



# Combining 2D image and point cloud deep learning to predict wheat above ground biomass

Shaolong Zhu<sup>1,2</sup> · Weijun Zhang<sup>1,2</sup> · Tianle Yang<sup>1,2</sup> · Fei Wu<sup>3</sup> · Yihan Jiang<sup>1,2</sup> · Guanshuo Yang<sup>1,2</sup> · Muhammad Zain<sup>1,2</sup> · Yuanyuan Zhao<sup>1,2</sup> · Zhaosheng Yao<sup>1,2</sup> · Tao Liu<sup>1,2</sup> · Chengming Sun<sup>1,2</sup>

Accepted: 23 August 2024 / Published online: 9 September 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

**Purpose** The use of Unmanned aerial vehicle (UAV) data for predicting crop above-ground biomass (AGB) is becoming a more feasible alternative to destructive methods. However, canopy height, vegetation index (VI), and other traditional features can become saturated during the mid to late stages of crop growth, significantly impacting the accuracy of AGB prediction.

**Methods** In 2022 and 2023, UAV multispectral, RGB, and light detection and ranging point cloud data of wheat populations were collected at seven growth stages across two experimental fields. The point cloud depth features were extracted using the improved PointNet++ network, and AGB was predicted by fusion with VI, color index (CI), and texture index (TI) raster image features.

**Results** The findings indicate that when the point cloud depth features were fused, the  $R^2$  values predicted from VI, CI, TI, and canopy height model images increased by 0.05, 0.08, 0.06, and 0.07, respectively. For the combination of VI, CI, and TI,  $R^2$  increased from 0.86 to a maximum of 0.9, while the root-mean-square error (RMSE) and mean absolute error were  $1.80 \text{ t ha}^{-1}$  and  $1.36 \text{ t ha}^{-1}$ , respectively. Additionally, our findings revealed that the hybrid fusion exhibits the highest accuracy, it demonstrates robust adaptability in predicting AGB across various years, growth stages, crop varieties, nitrogen fertilizer applications, and densities.

**Conclusion** This study effectively addresses the saturation in spectral and chemical information, provides valuable insights for high-precision phenotyping and advanced crop field management, and serves as a reference for studying other crops and phenotypic parameters.

**Keywords** Wheat · Biomass prediction · Unmanned aerial vehicle · Point cloud deep learning · Multimodal data fusion

Extended author information available on the last page of the article

## Introduction

Above-ground biomass (AGB) serves as a crucial indicator, reflecting crop growth conditions and the material basis for crop yield and quality (Liu et al., 2019). The rapid and accurate prediction of AGB holds significant importance in optimizing field management strategies and evaluating crop yield status (Lu et al., 2019). The traditional AGB measurement method entails harvesting, drying, and weighing the samples, while this method is destructive, labor-intensive, time-consuming, and unsuitable for large-scale crops (Yue et al., 2017). Therefore, non-destructive AGB measurement strategies are emerging as ideal alternatives, especially in era of high-throughput multi-omics (Han et al., 2019).

Unmanned aerial vehicle (UAV) platforms offers an ideal method for the rapid and non-destructive acquisition of crop growth information (Li et al., 2020a, 2020b). This platform can be outfitted with various sensors, such as RGB, multispectral (MS), hyperspectral (HS), thermal infrared (TIR), and light detection and ranging (LIDAR). RGB cameras, when combined with high-precision real-time kinematic (RTK) modules, can rapidly acquire farmland color, texture, structure, and elevation information, and predict canopy height (CH), coverage, leaf area index (LAI), AGB, and other parameters (Maimaitijiang et al., 2017). MS and HS cameras are capable of capturing the spectral absorption and reflection characteristics of crops, and are frequently utilized to predict leaf water content (LWC), leaf nutrient content (LNC), leaf chlorophyll content (LCC), and yield (Lu et al., 2021). TIR cameras have the ability to convert the infrared radiation energy received from crops into temperature. Given that canopy temperature is closely related to stomatal conductance, photosynthetic characteristics, and evaporation rate, TIR imaging technology can effectively be employed to detect the response of crops to stress environments (Kerkech et al., 2020). LIDAR as an active sensor, is the primary approach for obtaining CH and 3D information by emitting a laser beam to measure the distance to a target (Zhu et al., 2021a, 2021b).

Different sensors have limitations in their responses and the information they capture. Utilizing multiple sensors to acquire and integrate data comprehensively is essential to avoid issues arising from insufficient feature availability (Cen et al., 2019; Orusa et al., 2023). Previous study has simultaneously acquired RGB and MS images of rice, extracted CH and vegetation index (VI), and integrated these feature types for AGB prediction. The prediction accuracy proved to be significantly superior to that of models based on a single sensor (Wan et al., 2020). However, traditional features are derived from manually designed extraction rules, including NDVI, SAVI. Subsequently, feature screening or mapping methods are employed for dimensionality reduction, culminating in the construction of a prediction model (Li et al., 2020a, 2020b). Manually extracted features often remain at a basic level, only reflecting certain aspects of the original data and failing to capture its intrinsic nature fully and effectively (Quan et al., 2021).

The deep learning (DL) principle is to learn and extract features in data through multi-layer neural networks (Orusa et al., 2024). DL is not only capable of uncovering complex internal structures hidden in data, but also efficiently extracts pertinent features from the original data for specific tasks (Gill et al., 2022). Recent research has delved into the potential of canopy spectrum, structure, temperature, and texture information, acquired through UAV multi-sensors, for predicting soybean yield. The results indicate that the convolution of each feature vector followed by their fusion enhances prediction accuracy and robustness as compared to direct concatenate feature vectors (Maimaitijiang et al., 2020). However, this study initially involved manual extraction of features (e.g., VIs, texture indices (TIs)

from UAV images, before inputting them into a CNN for convolution and training, not fully containing the original or feature images information (Yu et al., 2023). Further, DL is frequently applied in the feature extraction of one-dimensional (1D) vectors and three-channel RGB images, but its usage in MS and HS original images and 3D point cloud data is very scarce (Ghahremani et al., 2021). Additionally, both MS and RGB images constitute structured data, with each pixel point being ordered, whereas point clouds represent unstructured data, characterized by their unordered and irregular format. The differing modalities of point clouds and raster images present challenges in data fusion. Currently, the combination of 2D images and 3D point clouds has been employed in the structure estimation of maize (Liu et al., 2021), vineyards (Comba et al., 2020), soybean (Maimaitijiang et al., 2019), potato (Wang et al., 2023), and certain vegetable crops (Astor et al., 2020). In these studies, point cloud data is primarily used to extract volume parameters or convert elevation information into raster data, with the extracted features often including maximum, minimum, and average values, as well as percentiles and coefficients of variation of elevation (Tian et al., 2022). However, a systematic investigation into the deep features of original point clouds remains significantly lacking.

This study constructs a diverse wheat dataset comprising of various cultivars, nitrogen fertilizer levels, and planting densities. Canopy data was obtained from UAV RGB, MS, and LIDAR sensors during the principal growth stages. Finally, a convolutional neural network was constructed that integrates RGB and MS feature images with 3D point clouds for rapid, end-to-end prediction of wheat AGB. The main objective of this research was to introduce novel methodologies for the comprehensive utilization of UAV images, particularly the depth features of LIDAR point cloud data, to enhance AGB prediction accuracy.

## Materials and methods

### Study area and experimental setup

The experiment involving wheat cultivation spanned two years and was conducted in two distinct areas. Field experiment 1 was carried out in 2022 in Yangzhou, China ( $32^{\circ}27'N$ ,  $119^{\circ}40'E$ ). This region is characterized by a transition area between a subtropical monsoon humid climate and a temperate monsoon climate. In 2022, this region recorded a rainfall of 653.3 mm and an annual sunshine duration of 2343 h. Annual production follows a double cropping cycle with winter wheat and rice grown in rotation in each year. Field experiment 2 was carried out in 2023 in Zhenjiang, China ( $32^{\circ}16'N$ ,  $119^{\circ}33'E$ ). This region has a northern subtropical monsoon climate. In 2023, this region recorded a rainfall of 780.6 mm and an annual sunshine duration of 1724 h. Annual production follows a double cropping cycle with the previous crop was maize.

To generate diverse wheat populations dataset with varying biomass, this study utilized a blend of various cultivars, nitrogen fertilizer levels, and density treatments. Experiment 1 comprised of five planting densities:  $135 \times 10^4$  plants  $hm^{-2}$  (D1),  $180 \times 10^4$  plants  $hm^{-2}$  (D2),  $225 \times 10^4$  plants  $hm^{-2}$  (D3),  $270 \times 10^4$  plants  $hm^{-2}$  (D4), and  $315 \times 10^4$  plants  $hm^{-2}$  (D5), and four nitrogen fertilizer treatments as  $0 kg\ hm^{-2}$  (N1),  $135 kg\ hm^{-2}$  (N2),  $225 kg\ hm^{-2}$  (N3), and  $300 kg\ hm^{-2}$  (N4). While, experiment 2 contains three planting densities: D1, D3, and D5, and the nitrogen fertilizer treatments were same as in experiment 1. Experiment 1 and 2 included 60 and 36 treatments, respectively, each with three replicates. All sampling and UAV data collection were conducted in the first replicate, while the other

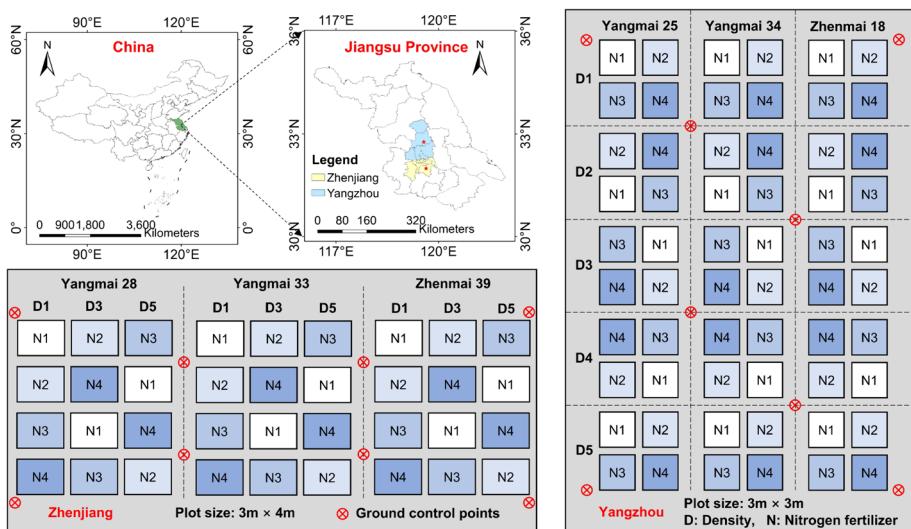
two were used for yield measurement. The seeding date for experiment 1 was November 10, 2022, while for experiment 2, it was November 20, 2023. The detailed experiment design is depicted in Fig. 1. The dataset from experiment 1 was primarily used for model construction and validation, whereas experiment 2 dataset was only employed to assess the model adaptability.

## Data acquisition

Field campaigns were undertaken to acquire UAV images, CH, and AGB data at important growth stages (Table 1).

