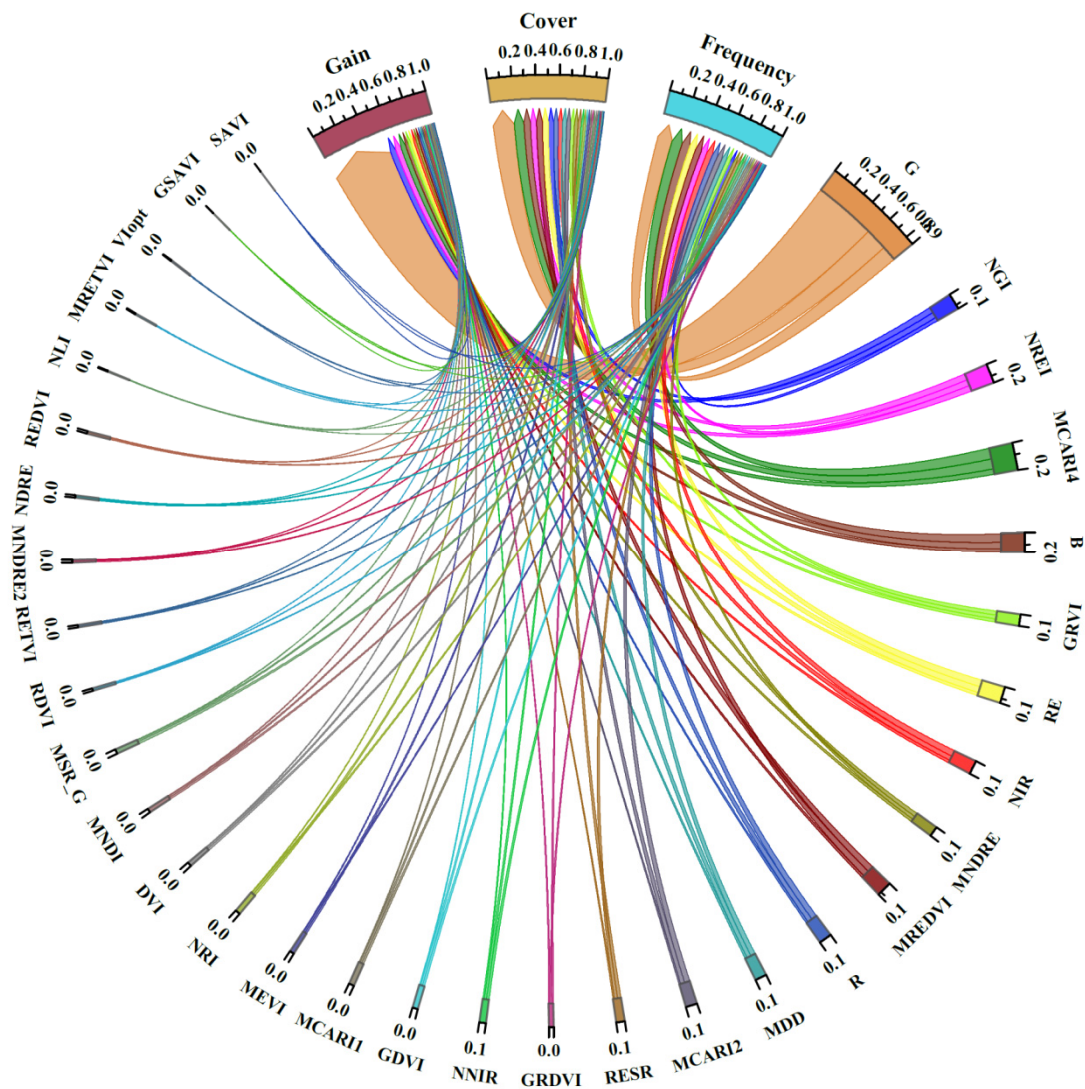
Figure 4. Scatter plot of AGB estimated by eight machine learning models vs. measured AGB.