### UAV image acquisition

The MS and RGB data were collected using a DJI Mavic 3 multispectral (Da-Jiang Innovations Science and Technology Co., Ltd., Shenzhen, China), which integrates one RGB camera and four MS cameras: green:  $560 \pm 16$  nm, red:  $650 \pm 16$  nm, red edge:  $730 \pm 16$  nm, near-infrared:  $840 \pm 26$  nm. Flight plans and parameters were established using DJI Pilot 2 software, setting the UAV altitude at 12 m and the lens angle set at  $90^\circ$  straight down. Forward and side overlaps were maintained at 80% and 70%, respectively. Prior to the flight mission, three reflectance boards with predetermined reflectance values (25%, 50%, and 75%) were manually captured for ensuing radiometric correction. LiDAR data was acquired using a DJI Matrice 300 RTK containing Zenmuse L1 sensor. The flight mode involves oblique photography, with the lens angle set at  $45^\circ$ , and each flight is required to fly five routes around the experiment field. The flying altitude was set to 15 m, while the forward and side overlaps were the same as in Mavic 3 Multispectral. Both UAVs operated concurrently from 11:00 a.m. to 1:00 p.m., under clear-sky and low-wind conditions.



**Fig. 1** Illustration of experimental location and design of study area

**Table 1** Summary of field campaigns for the wheat experiments

Experiment	Date of UAV flights and field sampling	Growth stage
Exp. 1	8 March, 2022	Returning (GS 30)
	15 March, 2022	Jointing (GS 32)
	27 March, 2022	Booting (GS 45)
	9 April, 2022	Heading (GS 55)
	20 April, 2022	Flowering (GS 65)
	2 May, 2022	Milking (GS 75)
	15 May, 2022	Dough (GS 85)
Exp. 2	12 March, 2023	Returning (GS 30)
	3 April, 2023	Booting (GS 48)
	26 April, 2023	Flowering (GS 65)
	20 May, 2023	Dough (GS 85)

The Zadoks growth scale (GS) is an established system for describing the stages of growth and development of cereals, particularly wheat and barley

## CH and AGB measurement

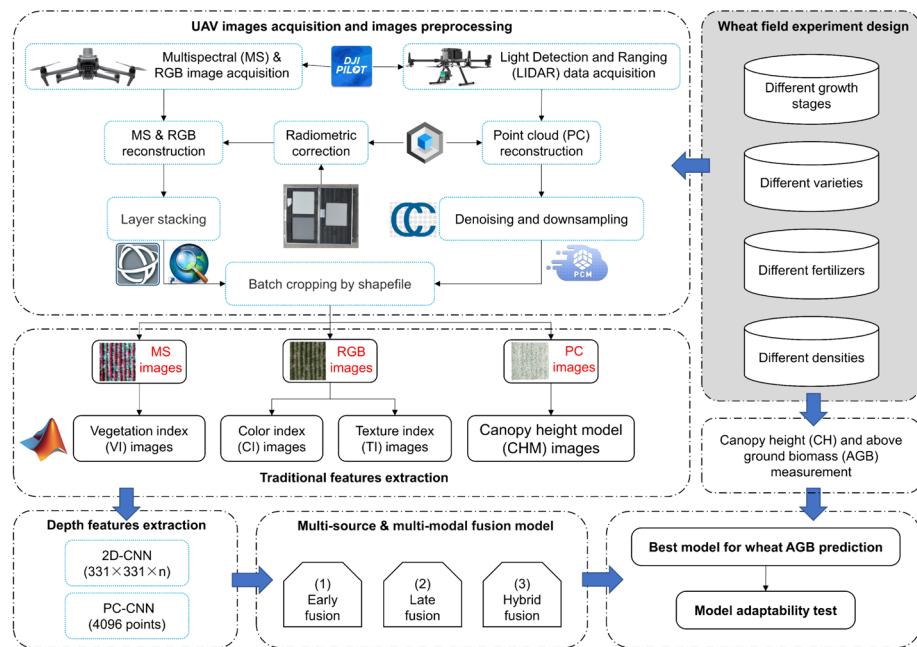
The CH measurements and sampling were conducted concurrently with the acquisition of UAV images. A ruler was used to measure the distance from ground to highest point of the leaf or ear in its natural position, excluding the awns. Within each plot, ten plants were randomly selected for measurement, and the mean CH was thereafter calculated.

A frame of  $0.3\text{ m} \times 0.3\text{ m}$  dimensions was placed in every plot, and all the fresh wheat plant samples within this frame were collected and transferred to laboratory. The samples were subjected to water cleansing, and their roots were excised. Subsequently, the residual samples were dried in an oven at  $105\text{ }^{\circ}\text{C}$  for 30 min, followed by further drying at  $80\text{ }^{\circ}\text{C}$  until a constant weight. The weight of each sample was recorded using electronic balance to quantify the AGB in standard units (tons per hectare).

## Image preprocessing

The MS and point cloud reconstructions were performed using DJI Terra software. Prior to MS reconstruction, calibrated board data was imported for radiometric correction to enhance the output reliability. After reconstruction, four single-band maps and an RGB images were generated. The “build layer stack tool” in ENVI was utilized to create the MS image from the four single-band maps. The point cloud density was configured to its maximum to ensure high-quality outcomes in point cloud reconstruction, and point cloud accuracy optimization was activated. CloudCompare was employed for point cloud denoising and downsampling, followed by the construction of a digital surface model (DSM) using the elevation data within the point cloud. The overall workflow is illustrated in Fig. 2.

Raster data alignment was performed using ground control points to rectify positional discrepancies in orthophotos, obtained from varied sensors and at different times. Batch clipping of the MS, RGB, and point cloud images was conducted based on the shapefiles corresponding to each plot. The aforementioned shapefiles were



**Fig. 2** A workflow diagram of data acquisition, data processing, feature extraction, and modeling

generated utilizing the “create feature class tool” in ArcGIS. After rotation and clipping, the MS and RGB images exhibited a resolution of  $331 \times 331$  pixels. The number of point clouds in each plot were 4096.

## Features extraction

### Vegetation indices.

The VIs can be constructed through linear or non-linear combinations of two or more characteristic spectral bands, enhancing the interpretation of remote sensing images (Maimaitijiang et al., 2019). Twenty VIs of the MS images were extracted in accordance with prior research (Table 2).

### Color indices

The RGB image comprising of three color bands (red, green, and blue) provides color information within the visible wavelength band (Woebbecke et al., 1994). In this study, fifteen color indices (CIs) were calculated, as presented in Table 3.

**Table 2** VIs used in this study for above ground biomass prediction

Acronym	Vegetation index	Formula	References
Clg	Chlorophyll index-green	$(\text{Nir}/G) - 1$	Ahammed et al. (2011)
GNDVI	Green normalized difference vegetation index	$(\text{Nir} - G)/(\text{Nir} + G)$	Candago et al. (2015)
NDVI	Normalized difference vegetation index	$(\text{Nir} - R)/(\text{Nir} + R)$	Huei et al. (2002)
NDRE	Normalized difference red-edge index	$(\text{Nir} - \text{RedE})/(\text{Nir} + \text{RedE})$	Tian et al. (2011)
RTVI	Red-edge triangulated vegetation index	$100*(\text{Nir} - \text{RedE}) - 10*(\text{Nir} - G)$	Chen et al. (2010)
SAVI	Soil-adjusted vegetation index	$1.5*(\text{Nir} - R)/(\text{Nir} + R + 0.5)$	Haboudane et al. (2004)
SR	Simple ratio index	$\text{Nir}/R$	Gamon et al. (1995)
NRI	Nitrogen reflectance index	$(G - R)/(G + R)$	Bendig et al. (2015)
Datt		$(\text{Nir} - \text{RedE})/(\text{Nir} - R)$	Datt (1999)
DVI	Difference vegetation index	$\text{Nir} - R$	Zhu et al. (2008)
GDVI	Green difference vegetation index	$\text{Nir} - G$	Wu (2014)
GOSAVI	Green optimized soil adjusted vegetation index	$1.16*(\text{Nir} - G)/(\text{Nir} + G + 0.16)$	Cao et al. (2015)
Clre	Chlorophyll index-red-edge	$(\text{Nir}/\text{RedE}) - 1$	Ahammed et al. (2011)
MTCI	MERIS terrestrial chlorophyll index	$(\text{Nir} - \text{RedE})/(\text{RedE} - R)$	Dash and Curran (2004)
OSAVI	Optimized soil-adjusted vegetation index	$1.16*(\text{Nir} - R)/(\text{Nir} + R + 0.16)$	Daughtry et al. (2000)
RV1	Ratio vegetation index	$\text{Nir}/R$	Broge and Mortensen (2002)
CRI	Carotenoid reflectance index	$(1/G) + (1/\text{NIR})$	Stagakis et al. (2010)
MSR	Modified simple ratio index	$((\text{NIR}/R) - 1)/\sqrt{\text{NIR}/R + 1}$	Haboudane et al. (2004)
TVI	Triangular vegetation index	$60*(\text{Nir} - G) - 100*(\text{Nir} + G)$	Broge and Leblanc (2001)
GI	Green index	$G/R$	Zarco-Tejada et al. (2001)

NIR, RedE, R, and G, represent the reflectance of near-infrared, red-edge, red, and green bands, respectively

**Table 3** CIs used in this study for above ground biomass prediction

Acronym	Color index	Formula	References
r	Normalized red	$R/(R + G + B)$	Kawashima and Nakatani (1998)
g	Normalized green	$G/(R + G + B)$	Kawashima and Nakatani (1998)
b	Normalized blue	$B/(R + G + B)$	Kawashima and Nakatani (1998)
GRRI	Green red ratio index	$G/R$	Gamon and Surfus (1999)
GBRI	Green blue ratio index	$G/B$	Maimaitijiang et al. (2019)
RBRI	Red blue ratio index	$R/B$	Maimaitijiang et al. (2019)
INT	Color intensity index	$(R + G + B)/3$	Panneton and Brouillard (2009)
NDI	Normalized difference index	$(r - g)/(r + g + 0.01)$	Woebbecke et al. (1993)
IKAW	Kawashima index	$(R - B)/(R + B)$	Kawashima and Nakatani (1998)
GLA	Green leaf index	$(2*G - R - B)/(2*G + R + B)$	Louhaichi et al. (2001)
ExR	Excess red vegetation index	$1.4*r - g$	Woebbecke et al. (1994)
ExG	Excess green vegetation index	$2*g - r - g$	Woebbecke et al. (1994)
ExB	Excess blue vegetation index	$1.4*b - g$	Guijarro et al. (2011)
ExGR	Excess green minus excess red index	$ExG - ExR$	Guijarro et al. (2011)
CIVE	Color index of vegetation	$0.441*R - 0.881*G + 0.385*B + 18.787$	Guijarro et al. (2011)

R, G, and B represent the DN variables from digital image red, green, and blue bands, respectively

## Texture indices

Texture features can provide canopy structure information, which is an effective supplement to spectral and color information (Hall-Beyer, 2017). In this study, texture features were extracted from RGB images utilizing a gray-level co-occurrence matrix (GLCM) (Haralick et al., 1973). The meanings and formulas of the texture indices are delineated in Table 4.

## Canopy height

CH was determined through the development of a canopy height model (CHM) derived from LiDAR elevation maps. A bare-earth digital elevation model (DEM) was established before wheat emergence. The CHM was computed as follows:

$$\text{CHM} = \text{DSM} - \text{DEM} \quad (1)$$

Subsequently, the CHM maps were clipped using the same shapefile that was used for cropping the MS and RGB images. It was then segmented into 10 smaller, equally sized regions. The mean of the maximum values among these 10 regions denotes the average CH of the respective plot.

## CNN model construction

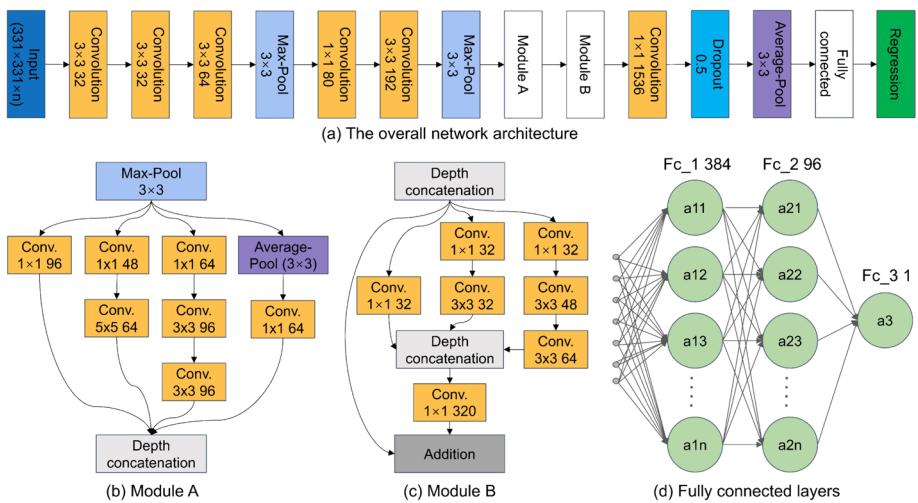
### 2DCNN model construction

The Inception-ResNet-V2 network architecture was utilized as a reference, renowned for its robust performance on the ImageNet dataset (Szegedy et al., 2016). The network was simplified and fine-tuned to adapt to the specific dataset used in this study. The overall architecture of the modified network is given in Fig. 3a. Upon data input, the network underwent

**Table 4** TIs used in this study for above ground biomass prediction

Acronym	Texture Index	Formula	References
Mean	Mean	$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j}), \mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j})$	Hall-Beyer (2017)
Var	Variance	$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i - \mu_i)^2, \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}(j - \mu_j)^2$	Hall-Beyer (2017)
Hom	Homogeneity	$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$	Hall-Beyer (2017)
Con	Contrast	$\sum_{i,j=0}^{N-1} P_{i,j}(i - j)^2$	Hall-Beyer (2017)
This	Dissimilarity	$\sum_{i,j=0}^{N-1} P_{i,j} i - j $	Rodriguez-Galiano et al. (2012)
Ent	Entropy	$\sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j})$	Rodriguez-Galiano et al. (2012)
Sec	Second moment	$\sum_{i,j=0}^{N-1} P_{i,j}^2$	Rodriguez-Galiano et al. (2012)
Cor	Correlation	$\sum_{i,j=0}^{N-1} P_{i,j} \left[ (i - \mu_i)(j - \mu_j) / \sqrt{(\sigma_i^2)(\sigma_j^2)} \right]$	Rodriguez-Galiano et al. (2012)

N is the number of gray levels. i and j are the column and row labels of the GLCM, respectively.  $P_{i,j}$  is the probability that values i and j appear in the adjacent pixels of the original image within the window that defines the neighborhood.  $\mu$  is the mean and  $\sigma$  is the standard deviation, defined by the GLCM mean and the GLCM variance equation in the table

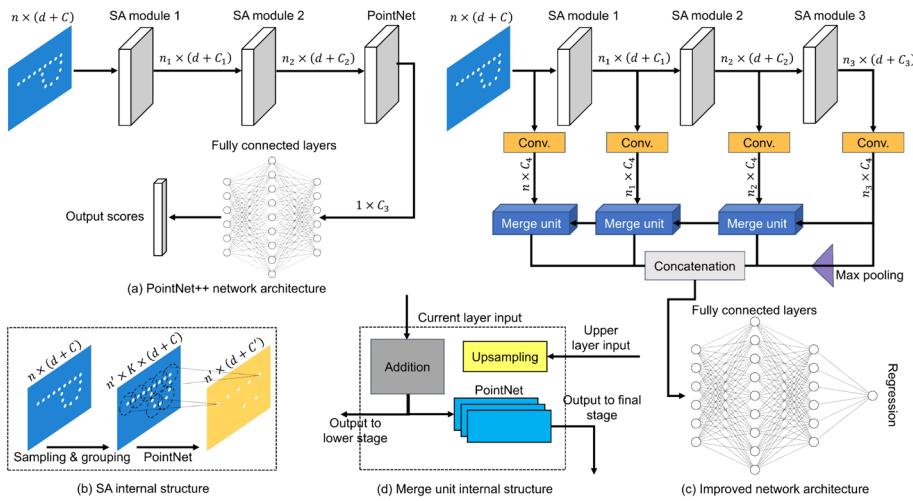


**Fig. 3** 2DCNN architecture. *Conv* convolution layer, *Bc* fully connected layer

a sequence of operations, including three convolutional layers, one max-pooling layer, two additional convolutional layers, another max-pooling layer, module A, module B, one  $1 \times 1$  convolutional layer, one dropout layer, one average-pooling layer, three fully connected layers, and a final regression step. Module A structure comprises of four branches, as shown in Fig. 3b: (1) A  $1 \times 1$  convolutional layer with 96 channels. (2) A combination of a  $1 \times 1$  convolutional layer with 48 channels and a  $5 \times 5$  convolutional layer with 64 channels. (3) A combination of a  $1 \times 1$  convolutional layer with 64 channels and two  $3 \times 3$  convolutional layers with 96 channels. (4) An average-pooling layer and a  $1 \times 1$  convolutional layer with 64 channels. These four branches were subsequently merged using depth concatenation. Module B was similarly segmented into four branches, which essentially is the residual network structure. The detailed structure is depicted in Fig. 3c. The output sizes of the final three fully connected layers in the network are 384, 96, and 1, respectively, as illustrated in Fig. 3d.

## Point Cloud CNN model construction

PointNet is a deep neural network that directly processes unordered raw point cloud data, thereby eliminating the need for voxelization or projection preprocessing of the point cloud. This approach not only enhances the applicability of the network but also maintains superb feature extraction capabilities (Li et al., 2022). PointNet++ is an enhanced version of PointNet (Fig. 4a). A key element of PointNet++ is its hierarchical set abstraction, comprising three essential layers: the sampling layer, the grouping layer, and the PointNet layer (Fig. 4b). PointNet primarily focuses on global features, whereas PointNet++ is designed to optimize the extraction of local features (Qi et al., 2017). This paper introduces an improved point cloud feature extraction network based on PointNet++, primarily through the addition of a network branch that connects with the original network architecture in a layer-by-layer manner, enabling the final output of the network to more effectively integrate features from local to global. The improved overall architecture is shown in Fig. 4c. The Merge Unit module is primarily responsible for merging the features from the upper layer

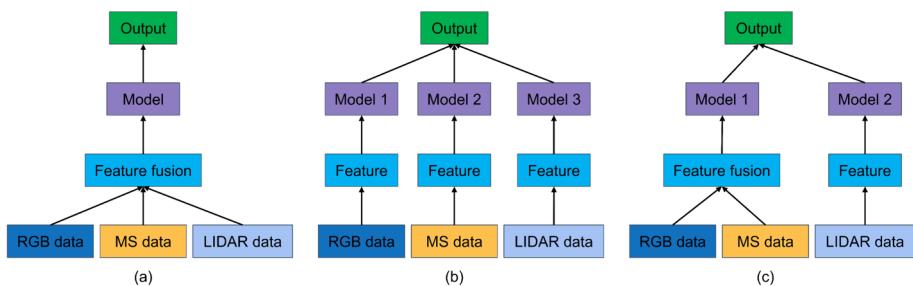


**Fig. 4** Point cloud CNN architecture. Assume the input is an  $n \times (d + C)$  matrix, where  $n$  represents the number of points,  $d$  is the dimension (e.g., if xyz coordinates are used,  $d$  equals 3), and  $C$  is the number of feature dimensions. The corresponding output is an  $n' \times (d + C')$  matrix, where  $n'$  represents the number of sampled points,  $d$  is the dimension, and  $C'$  is the number of new feature dimensions.  $K$  represents the number of neighboring points

with those of the current layer and outputting the processed features to the subsequent unit (Fig. 4d). There are two pathways for the new features: one continues downward, serving as input to the Merge Unit in the lower layer, and the other is directed to the concatenation layer after processing by PointNet. Following these layers, modifications were made to the fully connected layer, transitioning the task from classification or segmentation to regression. This adjustment enabled the application of the network to the specific task of predicting wheat AGB, aligning it closely with the objectives of our study.

## Data fusion model

Multi-source data fusion methods are categorized into early fusion, late fusion, and hybrid fusion (Boulahia et al., 2021). As illustrated in Fig. 5, early fusion involves fusing data immediately after feature extraction, late fusion performs fusion only after generating data



**Fig. 5** Three multimodal fusion methods. **a** early fusion, **b** late fusion, **c** hybrid fusion

outputs (such as classification or regression results), and hybrid fusion integrates the outputs of early fusion methods with those of single-modal predictors.

## Precision evaluation

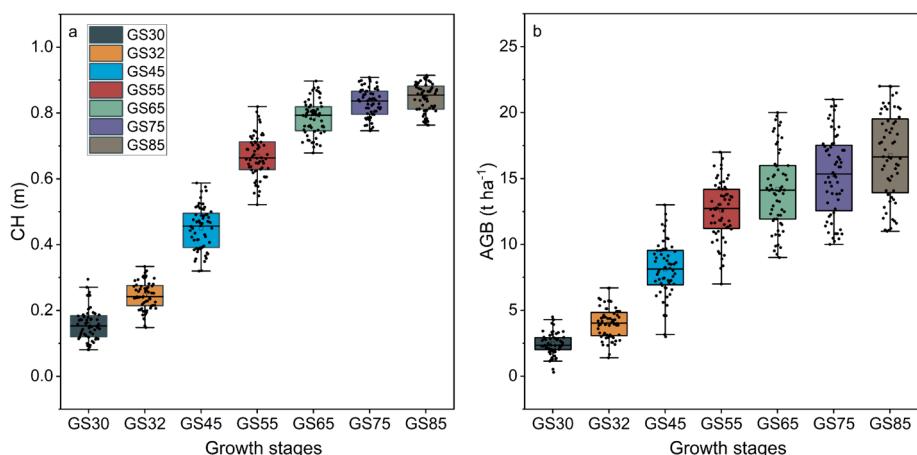
The accuracy of the AGB prediction model was evaluated using three key evaluation metrics: the coefficient of determination ( $R^2$ ), the root-mean-square error (RMSE), and the mean absolute error (MAE). The  $R^2$ , RMSE, and MAE values were calculated as follows:

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

$$cRMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (3)$$

$$cMAE = \frac{\sum_{i=1}^N | \hat{y}_i - y_i |}{N} \quad (4)$$

where  $N$  is the number of samples, and  $y_i$  and  $\hat{y}_i$  the measured and predicted AGB, respectively.  $\bar{y}_i$  is the average measured AGB.