3.4. ML Models with Optimized Features as Input

When there are too many input factors in an ML model, especially irrelevant factors, the model will try to explain the relationship between noises and targets, resulting in the decline of model prediction ability. To further select important factors, we used the factors of the GOA-XGB1 model (whose gain value was greater than 0.01) as input to reinsert the eight ML models and evaluate whether the model accuracy improved. The results are shown in Table 7 and Figure 5. All eight ML models had good accuracy ($\text{NRMSE} < 0.2$). In general, each model did not have an overestimation or underestimation problem. The GOA-XGB2 model was superior to other models, with $\text{RMSE} = 0.226 \text{ kg m}^{-2}$, $R^2 = 0.855$, and $\text{MAE} = 0.172 \text{ kg m}^{-2}$. The RMSE and MAE of the XGBoost2 model were about 5% higher than those of the GOA-XGB2 model, and the XGBoost2 ranked second in accuracy. The RF2 and GPR2 models performed slightly worse than the XGB2 model and slightly better than the GBDT2 model. The performances of the SVM2 and GOA-SVM2 models were comparable and only better than that of the MLP2 model. Figure 6 shows that the importance rankings of the top 16 features from GOA-XGB2 were entirely consistent with GOA-XGB1 (Figure 6). This demonstrates that removing redundant features improves the model's interpretability.

Table 7. Statistics indicators of twelve ML models with optimized features as input.

Model	RMSE (kg m ⁻²)	R ²	MAE (kg m ⁻²)	NRMSE	PBIAS (kg m ⁻²)
MLP2	0.355	0.647	0.281	0.193	0.0258
GPR2	0.256	0.826	0.201	0.140	0.015
SVM2	0.288	0.770	0.219	0.157	−0.001
PSO-SVM2	0.287	0.780	0.226	0.261	0.206
WOA-SVM2	0.287	0.771	0.222	0.263	0.201
GOA-SVM2	0.284	0.771	0.217	0.155	−0.001
RF2	0.253	0.831	0.190	0.138	0.032
GBDT2	0.272	0.805	0.194	0.148	0.035
XGBoost2	0.243	0.858	0.181	0.133	−0.043
PSO-XGB2	0.249	0.842	0.193	0.226	0.175
WOA-XGB2	0.236	0.849	0.179	0.214	0.162
GOA-XGB2	0.226	0.855	0.172	0.123	−0.001

**Figure 5.** Relationship between features and AGB based on the GOA-XGB1 model.

By comparing Tables 4 and 5, 75% of the models achieve higher accuracy after using fewer input factors. Compared with the GOA-XGB1 model, the RMSE and MAE of the GOA-XGB2 model decreased by 2.6% and 3.4%, respectively. Compared with the GOA-SVM1 model, the RMSE and MAE of the GOA-SVM2 model decreased by 4.7% and 9.2%,

respectively. The XGB, RF, and GPR models had similar results. However, the accuracy of the GBDT and MLP models did not change or even decrease after using fewer input factors.