**Fig. 6** Distributions of measured manually PH and AGB

## Results

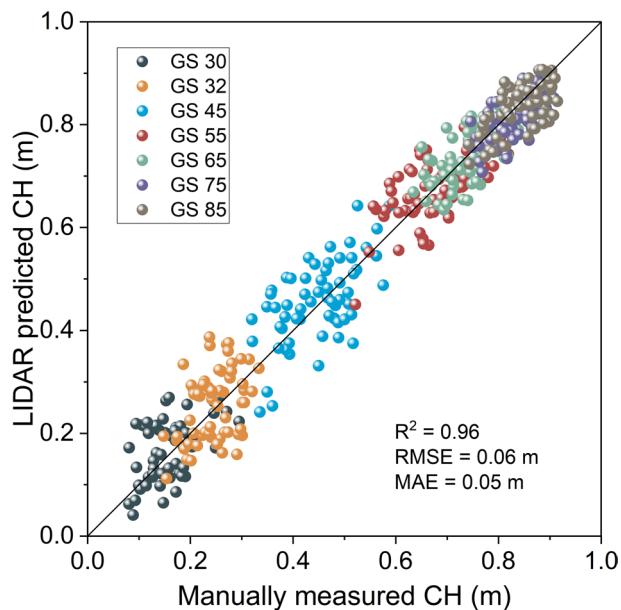
### Descriptive statistics

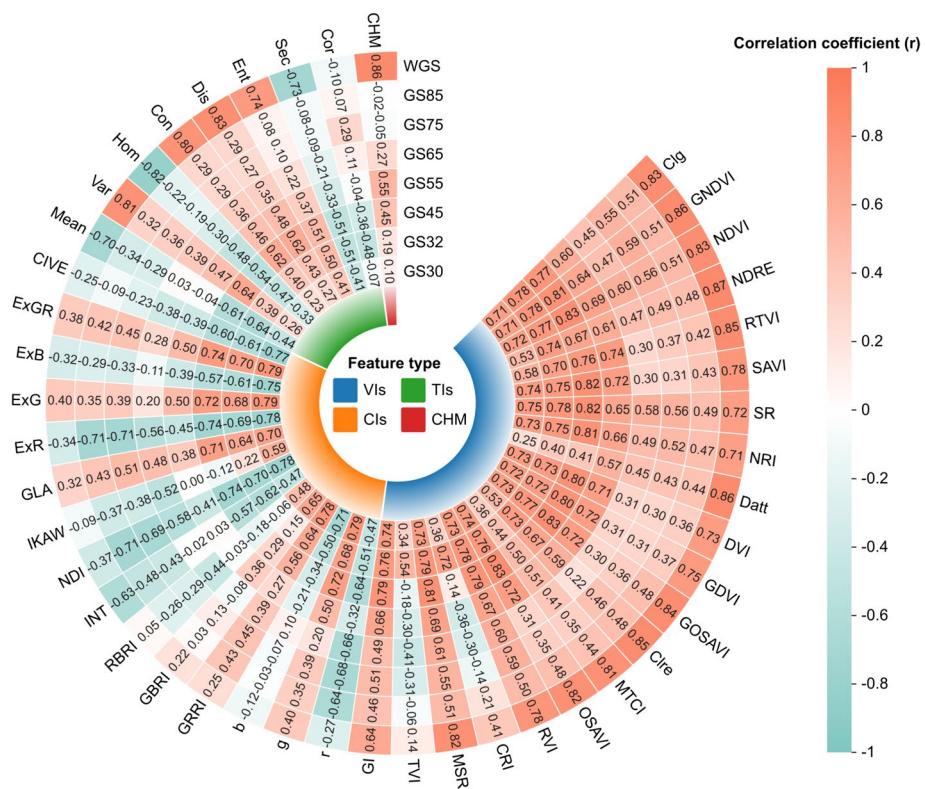
The PH and AGB were measured manually in 60 plots throughout the wheat growth stage. The results are shown in Fig. 6. The average CH values measured at GS30, GS32, GS45, GS55, GS65, GS75, and GS85 were 0.16 m, 0.25 m, 0.45 m, 0.67 m, 0.79 m, 0.83 m, and 0.84 m, respectively. The average AGB values measured at GS30, GS32, GS45, GS55, GS65, GS75, and GS85 were  $2.46 \text{ t ha}^{-1}$ ,  $4.02 \text{ t ha}^{-1}$ ,  $8.13 \text{ t ha}^{-1}$ ,  $12.57 \text{ t ha}^{-1}$ ,  $14.18 \text{ t ha}^{-1}$ ,  $15.26 \text{ t ha}^{-1}$ , and  $16.73 \text{ t ha}^{-1}$ , respectively. Both CH and AGB transitioned from gradual (GS30–32) to rapid (GS45–65) growth. However, while CH peaked at GS65 and then remained nearly constant, AGB continued to increase due to grain filling.

### Canopy height extraction and validation

Accurate measurement of CH is essential for predicting AGB. CH data extracted from UAV-based LIDAR point clouds was compared with manual ground-based CH measurements (Fig. 7). Results demonstrated a significant linear correlation ( $R^2 = 0.96$ ) between ground-based and LIDAR-extracted CH measurements, and correlation consistency was intensified with growth of winter wheat. During early growth stages ( $\text{CH} < 0.6 \text{ m}$ ), data points appear more scattered near the 1:1 line, while in the later stages ( $\text{CH} > 0.6 \text{ m}$ ), they exhibit a tendency to cluster more closely. Overall, high prediction accuracy was observed in 420 CH samples collected during seven growth stages.

**Fig. 7** Relationship between LIDAR predicted and manually measured CH





**Fig. 8** Pearson's correlation coefficient ( $r$ ) between AGB versus VIs, CIs, TIs, and CHM.  $r$  (0.05, 60)=0.254,  $r$  (0.01, 60)=0.330,  $r$  (0.05, 420)=0.096,  $r$  (0.01, 420)=0.126

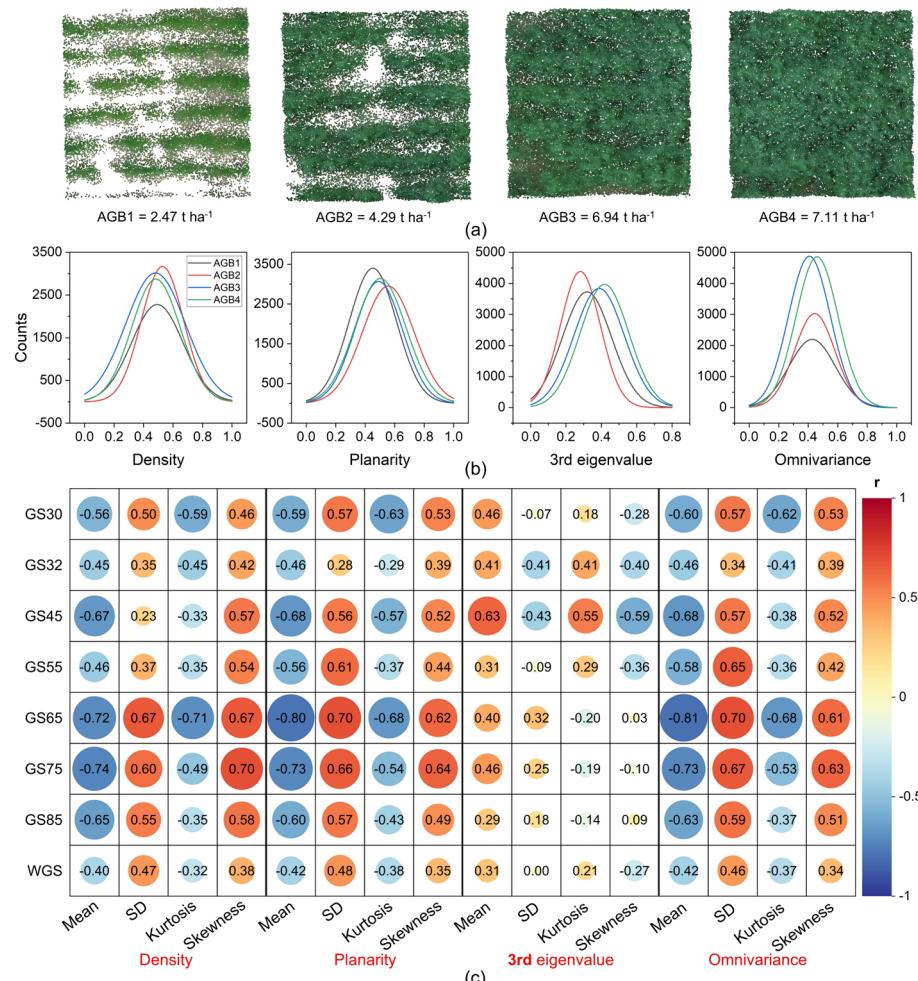
## Correlation analysis

The correlation between feature variables and AGB was examined through Pearson correlation analysis using IBM SPSS software. Overall, the combined data showed stronger correlations than data captured on a single growth stage (Fig. 8). NDRE produced the highest correlation ( $r=0.87$ ) among all features when using whole growth stage (WGS) data. This might be because of the relationship between features of a single growth stage and AGB is influenced by various factors including sensor accuracy and light intensity fluctuations. The features of UAV images may not accurately represent the actual state of AGB. However, when data from various growth stages was combined, these seasonal and transient factors effects were alleviated. Consequently, the long-term trends in image feature and AGB become clearer.

Comparing VIs with CIs, TIs, and CHM, the VIs exhibited a stronger correlation with AGB than other indices in most growth stages. This can be attributed to VIs typically containing the red-edge and near-infrared bands, effectively reflecting plant pigments content and therefore yielding a more accurate indication of AGB. Further, the correlation between most features and AGB initially intensifies and then diminishes with growth of wheat. This indicates that VIs is the most effective indicator for assessing AGB in the early and middle growth stages. However, as the growth stage progresses, the growth rate of the wheat

decelerates, and VIs tend to reach saturation. Such as during dough stage, the leaves, ears, stems, and leaves of some plots began to yellow, and VIs showed rapid changes and were unable to accurately depict the true condition of the AGB.

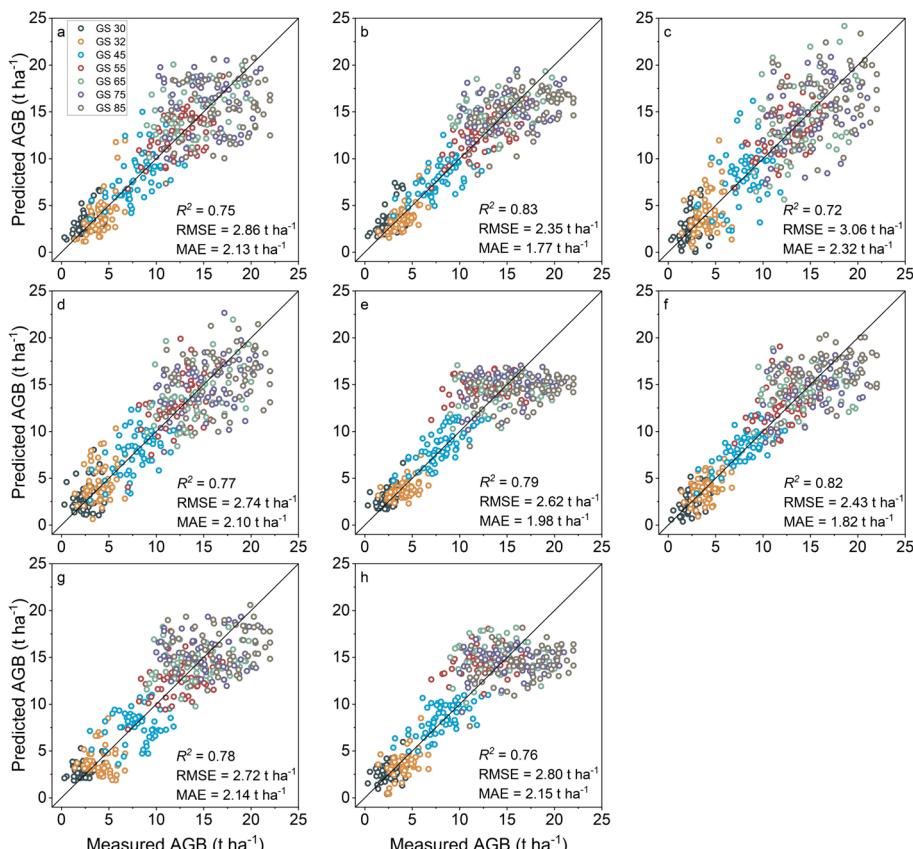
Comparing CIs, TIs, and CHM, the correlation between CIs and AGB was notably stronger in a single growth stage, but during the WGS correlation was comparatively lower, and  $r$  value reached a maximum of -0.63. However, TIs and CHM exhibits better performance during the WGS. The CHM exhibits the highest correlation, registering a  $r$  of 0.86, while the dissimilarity exhibits the highest correlation among the eight TIs, with a  $r$  of 0.83. CIs primarily indicate the appearance and health of the plant during specific growth stages. The color changes in the canopy between different treatments tend to be



**Fig. 9** Feasibility analysis of AGB prediction using point cloud depth features. **a** the top view of point clouds with different AGB values, **b** normal distribution curves of four traditional features with different AGB values, **c** the  $r$  between AGB and the features of normal distribution curves at different growth stages,  $SD$  standard deviation

quite pronounced. Consequently, CIs were found to be more appropriate for assessing the AGB during single growth stage. Given that TIs and CHM effectively capture the spatial distribution and structural changes in vegetation, and they are deemed more suitable for evaluating AGB changes during the WGS.

Four plots with different AGB values were selected during the GS45 stage (Fig. 9a), and four feature parameters were computed using the traditional method of point cloud feature extraction (density, planarity, third eigenvalue, and omnivariance). The normal distribution curves of these features are depicted in Fig. 9b. AGBs with different values exhibit significant differences in top views and curves, while those with nearly identical values show minor differences in point clouds. Mean, standard deviation, kurtosis, and skewness were extracted from the curves, and Pearson's correlation analysis was performed with AGB, as illustrated in Fig. 9c. Most curve characteristics exhibit a high correlation with AGB, and the mean of planarity and omnivariance curve shows the strongest correlation, with the  $r$  values were all greater than 0.8. The 3rd eigenvalue exhibits the weakest correlation with AGB, with a maximum  $r$  of only 0.63. Consequently, besides CH, other features of LIDAR



**Fig. 10** The scatter plots of predicted and measured AGB based on different inputs and fusion methods. **a** MS images, **b** VI images, **c** RGB images, **d** CI images, **e** TI images, **f** CI and TI images, **g** point cloud images, **h** CHM images

point clouds contribute to AGB prediction, which also shows the feasibility of using DL to extract depth features of point clouds for AGB prediction.

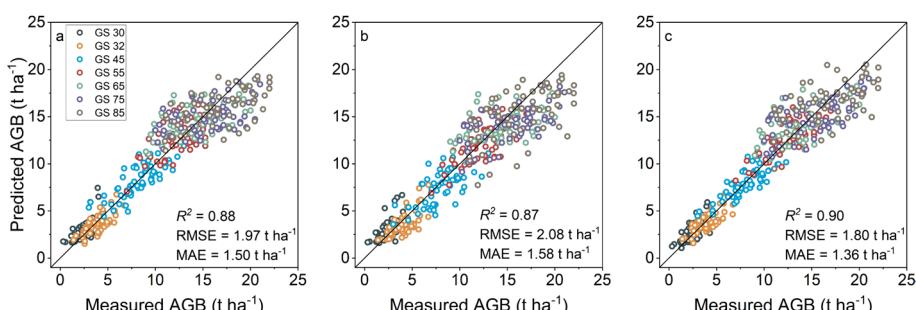
### AGB prediction based on single-source data

The wheat AGB prediction models were established by using the original images and the extracted feature images, and were then compared with the manually field measured AGB. The results of the 10-fold cross-validation are presented in Fig. 10. The results indicate that all features exhibited varying degrees of underprediction when predicting AGB at later growth stages, as the predicted values being lower than the measured values. Among them, TI and CHM showed the high underprediction. This might be attributed to the saturation of these features before flowering, leading to minimal changes in their values, while the AGB continues to increase. VIs images have the highest prediction accuracy, with an  $R^2$  of 0.83, and RMSE and MAE values of  $2.35 \text{ t ha}^{-1}$  and  $1.77 \text{ t ha}^{-1}$ , respectively. The RGB images possess the lowest prediction accuracy, with an  $R^2$  of only 0.72. Comparing original and feature images, the prediction accuracy of the MS images is found to be lower than that of the VI images, while the RGB images are less accurate than both the combined and individual CI and TI images. However, point cloud images exhibited slightly higher prediction accuracy than CHM images.

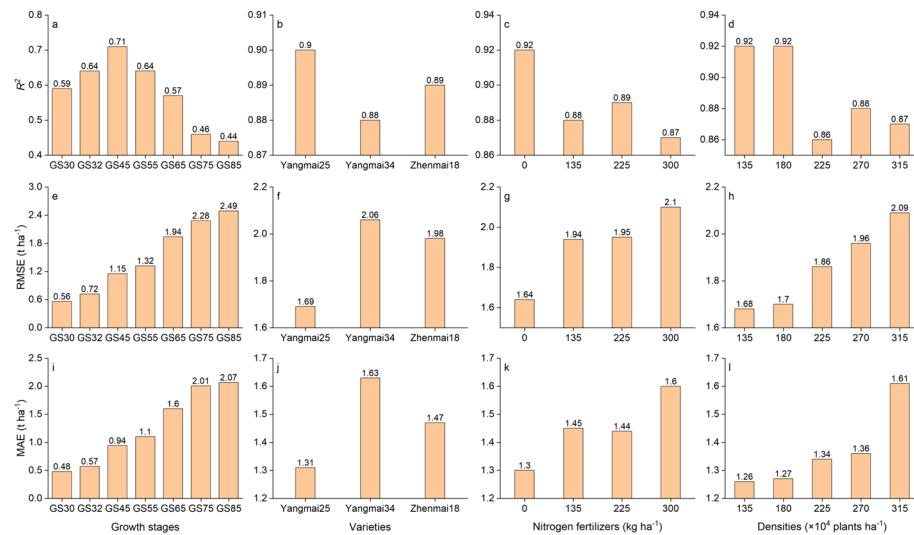
### AGB prediction based on multi-source data fusion

The wheat AGB prediction models were established using VI, CI, TI, and point cloud images and compared with the manually measured AGB. As depicted in Fig. 11, the hybrid fusion method exhibits the highest  $R^2$  value (0.90) and the lowest RMSE ( $1.80 \text{ t ha}^{-1}$ ) and MAE ( $1.36 \text{ t ha}^{-1}$ ), indicating its high accuracy and stability in predicting AGB. In contrast, the  $R^2$  values of the early and late fusion methods were 0.88 and 0.87, respectively, with RMSE values of  $1.97 \text{ t ha}^{-1}$  and  $2.08 \text{ t ha}^{-1}$ , and MAE values of  $1.50 \text{ t ha}^{-1}$  and  $1.58 \text{ t ha}^{-1}$ , respectively. This may be attributed to the fact that the early and late fusion methods were unable to fully capitalize on the complementarity of multi-source data, whereas the hybrid fusion method effectively integrates these advantages to enhance the prediction accuracy.

The robustness of the AGB prediction model was confirmed across different growth stages, varieties, nitrogen fertilizers, and densities. The results of a 10-fold cross-validation

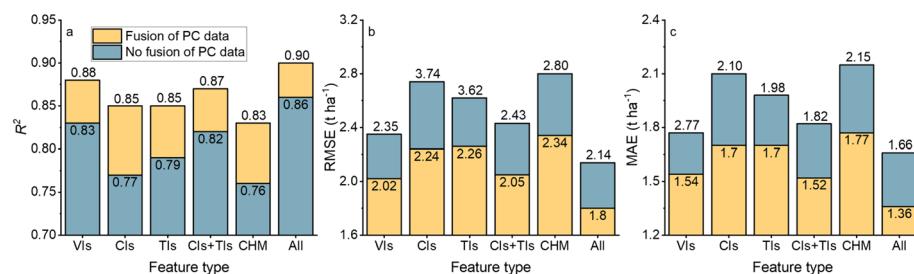


**Fig. 11** The scatter plots of predicted and measured AGB based on different fusion methods. **a** early fusion, **b** late fusion, **c** hybrid fusion



**Fig. 12** AGB prediction performance across different growth stages, varieties, nitrogen fertilizers, and planting densities. **a–d**  $R^2$ , **e–h** RMSE, **i–l** MAE

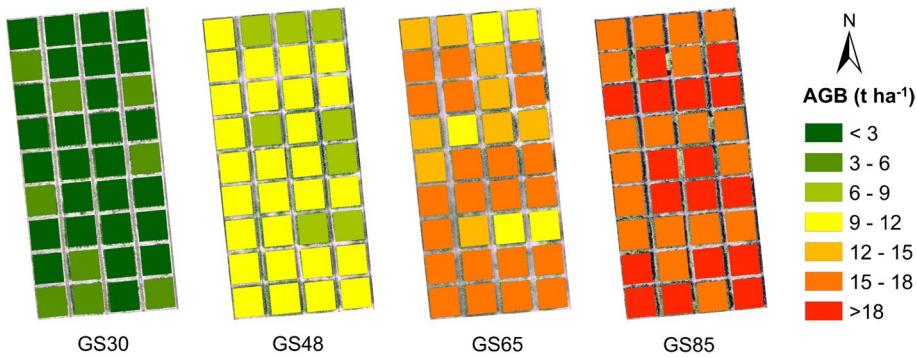
showed that as the growth process progresses, the  $R^2$  value demonstrates a pattern of initially increasing and then decreasing (Fig. 12). The highest  $R^2$  value (0.71) was observed at GS45, while the lowest (0.44) was noticed at GS85 (Fig. 12a). During the early growth stage, minimal variation in wheat growth between different plots is noted, with accuracy being significantly impacted by mixed pixels. After GS45, canopy density increases, the VI becomes saturated, and CH ceases to increase, while AGB continues to increase. These factors may contribute to inaccurate AGB prediction at relatively late growth stages. Moreover, yellowing of ears and upper part of the stem at GS85 results in substantial changes in the VI, further decreasing AGB prediction accuracy. The model consistently excelled in predicting AGB across various wheat varieties as  $R^2$  values for all varieties were above 0.88. Among the three varieties, the highest prediction accuracy ( $R^2=0.90$ ) was observed in Yangmai25 (Fig. 12b). This could be due to Yangmai25 taller height and smaller canopy density (Average CH of Yangmai25 was 0.83 m, Yangmai34 have 0.77 m, and Zhenmai18 was 0.79 m). The limited penetration capability of optical sensors renders them prone to saturation in AGB prediction. Generally, with the increase in fertilization



**Fig. 13** Influence of point cloud (PC) deep learning on AGB prediction

**Table 5** Results of adaptability test

Zadoks Scale	$R^2$	RMSE ( $t\ ha^{-1}$ )	MAE ( $t\ ha^{-1}$ )
GS30	0.63	0.61	0.50
GS48	0.66	1.03	0.88
GS65	0.57	1.92	1.72
GS85	0.46	2.43	2.05
All	0.82	2.62	2.09

**Fig. 14** Spatial distribution of prediction AGB in different growth stages

and planting density, there was rise in canopy density, which in turn leads to a reduction in  $R^2$  (Fig. 12c–d).