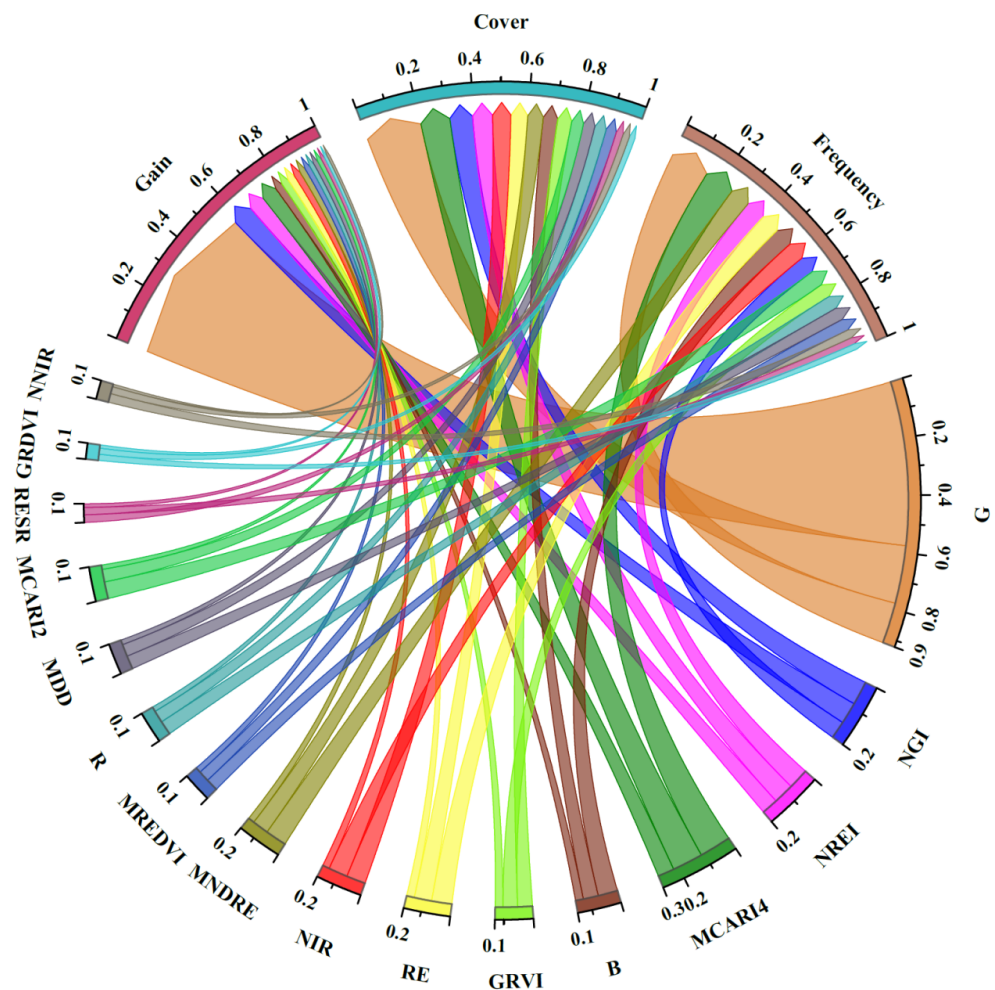


Figure 6. Relationship between features and AGB based on the GOA-XGB2 model.

3.5. Mapping AGB at Field Scale

The knowledge of wheat biomass at the field scale is helpful to evaluate crop growth potential, rationally plan the management of water and fertilizer, and achieve accurate management of farmland. In this study, the model with the highest accuracy, the GOA-XGB2 model, was used to simulate various dynamics of wheat AGB. The results are shown in Figure 7. The mean value of the plot was $1.49 \pm 0.34 \text{ kg m}^{-2}$. AGB was the smallest in the central northern part of the plot, with an AGB value of 1 kg m^{-2} , whereas the biomass was larger in the southern part, with a value of more than 1.6 kg m^{-2} , mainly because there was only a small amount of irrigation in 2021. The entire plot had a certain slope, and the surface runoff mainly gathered in the southern region.