In order to explore the improvement in AGB prediction accuracy by point cloud deep learning, VI image, CI image, TI image and CHM image were fused with point cloud data respectively, and the results before and after fusion were compared as shown in Fig. 13. The results show that prediction accuracy for all feature images was improved after the fusion of point cloud data. Among them, accuracy in CI image improves the most after the fusion of point cloud data, as  $R^2$  increases from 0.77 to 0.85, while RMSE and MAE were decreased by 0.50  $t\ ha^{-1}$  and 0.40  $t\ ha^{-1}$ , respectively. Following the fusion of CI, TI, and CHM images with the VI images, the  $R^2$  value only increased from 0.83 to 0.86, whereas fusing the original point cloud images with VI images resulted in an  $R^2$  increase to 0.88. The contribution of each feature type was analyzed using the control variable method. Sequentially removing VIs, CIs, TIs, CHM, and PCs resulted in reductions in  $R^2$  by 0.03, 0.02, 0, 0, and 0.04, respectively. This indicate that PCs contribute the most to improved predictions and encompasses information absent in MS and RGB images.

## Adaptability test

To evaluate the accuracy and adaptability of the AGB prediction model, data from experiment 2 in 2023 was imported into the hybrid fusion model, and the results are presented in Table 5. When all data was modeled together, the  $R^2$  attains 0.82, depicting a slight decrease in prediction accuracy as compared to the training data results ( $R^2=0.90$ ). Considering the specific growth stage, GS48 stage exhibits the highest accuracy, with an  $R^2$  of 0.66, and RMSE and MAE values of  $1.03 \text{ t ha}^{-1}$  and  $0.88 \text{ t ha}^{-1}$ , respectively. However, at GS85 stage, the  $R^2$  value decreased to 0.46. Observing the overall trend, the model prediction accuracy aligns with the training data results (Fig. 12a). Further, the spatial distribution of predicted AGB is illustrated in Fig. 14, effectively presenting the overall growth status of wheat during different growth stages, thereby offering crucial decision-making information for field management.

## Discussion

### Influence of fusion methods on AGB prediction

Different sensors can obtain identical agronomic parameters, though their prediction accuracy varies. Typically, multi-modal data fusion modeling outperforms single-modal data approaches (Yu et al., 2023; Zhang et al., 2023). Zhu et al. (2021a, 2021b) found that the most effective multi-source UAV agricultural monitoring solution for LAI, AGB, and CH prediction is LiDAR+RGB+HS. However, for LWC and LCC prediction, single-mode Thermal imaging and hyperspectral imaging surpass multi-source data fusion. However, Different fusion methods have their own advantages and disadvantages, and the best method depends on the data.

Early fusion entails the incorporation of raw data sources at an initial stage of the data processing cycle, followed by subsequent processing and analysis of the comprehensive dataset (Maimaitijiang et al., 2017). To address inconsistencies among the original data in each modality, it is essential to separately extract features from the original data and then integrate these features (Boulahia et al., 2021). Early fusion yields comprehensive multimodal information, facilitating more in-depth analysis and decision-making (Gao et al., 2020). However, it can lead to high-dimensional data, elevated computing and storage costs (Audebert et al., 2018). Researchers frequently utilize dimensionality reduction techniques to mitigate redundancy in input features (Yue et al., 2023).

Late fusion entails independent feature extraction and analysis for each data source, followed by the fusion of these independent results at a later stage (Boulahia et al., 2021). This approach provides greater flexibility, permitting the selection of different analysis methods and models for each data source. Previous research indicates that early fusion outperforms late fusion when there is a significant correlation among modalities. In cases where the various modalities are largely unrelated, late fusion is a more appropriate method (Murphy, 2019).

Hybrid fusion integrates both early and late fusion, and was found to be particularly suitable for our study, which might be due to diverse and flexible nature of DL model structures. Hybrid fusion enables the utilization of different data fusion strategies at various stages, leveraging the advantages of each fusion approach (Boulahia et al., 2021). However, thorough problem analysis is required to ascertain when early and late fusion should be implemented. In this study, raster data (VI, CI, and TI images) underwent initial early

fusion and were subsequently fused with LIDAR point cloud data, resulting in the highest prediction accuracy.

## Factors influencing the AGB prediction accuracy

The uniformity of wheat plants and solar radiation are primary factors for accurate AGB prediction. Uniform plants facilitate the selection of regions of interest for sampling and the creation of shapefiles for batch clipping. When wheat plants are uneven, markers can ensure the shapefile and sampling locations overlap, thereby improving the alignment of variables and responses. Yang et al. (2024) also demonstrated that plant uniformity affects model accuracy. Although the technology in this study has been validated at two locations with similar climatic conditions, further verification is needed across different climatic and years. VIs primarily utilizes spectral reflectance information, closely related to solar radiation (Broge & Mortensen, 2002). Therefore, multispectral image acquisition should be conducted on clear days between 11:00 a.m. to 1:00 p.m., avoiding cloudy weather. Additionally, since solar radiation changes continuously, extensive farmland should be imaged in segments with radiometric calibration every half hour. LIDAR is unaffected by solar radiation; however, it should be avoided in foggy, hazy, rainy, or snowy weather (Zhu et al., 2021a, 2021b).

The image resolution, point cloud density, and vertical distribution of canopy structure are major factors influencing the accuracy of AGB prediction. The results indicated a decrease in AGB prediction accuracy as the image resolution and point cloud density decreased, as summarized in Table 6. Specifically,  $R^2$  declined from 0.90 to 0.82, while RMSE escalated from  $1.80 \text{ t ha}^{-1}$  to  $2.44 \text{ t ha}^{-1}$ , and MAE rose  $1.36 \text{ t ha}^{-1}$  to  $1.88 \text{ t ha}^{-1}$ . This decline in prediction accuracy can be attributed to several factors: (1) Low image resolution and point cloud density led to a decrease in the accuracy of that specific canopy surface details, and the 3D properties of canopy spatial structure (Lu et al., 2019). (2) Reduced image resolution and point cloud density contributed to an increased presence of mixed pixels (Ghahremani et al., 2021). This mixing influences the values of VI, CI, and TI image, further affecting AGB prediction accuracy.

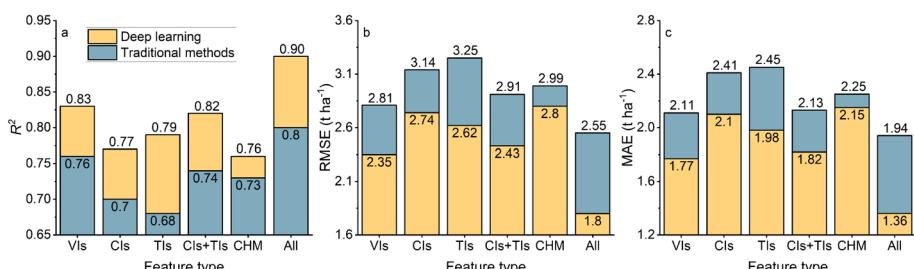
Tian et al. (2022) demonstrated that lower UAV flight altitudes result in higher image resolution, greater point cloud density, and more accurate feature extraction from images. However, this effect is primarily observed at the 2D level. The nature of LIDAR sensors makes them susceptible to LAI. As LAI increases, it becomes more challenging for the

**Table 6** AGB prediction based on different image resolution and point cloud density

Image resolution	Point cloud density	$R^2$	RMSE ( $\text{t ha}^{-1}$ )	MAE ( $\text{t ha}^{-1}$ )
$331 \times 331$	4096	0.90	1.80	1.36
$300 \times 300$	3584	0.89	1.92	1.46
$250 \times 250$	3072	0.87	2.10	1.61
$200 \times 200$	2560	0.85	2.22	1.67
$150 \times 150$	2048	0.84	2.28	1.71
$100 \times 100$	1536	0.83	2.37	1.80
$50 \times 50$	1024	0.82	2.44	1.88

LIDAR signal to penetrate the canopy and reach the ground. LAI peaks after the flowering stage, with higher density and fertilizer application further elevating LAI. This leads to insufficient canopy information acquisition by LIDAR, affecting the accuracy of AGB predictions. Consequently, the flight altitude of the Mavic 3 multispectral should ideally not exceed 20 m, and the Matrice 300 RTK should not exceed 30 m. However, low flight altitudes limit the technology's application in large-scale agricultural environments. Therefore, further research is needed to determine the optimal flight altitude that balances prediction accuracy and efficiency of AGB (Zhu et al., 2021a, 2021b). It's important to note that the reduction of image resolution and point cloud density may have different effects as compared to increasing flight altitude (Maimaitijiang et al., 2019). For instance, an increase in flight altitude may impair LiDAR ability to capture fine details at the plot edge. However, following point cloud downsampling, the points at the plot edge are retained in the data. Wan et al. (2020) also believed that elevating flight altitude may potentially exacerbate the error. The optimal LAI of crops generally approximates 4, with specific values ranging from 6 to 8.8 for wheat, 4 to 7 for rice, 5 for corn, and 3.2 for soybean (Li et al., 2023). This study achieved higher accuracy in predicting wheat AGB. Therefore, the method employed in this study holds potential applicability to other crops, potentially yielding higher prediction accuracy. Currently, corn-related data is being collected, and future research will investigate this method's applicability to corn.

Feature extraction and model selection are direct factors influencing the accuracy of AGB prediction. The features determine the upper limit of the effect, while the selection of models can only approach this upper limit asymptotically (Guo & Lin, 2021). Walter et al. (2018) utilized a consumer-level digital camera to image each plot, accurately estimating wheat AGB and harvest index using only synthesized 3D point cloud data. Fei et al. (2023) incorporated 3D point cloud data obtained through oblique photogrammetry with multispectral, RGB, and thermal infrared imaging, enhancing maize LAI and AGB estimation accuracy by 6.1% and 3.7%, respectively. Multispectral image processing is simpler than traditional methods as it omits the feature averaging step. Conversely, point cloud processing is more complex, requiring operations such as denoising, cropping, and downsampling. However, these tasks can be automated via code or software. Deep learning outperforms traditional machine learning, particularly in feature extraction, a finding substantiated in the prediction of AGB (Yu et al., 2023), LAI (Fei et al., 2023), nitrogen content (Tian et al., 2011), and yield (Wan et al., 2020).



**Fig. 15** Performance comparison of traditional methods and image deep learning

## Performance comparison of traditional methods and image deep learning

Prior to the rapid development in deep learning approaches, traditional AGB prediction methods seldom utilized images as input, particularly point cloud images (Ma et al., 2019). To compare the performance of AGB prediction through traditional methods and deep learning, the average values of VIs, CIs, TIs, and CHM for each plot were calculated, and these selected features were then input into the regression model as one-dimensional vectors. Six statistical or machine learning regression models were employed: multiple linear regression (MLR), support vector machine regression (SVM), Gaussian process regression (GPR), random forest (RF), gradient boosting decision tree (GBDT), and one-dimensional convolutional neural network (1DCNN). Figure 15 displays the results of the best-performing model based on six different datasets. The results indicate that among single-source data, VIs attained the highest prediction accuracy, with an  $R^2$  of 0.76, RMSE and MAE of  $2.81 \text{ t ha}^{-1}$  and  $2.11 \text{ t ha}^{-1}$ , respectively. Following data fusion, the  $R^2$  value escalated to its apex of 0.80. However, the  $R^2$  value is noticeably diminished in traditional methods as compared to deep learning methods. The potential explanations for this phenomenon are as follows: (1) Heterogeneity in spatial distribution is frequently overlooked when utilizing the mean value of feature images for AGB prediction. For instance, the difference in fractional vegetation cover across different segments of a plot often goes unrepresented, as averages do not adequately reflect this disparity. (2) Deep learning, particularly convolutional neural networks, exhibits exceptional efficacy in processing image data (Yu et al., 2023). Additionally, these algorithms significantly improve AGB prediction accuracy by a thorough analysis of multiple features and the spatial relationships within images.

## Conclusion

The research focused on the potential of wheat AGB prediction using the 2D image and point cloud deep learning. In single-sensor AGB prediction scenarios, VI images outperform CI, TI, and CHM images. However, feature fusion of CI and TI images emerged as a viable alternative to VI images, offering comparable prediction accuracy with  $R^2$  only decreasing by 0.01. The deep learning fusion model utilizing VI, CI, TI, and point cloud images as inputs yielded superior predictions for wheat AGB, achieving an  $R^2$  value exceeding 0.88. Relative to traditional eigenvector-based AGB prediction, this approach increased  $R^2$  by 12.5% and reduced the RMSE and MAE by 29.4% and 29.9%, respectively.

The research evaluated the impact of various fusion methods on AGB prediction, ranking their accuracy as follows: hybrid fusion > early fusion > late fusion, with hybrid achieving the highest  $R^2$  of 0.90. This fusion model demonstrates high robustness and accuracy across varying years, growth stages, crop varieties, nitrogen fertilizers, and densities. Considering the demonstrated success in wheat AGB prediction, the depth features of LIDAR point cloud data show significant potential for predicting wheat AGB, and this worth further exploration.

**Acknowledgements** This work was supported by Key Research and Development Program (Modern Agriculture) of Jiangsu Province (BE2022335, BE2022338), National Natural Science Foundation of China (32261143462, 321722110, 32071945), National Key Research and Development Program of China (2018YFD0300805, 2023YFD1202200), and Jiangsu Provincial Postgraduate Scientific Research Innovation Program (KYCX21\_3240).

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Ahamed, T., Tian, L., Zhang, Y., & Ting, K. C. (2011). A review of remote sensing methods for biomass feedstock production. *Biomass and Bioenergy*, 35(7), 2455–2469. <https://doi.org/10.1016/j.biombioe.2011.02.028>
- Astor, T., Dayananda, S., Nautiyal, S., & Wachendorf, M. (2020). Vegetable crop biomass estimation using hyperspectral and RGB 3D UAV data. *Agronomy-Basel*, 10(10), 1600. <https://doi.org/10.3390/agronomy10101600>
- Audebert, N., Le Saux, B., & Lefèvre, S. (2018). Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140, 20–32. <https://doi.org/10.1016/j.isprsjprs.2017.11.011>
- Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, 39, 79–87. <https://doi.org/10.1016/j.jag.2015.02.012>
- Boulahia, S. Y., Amamra, A., Madi, M. R., & Daikh, S. (2021). Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition. *Machine Vision and Applications*, 32(6), 121. <https://doi.org/10.1007/s00138-021-01249-8>
- Broge, N. H., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76(2), 156–172. [https://doi.org/10.1016/s0034-4257\(00\)00197-8](https://doi.org/10.1016/s0034-4257(00)00197-8)
- Broge, N. H., & Mortensen, J. V. (2002). Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data. *Remote Sensing of Environment*, 81(1), 45–57. [https://doi.org/10.1016/s0034-4257\(01\)00332-7](https://doi.org/10.1016/s0034-4257(01)00332-7)
- Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., & Gattelli, M. (2015). Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sensing*, 7(4), 4026–4047. <https://doi.org/10.3390/rs70404026>
- Cao, Q., Miao, Y. X., Feng, G. H., Gao, X. W., Li, F., Liu, B., et al. (2015). Active canopy sensing of winter wheat nitrogen status: An evaluation of two sensor systems. *Computers and Electronics in Agriculture*, 112, 54–67. <https://doi.org/10.1016/j.compag.2014.08.012>
- Cen, H. Y., Wan, L., Zhu, J. P., Li, Y. J., Li, X. R., Zhu, Y. M., et al. (2019). Dynamic monitoring of biomass of rice under different nitrogen treatments using a lightweight UAV with dual image-frame snapshot cameras. *Plant Methods*, 15, 32. <https://doi.org/10.1186/s13007-019-0418-8>
- Chen, P. F., Tremblay, N., Wang, J. H., Vigneault, P., Huang, W. J., & Li, B. G. (2010). New index for crop canopy fresh biomass estimation. *Spectroscopy and Spectral Analysis*, 30(2), 512–517. [https://doi.org/10.3964/j.issn.1000-0593\(2010\)02-0512-06](https://doi.org/10.3964/j.issn.1000-0593(2010)02-0512-06)
- Comba, L., Zaman, S., Biglia, A., Aimorino, D. R., Dabbene, F., & Gay, P. (2020). Semantic interpretation and complexity reduction of 3D point clouds of vineyards. *Biosystems Engineering*, 197, 216–230. <https://doi.org/10.1016/j.biosystemseng.2020.05.013>
- Dash, J., & Curran, P. J. (2004). The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing*, 25(23), 5403–5413. <https://doi.org/10.1080/0143116042000274015>
- Datt, B. (1999). Remote sensing of water content in eucalyptus leaves. *Australian Journal of Botany*, 47(6), 909–923. <https://doi.org/10.1071/bt98042>
- Daughtry, C. S. T., Walthall, C. L., Kim, M. S., de Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74(2), 229–239. [https://doi.org/10.1016/s0034-4257\(00\)00113-9](https://doi.org/10.1016/s0034-4257(00)00113-9)
- Fei, S. P., Xiao, S. F., Li, Q., Shu, M. Y., Zhai, W. G., Xiao, Y. G., et al. (2023). Enhancing leaf area index and biomass estimation in maize with feature augmentation from unmanned aerial vehicle-based nadir and cross-circling oblique photography. *Computers and Electronics in Agriculture*, 215, 108462. <https://doi.org/10.1016/j.compag.2023.108462>
- Gamon, J. A., & Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytologist*, 143(1), 105–117. <https://doi.org/10.1046/j.1469-8137.1999.00424.x>

- Gamon, J. A., Field, C. B., Goulden, M. L., Griffin, K. L., Hartley, A. E., Joel, G., et al. (1995). Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications*, 5(1), 28–41. <https://doi.org/10.2307/1942049>
- Gao, J., Li, P., Chen, Z. K., & Zhang, J. N. (2020). A survey on deep learning for multimodal data fusion. *Neural Computation*, 32(5), 829–864. [https://doi.org/10.1162/neco\\_a\\_01273](https://doi.org/10.1162/neco_a_01273)
- Ghahremani, M., Williams, K., Corke, F. M. K., Tiddeman, B., Liu, Y. H., & Doonan, J. H. (2021). Deep segmentation of point clouds of wheat. *Frontiers in Plant Science*, 12, 608732. <https://doi.org/10.3389/fpls.2021.608732>
- Gill, T., Gill, S. K., Saini, D. K., Chopra, Y., de Koff, J. P., & Sandhu, K. S. (2022). A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics (Cham Switzerland)*, 2(3), 156–183. <https://doi.org/10.1007/s43657-022-00048-z>
- Guizarro, M., Pajares, G., Riomoros, I., Herrera, P. J., Burgos-Artizú, X. P., & Ribeiro, A. (2011). Automatic segmentation of relevant textures in agricultural images. *Computers and Electronics in Agriculture*, 75(1), 75–83. <https://doi.org/10.1016/j.compag.2010.09.013>
- Guo, Z., & Lin, B. (2021). Machine learning stability and band gap of lead-free halide double perovskite materials for perovskite solar cells. *Solar Energy*, 228, 689–699. <https://doi.org/10.1016/j.solener.2021.09.030>
- Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3), 337–352. <https://doi.org/10.1016/j.rse.2003.12.013>
- Hall-Beyer, M. (2017). Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *International Journal of Remote Sensing*, 38(5), 1312–1338. <https://doi.org/10.1080/01431161.2016.1278314>
- Han, L., Yang, G. J., Dai, H. Y., Xu, B., Yang, H., Feng, H. K., et al. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods*, 15, 10. <https://doi.org/10.1186/s13007-019-0394-z>
- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC*, 3(6), 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2), 195–213. [https://doi.org/10.1016/s0034-4257\(02\)00096-2](https://doi.org/10.1016/s0034-4257(02)00096-2)
- Kawashima, S., & Nakatani, M. (1998). An algorithm for estimating chlorophyll content in leaves using a video camera. *Annals of Botany (London)*, 81(1), 49–54. <https://doi.org/10.1006/anbo.1997.0544>
- Kerkech, M., Hafiane, A., & Canals, R. (2020). Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. *Computers and Electronics in Agriculture*, 174, 105446. <https://doi.org/10.1016/j.compag.2020.105446>
- Li, Z. B., Guo, R. H., Li, M., Chen, Y. R., & Li, G. Y. (2020a). A review of computer vision technologies for plant phenotyping. *Computers and Electronics in Agriculture*, 176, 105672. <https://doi.org/10.1016/j.compag.2020.105672>
- Li, B., Xu, X. M., Zhang, L., Han, J. W., Bian, C. S., Li, G. C., et al. (2020b). Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162, 161–172. <https://doi.org/10.1016/j.isprsjprs.2020.02.013>
- Li, D. W., Shi, G. L., Li, J. S., Chen, Y. L., Zhang, S. Y., Xiang, S. Y., et al. (2022). PlantNet: A dual-function point cloud segmentation network for multiple plant species. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184, 243–263. <https://doi.org/10.1016/j.isprsjprs.2022.01.007>
- Li, Z. H., Zhan, C. S., Hu, S., Ning, L. K., Wu, L. F., & Guo, H. (2023). Implementation of a dynamic specific leaf area (SLA) into a land surface model (LSM) incorporated crop-growth model. *Computers and Electronics in Agriculture*, 213, 108238. <https://doi.org/10.1016/j.compag.2023.108238>
- Liu, Y. N., Liu, S. S., Li, J., Guo, X. Y., Wang, S. Q., & Lu, J. W. (2019). Estimating biomass of winter oilseed rape using vegetation indices and texture metrics derived from UAV multispectral images. *Computers and Electronics in Agriculture*, 166, 105026. <https://doi.org/10.1016/j.compag.2019.105026>
- Liu, F. S., Song, Q. F., Zhao, J. K., Mao, L. X., Bu, H. Y., Hu, Y., et al. (2021). Canopy occupation volume as an indicator of canopy photosynthetic capacity. *New Phytologist*, 232(2), 941–956. <https://doi.org/10.1111/nph.17611>
- Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, 16(1), 65–70. <https://doi.org/10.1080/10106040108542184>