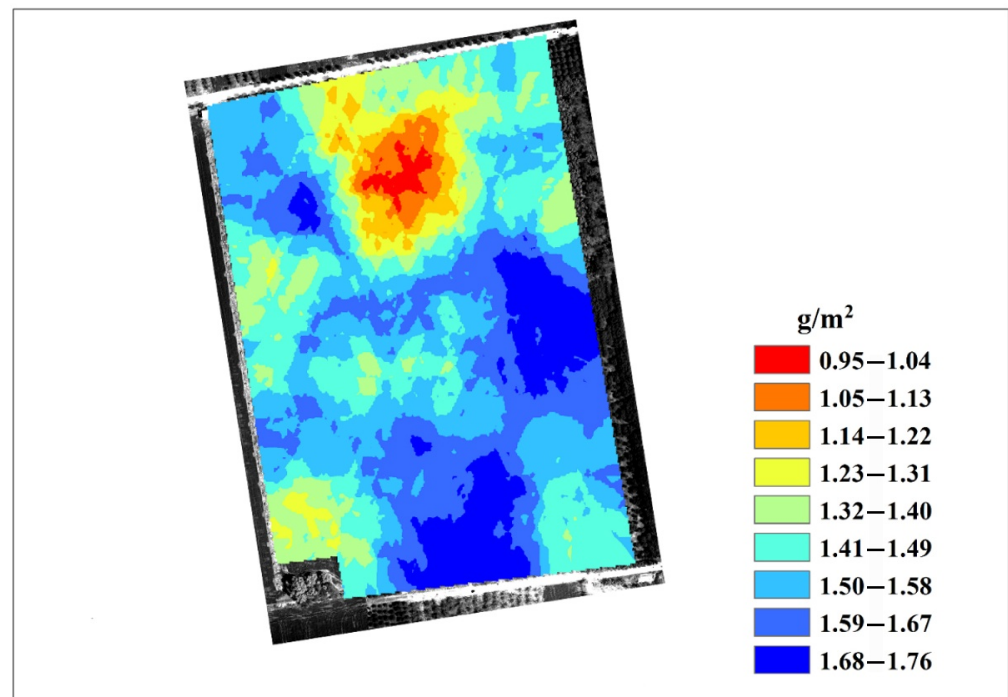


Figure 7. Mapping of AGB in the whole field.

4. Discussion

4.1. Uncertainty of Observed Data

Using UAV to quickly and nondestructively collect spectral, texture, point cloud, and digital elevation information for wheat prediction is proven to be an effective method. The applicability of this method to different crops varies to some extent. For example, sparse vegetation is suitable for point cloud features, whereas tall vegetation can be aided by digital elevation. The uncertainty of using UAV to simulate wheat AGB mainly comes from two aspects: crop and image. First, when calculating wheat AGB at the field scale, individual growth differences between wheat plants within a given plot are not considered; it is assumed that wheat is generally at a similar growth level within a given plot. For destructive sampling, simply multiplying the average biomass by the number of plants may result in systematic errors and outliers in data. In addition, there are multisource errors in UAV remote sensing images, which affect the accuracy of wheat AGB estimation. Because the growth of wheat can cause changes in canopy structure, leaf development, and senescence, differences in radiation intensity on different dates and the uniformity of UAV shooting angle induce considerable uncertainties to data consistency. As VIs are susceptible to the mixed influence of canopy greenness and soil reflectance [49], the accumulation of wheat AGB is directly related to the change in wheat's physical structure. Using hyperspectral cameras to obtain rich spectral features of narrow bands to estimate AGB can reduce collinearity and redundancy of spectral predictors due to similar calculation formulas [7,50]. However, the high-cost requirements of the program pose great difficulties for popularization. Moreover, when four or five narrowband multispectral sensors are used to calculate VIs, a large number of VIs highly correlating with each other will be generated due to the close calculation formula and partial spectral information. We found similar problems, but we attempted to use a coupling algorithm to reduce the collinearity effect and screen out an ideal combination of features that can reduce the redundancy and over-fitting problems of VIs. The results showed that the input features reduced from 5 bands plus 56 VIs (61 features) to 16 features, and the accuracy of the model further improved. Finally, farmlands are not completely controlled environments as laboratory experiments. Farmland can be affected by weeds, plant health, or seeding losses that involve canopy

changes in crop management; however, we did not consider this in this study, which is a limitation.

4.2. Comparison of Different Models

Certain studies report that the MLR model's obvious advantage is that it is highly explicable [51]. Therefore, the MLR model can use the standardized partial regression coefficient to determine the impact strength of one or more predictive variables on response variables. However, we found that when the MLR model had more input features, its accuracy improved significantly, indicating that MLR is only interpretable in finite dimensions and is not strong for complex problems.

In this study, the performance of ML models is significantly better than that of the MLR model because the nonlinear regression model can judge the nonlinear relationship in data. We explored twelve ML regression algorithms, all of which yielded acceptable accuracies. Tree-based algorithms achieved high precision. This is similar to the results by Han et al. [1] who compared four ML algorithms and recommended the RF model. Notably, there are limitations in model comparisons. Owing to the small sample size, the advantages and disadvantages of using different modeling strategies have not been fully demonstrated. For example, the ANN model has a high requirement for the number of samples. When the number of neurons in the hidden layer is large, small sample data will make the model overfit. SVM also lacks a clear internal working mechanism, especially when it contains enormous noise, its parameters become very sensitive, resulting in great uncertainties. The GPR model has a good effect when the data conform to the Gaussian distribution, but when the data distribution is irregular, the model produces considerable uncertainties. The tree-based models (GBDT, RF, and XGBoost) were more dependent on feature selection and could sort the contribution of features and eliminate features with low contribution, which has obvious advantages when the data dimensions are high and there is correlation between them. Particularly, after the optimization algorithm is used, the model has higher precision, and its structure becomes more simplified.

4.3. Determining the Most Important VIs

Han et al. [1] found that BIOVP was a volume indicator to estimate corn AGB, and this indicator was highly correlated with some spectral indicators, such as NGRDI and VARI. Because BIOVP contains both spectral and plant height information, they affect the BIOVP calculation accuracy. Jannoura et al. [52] obtained a significant relationship between NGRDI of peas and oats and AGB. However, other authors have described saturation of maximum leaf-area index values for corn, soybean, and alfalfa [53]. There are also studies that used point cloud computing based on digital images. Compared with other commonly used indicators, the CSM has certain applicability because it can avoid the influence of saturation, i.e., the saturation of NDVI in a later period. In this study, owing to the low plant height and high density of wheat, we found that the applicability of this index was low after trying. The top five contribution degrees of the optimal model parameter combination determined in this study were B, G, NGI, NREI, and MCARI4. Notably, except for MCARI4, all other parameters reached their maximum value at the end of growth, which differed from NDVI and other indices, which decreased with a decrease in chlorophyll content of leaf species. In addition, some authors [54,55] reported that it was difficult to obtain good growth monitoring results at early stages due to low biomass rates, mainly because the differences in VI characteristics from mid-late growth were relatively pronounced. We obtained similar results. However, this problem can be alleviated by choosing a better model, such as the GOA-XGB2 model.

4.4. Effect of Growth Stage on VIs-AGB Model

The most common method for developing AGB models is to create multiple models for different crop growth stages [56].

This is due to the fact that wheat exhibits different spectroscopic characteristics at different growth stages, and it is extremely difficult to establish an AGB prediction model for the entire growth stage using just one or few vegetation indices. For example, while NDVI is thought to have a good relationship with AGB, it also suffers from the problem of light saturation, which means that NDVI no longer increases during the flowering and grain-filling stages of wheat while biomass continues to increase. Li et al. [13] achieved good results by incorporating accumulated temperature, days after sowing, and other data to assist the vegetation index in establishing the AGB model of the entire growth stage. However, one potential limitation of this method is that the growth accumulation temperatures of different wheat varieties vary, and a lack of water and fertilizer can also affect wheat growth, so information such as GDD may also be a model uncertainty factor. This study proposes a general AGB prediction model (GOA-XG2) for wheat at all growth stages that is not dependent on information, such as GDD. It mines various vegetation index combinations using machine learning to simulate the characteristics of wheat at various stages, and it achieves extremely high accuracy. The new GOA-XGB2 method has screened out 16 combinations of vegetation indices and bands that could reflect the growing trend of AGB during the growth period, which is the main reason for its feasibility. Geng et al. [9] used the XGBoost model to simulate AGB and reduced the number of features from 27 to 9.

5. Conclusions

Rapid and nondestructive determination of wheat AGB is crucial for accurate and efficient agricultural management. In this study, we established a new model, named XGBoost optimization by GOA (GOA-XGB), which can accurately determine an ideal VI combination for inversion of wheat AGB. GOA-XGB was compared with many state-of-the-art models. The results showed that SMLR and GOA-MLR models had poor prediction accuracy for AGB, and the accuracy did not improve significantly when the input factors were more than three. Among single-factor-driven ML models, the GPR model had the highest accuracy, followed by the XGBoost model. When the input combination of multispectral bands and VIs were used, the GOA-XGB model with 37 input factors had the highest accuracy, with $RMSE = 0.232 \text{ kg m}^{-2}$, $R^2 = 0.847$, $MAE = 0.178 \text{ kg m}^{-2}$, and $NRMSE = 0.127$. When the XGBoost feature selection was used to reduce the input factors to 16, the model accuracy further improved to $RMSE = 0.226 \text{ kg m}^{-2}$, $R^2 = 0.855$, $MAE = 0.172 \text{ kg m}^{-2}$, and $NRMSE = 0.123$. Based on the developed model, the average AGB of the plot was $1.49 \pm 0.34 \text{ kg}$. The above results show that the newly established model has high accuracy and is helpful in accurately estimating the AGB of wheat at any growth stage at the farmland scale.

The results of this study confirm that the GOA-XGB model (driven by the combination of vegetation indices) can accurately predict the AGB in the whole growth stage of wheat. The limitation of this study is that it was only evaluated in one wheat field. In the future, we hope to apply the model to satellite remote sensing platforms to conduct large-scale wheat AGB and related estimates. In addition, we wish to evaluate the GOA-XGB model and related models for their ability to predict leaf nitrogen concentration and yield, and estimate nitrogen dynamics in crop ecosystems.

Author Contributions: Methodology, Z.L.; software, writing—original draft preparation, Y.H.; formal analysis, validation, R.T.; writing—review and editing, funding acquisition, B.Z.; visualization, supervision, J.F. All authors have read and agreed to the published version of the manuscript.

Funding: The work reported in this manuscript was funded by the National Natural Science Foundation of China (31772389); the National Key Research and Development Program of China (2018YFD0200403); the National Science and Technology Support Program (2015BAD23B04); the Special Fund for Agricultural Research in Public Welfare Industry (201503124); and the National Wheat Modern Industrial Technology System Construction Special Project (Z225020803).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Han, L.; Yang, G.; Dai, H.; Xu, B.; Yang, H.; Feng, H.; Li, Z.; Yang, X. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods* **2019**, *15*, 10. [[CrossRef](#)] [[PubMed](#)]
- Kumar, L.; Mutanga, O. Remote sensing of above-ground biomass. *Remote Sens.* **2017**, *9*, 935. [[CrossRef](#)]
- Zhou, Y.; Xiao, X.; Qin, Y.; Dong, J.; Zhang, G.; Kou, W.; Li, X. Mapping paddy rice planting area in rice-wetland coexistent areas through analysis of Landsat 8 OLI and MODIS images. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *46*, 1–12. [[CrossRef](#)]
- Jin, X.; Li, Z.; Feng, H.; Ren, Z.; Li, S. Deep neural network algorithm for estimating maize biomass based on simulated Sentinel 2A vegetation indices and leaf area index. *Crop J.* **2020**, *8*, 87–97. [[CrossRef](#)]
- Dong, T.; Liu, J.; Qian, B.; He, L.; Liu, J.; Wang, R.; Shang, J. Estimating crop biomass using leaf area index derived from Landsat 8 and Sentinel-2 data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 236–250. [[CrossRef](#)]
- Zhang, L.X.; Chen, Y.Q.; Li, Y.X.; Du, K.M.; Zheng, F.X.; Sun, Z.F. Estimating Above Ground Biomass of Winter Wheat at Early Growth Stages Based on Visual Spectral. *Spectrosc. Spectr. Anal.* **2019**, *39*, 2501.
- Yue, J.; Yang, G.; Tian, Q.; Feng, H.; Xu, K.; Zhou, C. Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 226–244. [[CrossRef](#)]
- Jia, M.; Li, W.; Wang, K.; Zhou, C.; Cheng, T.; Tian, Y.; Zhu, Y.; Cao, W.; Yao, X. A newly developed method to extract the optimal hyperspectral feature for monitoring leaf biomass in wheat. *Comput. Electron. Agric.* **2019**, *165*, 104942. [[CrossRef](#)]
- Geng, L.; Che, T.; Ma, M.; Tan, J.; Wang, H. Corn Biomass Estimation by Integrating Remote Sensing and Long-Term Observation Data Based on Machine Learning Techniques. *Remote Sens.* **2021**, *13*, 2352. [[CrossRef](#)]
- Liu, S.; Zeng, W.; Wu, L.; Lei, G.; Chen, H.; Gaiser, T.; Srivastava, A.K. Simulating the Leaf Area Index of Rice from Multispectral Images. *Remote Sens.* **2021**, *13*, 3663. [[CrossRef](#)]
- Wang, M.; Wan, Y.; Ye, Z.; Lai, X. Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm. *Inf. Sci.* **2017**, *402*, 50–68. [[CrossRef](#)]
- Dong, J.; Liu, X.; Huang, G.; Fan, J.; Wu, L.; Wu, J. Comparison of four bio-inspired algorithms to optimize KNEA for predicting monthly reference evapotranspiration in different climate zones of China. *Comput. Electron. Agric.* **2021**, *186*, 106211. [[CrossRef](#)]
- Li, Z.; Zhao, Y.; Taylor, J.; Gaulton, R.; Jin, X.; Song, X.; Yang, G. Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data. *Remote Sens. Environ.* **2022**, *273*, 112967. [[CrossRef](#)]
- Zadoks, J.C.; Chang, T.T.; Konzak, C.F. A decimal code for the growth stages of cereals. *Weed Res.* **1974**, *14*, 415–421. [[CrossRef](#)]
- Buschmann, C.; Nagel, E. In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *Int. J. Remote Sens.* **1993**, *14*, 711. [[CrossRef](#)]
- Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127. [[CrossRef](#)]
- Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
- Cao, Q.; Miao, Y.; Wang, H.; Huang, S.; Cheng, S.; Khosla, R.; Jiang, R. Non-destructive estimation of rice plant nitrogen status with Crop Circle multispectral active canopy sensor. *Field Crops Res.* **2013**, *154*, 133. [[CrossRef](#)]
- Gitelson, A.A. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* **2005**, *32*, L08403. [[CrossRef](#)]
- Sripada, R.P.; Heiniger, R.W.; White, J.G.; Meijer, A.D. Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agron. J.* **2006**, *98*, 968. [[CrossRef](#)]
- Lu, J.; Miao, Y.; Shi, W.; Li, J.; Yuan, F. Evaluating different approaches to non-destructive nitrogen status diagnosis of rice using portable RapidSCAN active canopy sensor. *Sci. Rep.* **2017**, *7*, 14073. [[CrossRef](#)] [[PubMed](#)]
- Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352. [[CrossRef](#)]
- Rouse, J.W., Jr.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In *NASA. Goddard Space Flight Center 3d ERTS-1 Symphony*; NASA: Washington, DC, USA, 1974; p. 309.
- Jordan, C.F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology* **1969**, *50*, 663–666. [[CrossRef](#)]
- Roujean, J.; Breon, F. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* **1995**, *51*, 375–384. [[CrossRef](#)]
- Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295. [[CrossRef](#)]
- Sandham, L. Surface temperature measurement from space: A case study in the south western cape of south Africa. *S. Afr. J. Enol. Vitic.* **1997**, *18*, 25. [[CrossRef](#)]
- Chen, J.M. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* **1996**, *22*, 229. [[CrossRef](#)]
- Reyniers, M.; Walvoort, D.J.; De Baardemaaker, J. A linear model to predict with a multi-spectral radiometer the amount of nitrogen in winter wheat. *Int. J. Remote Sens.* **2006**, *27*, 4159–4179. [[CrossRef](#)]
- Dash, J.; Curran, P. The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* **2004**, *25*, 5403–5413. [[CrossRef](#)]
- Goel, N.S.; Qin, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation. *Remote Sens. Environ.* **1994**, *10*, 309–347. [[CrossRef](#)]

32. Gong, P.; Pu, R.; Biging, G.S.; Larrieu, M.R. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Geosci. Remote Sens. Lett.* **2003**, *41*, 1355–1362. [\[CrossRef\]](#)
33. Barnes, E.; Clarke, T.; Richards, S.; Colaizzi, P.; Haberland, J.; Kostrzewski, M.; Waller, P.; Choi, C.; Riley, E.; Thompson, T.; et al. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA, 16–19 July 2000.
34. Jasper, J.; Reusch, S.; Link, A. Active sensing of the N status of wheat using optimized wavelength combination: Impact of seed rate, variety and growth stage. *Precis. Agric.* **2009**, *9*, 23–30.
35. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [\[CrossRef\]](#)
36. Elsayed, S.; Rischbeck, P.; Schmidhalter, U. Comparing the performance of active and passive reflectance sensors to assess the normalized relative canopy temperature and grain yield of drought-stressed barley cultivars. *Field Crops Res.* **2015**, *177*, 148–160. [\[CrossRef\]](#)
37. Erdle, K.; Misteale, B.; Schmidhalter, U. Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in wheat cultivars. *Field Crops Res.* **2011**, *124*, 74. [\[CrossRef\]](#)
38. Datt, B. Visible/near infrared reflectance and chlorophyll content in Eucalyptus leaves. *Int. J. Remote Sens.* **1999**, *20*, 2741–2759. [\[CrossRef\]](#)
39. Dai, H.; MacBeth, C. Effects of learning parameters on learning procedure and performance of a BPNN. *Neural Netw.* **1997**, *10*, 1505–1521. [\[CrossRef\]](#)
40. Quinonero-Candela, J.; Rasmussen, C.E. A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.* **2005**, *6*, 1939–1959.
41. Vapnik, V.; Izmailov, R. Knowledge transfer in SVM and neural networks. *Ann. Math. Artif. Intell.* **2017**, *81*, 3–19. [\[CrossRef\]](#)
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
43. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [\[CrossRef\]](#)
44. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
45. Mafarja, M.; Aljarah, I.; Heidari, A.A.; Hammouri, A.I.; Faris, H.; Ala'M, A.Z.; Mirjalili, S. Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems. *Knowl.-Based Syst.* **2018**, *145*, 25–45. [\[CrossRef\]](#)
46. Yue, J.; Feng, H.; Jin, X.; Yuan, H.; Li, Z.; Zhou, C.; Yang, G.; Tian, Q. A comparison of crop parameters estimation using images from UAV-mounted snapshot hyperspectral sensor and high-definition digital camera. *Remote Sens.* **2018**, *10*, 1138. [\[CrossRef\]](#)
47. Li, W.; Niu, Z.; Chen, H.; Li, D.; Wu, M.; Zhao, W. Remote estimation of canopy height and aboveground biomass of maize using high-resolution stereo images from a low-cost unmanned aerial vehicle system. *Ecol. Indic.* **2016**, *67*, 637–648. [\[CrossRef\]](#)
48. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [\[CrossRef\]](#)
49. Jimenezberni, J.A.; Deery, D.M.; Rozaslarraondo, P.; Condon, A.G.; Rebetzke, G.J.; James, R.A.; Bovill, W.D.; Furbank, R.T.; Sirault, X.R.R. High throughput determination of plant height, ground cover, and above-ground biomass in wheat with LiDAR. *Front. Plant Sci.* **2018**, *9*, 237. [\[CrossRef\]](#)
50. Chakraborty, A.; Goswami, D. Prediction of slope stability using multiple linear regression (MLR) and artificial neural network (ANN). *Arab. J. Geosci.* **2017**, *10*, 385. [\[CrossRef\]](#)
51. Xu, W.; Chen, P.; Zhan, Y.; Chen, S.; Zhang, L.; Lan, Y. Cotton yield estimation model based on machine learning using time series UAV remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102511. [\[CrossRef\]](#)
52. Jannoura, R.; Joergensen, R.G.; Bruns, C. Organic fertilizer effects on growth, crop yield, and soil microbial biomass indices in sole and intercropped peas and oats under organic farming conditions. *Eur. J. Agron.* **2014**, *52*, 259–270. [\[CrossRef\]](#)
53. Hunt, E.R.; Cavigelli, M.; Daughtry, C.S.; McMurtrey, J.E.; Walthall, C.L. Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precis. Agric.* **2005**, *6*, 359–378. [\[CrossRef\]](#)
54. Bosch Serra, A.D.; Casanova, D. Estimation of onion (*Allium cepa*, L.) biomass and light interception from reflectance measurements at field level. In Proceedings of the XXV International Horticultural Congress, Part 9: Computers and Automation, Electronic Information in Horticulture, Brussels, Belgium, 2–7 August 1998; pp. 53–64.
55. Alvino, A.; Marino, S. Remote sensing for irrigation of horticultural crops. *Horticulturae* **2017**, *3*, 40. [\[CrossRef\]](#)
56. Marshall, M.; Belgiu, M.; Boschetti, M.; Pepe, M.; Stein, A.; Nelson, A. Field-level crop yield estimation with PRISMA and Sentinel-2. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 191–210. [\[CrossRef\]](#)