- Lu, N., Zhou, J., Han, Z. X., Li, D., Cao, Q., Yao, X., et al. (2019). Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods*, 15, 17. <https://doi.org/10.1186/s13007-019-0402-3>
- Lu, J. S., Eitel, J. U. H., Engels, M., Zhu, J., Ma, Y., Liao, F., et al. (2021). Improving unmanned aerial vehicle (UAV) remote sensing of rice plant potassium accumulation by fusing spectral and textural information. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102592. <https://doi.org/10.1016/j.jag.2021.102592>
- Ma, J. C., Li, Y. X., Chen, Y. Q., Du, K. M., Zheng, F. X., Zhang, L. X., et al. (2019). Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *European Journal of Agronomy*, 103, 117–129. <https://doi.org/10.1016/j.eja.2018.12.004>
- Maimaitijiang, M., Ghulam, A., Sidike, P., Hartling, S., Maimaitiyiming, M., Peterson, K., et al. (2017). Unmanned aerial system (UAS)-based phenotyping of soybean using multi-sensor data fusion and extreme learning machine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134, 43–58. <https://doi.org/10.1016/j.isprsjprs.2017.10.011>
- Maimaitijiang, M., Sagan, V., Sidike, P., Maimaitiyiming, M., Hartling, S., Peterson, K. T., et al. (2019). Vegetation index weighted canopy volume model (CVMVI) for soybean biomass estimation from unmanned aerial system-based rgb imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 151, 27–41. <https://doi.org/10.1016/j.isprsjprs.2019.03.003>
- Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, 237, 111599. <https://doi.org/10.1016/j.rse.2019.111599>
- Murphy, R. R. (2019). Computer vision and machine learning in science fiction. *Science Robotics*, 4(30), eaax7421. <https://doi.org/10.1126/scirobotics.aax7421>
- Orusa, T., Viani, A., Cammareri, D., & Mondino, E. B. (2023). A Google earth engine algorithm to map phenological metrics in mountain areas worldwide with landsat collection and sentinel-2. *Geomatics*, 3(1), 221–238. <https://doi.org/10.3390/geomatics3010012>
- Orusa, T., Viani, A., Borgogno-Mondino, E., & Suziedelyte-Visockiene, J. (2024). Earth observation data and geospatial deep learning AI to assign contributions to European municipalities Sen4MUN: An empirical application in aosta valley (NW Italy). *Land*, 13(1), 80. <https://doi.org/10.3390/land13010080>
- Panneton, B., & Brouillard, M. (2009). Colour representation methods for segmentation of vegetation in photographs. *Biosystems Engineering*, 102(4), 365–378. <https://doi.org/10.1016/j.biosystemseng.2009.01.003>
- Qi, C., Yi, L., Su, H., & Guibas, L. J. (2017). PointNet++: Deep hierarchical feature learning on point sets in a metric space. Preprint at <https://doi.org/10.48550/arXiv.1706.02413>
- Quan, L. Z., Li, H. D., Li, H. L., Jiang, W., Lou, Z. X., & Chen, L. Q. (2021). Two-stream dense feature fusion network based on RGB-D data for the real-time prediction of weed aboveground fresh weight in a field environment. *Remote Sensing*, 13(12), 2288. <https://doi.org/10.3390/rs13122288>
- Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., & Jeganathan, C. (2012). Random forest classification of mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, 121, 93–107. <https://doi.org/10.1016/j.rse.2011.12.003>
- Stagakis, S., Markos, N., Sykioti, O., & Kyparissis, A. (2010). Monitoring canopy biophysical and biochemical parameters in ecosystem scale using satellite hyperspectral imagery: An application on a Phlomis fruticosa Mediterranean ecosystem using multiangular CHRIS/PROBA observations. *Remote Sensing of Environment*, 114(5), 977–994. <https://doi.org/10.1016/j.rse.2009.12.006>
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1602.07261>
- Tian, Y. C., Yao, X., Yang, J., Cao, W. X., Hannaway, D. B., & Zhu, Y. (2011). Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground- and space-based hyperspectral reflectance. *Field Crops Research*, 120(2), 299–310. <https://doi.org/10.1016/j.fcr.2010.11.002>
- Tian, Y. C., Zhang, Q., Huang, H., Huang, Y. J., Tao, J., Zhou, G. Q., et al. (2022). Aboveground biomass of typical invasive mangroves and its distribution patterns using UAV-LiDAR data in a subtropical estuary: Maoling River Estuary, Guangxi, China. *Ecological Indicators*, 136, 108694. <https://doi.org/10.1016/j.ecolind.2022.108694>
- Walter, J., Edwards, J., McDonald, G., & Kuchel, H. (2018). Photogrammetry for the estimation of wheat biomass and harvest index. *Field Crops Research*, 216, 165–174. <https://doi.org/10.1016/j.fcr.2017.11.024>

- Wan, L., Cen, H. Y., Zhu, J. P., Zhang, J. F., Zhu, Y. M., Sun, D. W., et al. (2020). Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer—A case study of small farmlands in the South of China. *Agricultural and Forest Meteorology*, 291, 108096. <https://doi.org/10.1016/j.agrformet.2020.108096>
- Wang, L. Y., Miao, Y. L., Han, Y. X., Li, H., Zhang, M., & Peng, C. (2023). Extraction of 3D distribution of potato plant CWSI based on thermal infrared image and binocular stereovision system. *Frontiers in Plant Science*, 13, 1104390. <https://doi.org/10.3389/fpls.2022.1104390>
- Woebbecke, D. M., Meyer, G. E., Von Bargen, K., & Mortensen, D. A. (1993). Plant species identification, size, and enumeration using machine vision techniques on near-binary images. In *Proceedings of Optics in Agriculture and Forestry* (pp. 208–219). The International Society for Optical Engineering. <https://doi.org/10.1117/12.144030>
- Woebbecke, D. M., Meyer, G. E., Bargen, K. V., & Mortensen, D. A. (1994). Color indices for weed identification under various soil, residue, and lighting conditions. *American Society of Agricultural and Biological Engineers*, 38, 259–269. <https://doi.org/10.13031/2013.27838>
- Wu, W. C. (2014). The generalized difference vegetation index (GDVI) for dryland characterization. *Remote Sensing*, 6(2), 1211–1233. <https://doi.org/10.3390/rs6021211>
- Yang, Y., Li, Q., Mu, Y., Li, H., Wang, H., Ninomiya, S., et al. (2024). UAV-assisted dynamic monitoring of wheat uniformity toward yield and biomass estimation. *Plant Phenomics (Washington D C)*, 6, 0191. <https://doi.org/10.34133/plantphenomics.0191>
- Yu, D. Y., Zha, Y. Y., Sun, Z. G., Li, J., Jin, X. L., Zhu, W. X., et al. (2023). Deep convolutional neural networks for estimating maize above-ground biomass using multi-source UAV images: A comparison with traditional machine learning algorithms. *Precision Agriculture*, 24(1), 92–113. <https://doi.org/10.1007/s11119-022-09932-0>
- Yue, J. B., Yang, G. J., Li, C. C., Li, Z. H., Wang, Y. J., Feng, H. K., et al. (2017). Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models. *Remote Sensing*, 9(7), 708. <https://doi.org/10.3390/rs9070708>
- Yue, J. B., Yang, H., Yang, G. J., Fu, Y. Y., Wang, H., & Zhou, C. Q. (2023). Estimating vertically growing crop above-ground biomass based on UAV remote sensing. *Computers and Electronics in Agriculture*, 205, 107627. <https://doi.org/10.1016/j.compag.2023.107627>
- Zarco-Tejada, P. J., Miller, J. R., Noland, T. L., Mohammed, G. H., & Sampson, P. H. (2001). Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7), 1491–1507. <https://doi.org/10.1109/36.934080>
- Zhang, R. P., Zhou, J. H., Guo, J., Miao, Y. H., & Zhang, L. L. (2023). Inversion models of aboveground grassland biomass in Xinjiang based on multisource data. *Frontiers in Plant Science*, 14, 1152432. <https://doi.org/10.3389/fpls.2023.1152432>
- Zhu, Y., Yao, X., Tian, Y. C., Liu, X. J., & Cao, W. X. (2008). Analysis of common canopy vegetation indices for indicating leaf nitrogen accumulations in wheat and rice. *International Journal of Applied Earth Observation and Geoinformation*, 10(1), 1–10. <https://doi.org/10.1016/j.jag.2007.02.006>
- Zhu, Y. L., Sun, G., Ding, G. H., Zhou, J., Wen, M. X., Jin, S. C., et al. (2021a). Large-scale field phenotyping using backpack LiDAR and CropQuant-3D to measure structural variation in wheat. *Plant Physiology*, 187(2), 716–738. <https://doi.org/10.1093/plphys/kiab324>
- Zhu, W. X., Sun, Z. G., Huang, Y. H., Yang, T., Li, J., Zhu, K. Y., et al. (2021b). Optimization of multi-source UAV RS agro-monitoring schemes designed for field-scale crop phenotyping. *Precision Agriculture*, 22(6), 1768–1802. <https://doi.org/10.1007/s11119-021-09811-0>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Shaolong Zhu<sup>1,2</sup>  · Weijun Zhang<sup>1,2</sup> · Tianle Yang<sup>1,2</sup> · Fei Wu<sup>3</sup> · Yihan Jiang<sup>1,2</sup> · Guanshuo Yang<sup>1,2</sup> · Muhammad Zain<sup>1,2</sup> · Yuanyuan Zhao<sup>1,2</sup> · Zhaosheng Yao<sup>1,2</sup> · Tao Liu<sup>1,2</sup> · Chengming Sun<sup>1,2</sup>

✉ Chengming Sun  
cmsun@yzu.edu.cn

<sup>1</sup> Key Laboratory of Crop Genetics and Physiology of Jiangsu Province, Key Laboratory of Crop Cultivation and Physiology of Jiangsu Province, College of Agriculture, Yangzhou University, Yangzhou 225009, China

<sup>2</sup> Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou 225009, China

<sup>3</sup> Precision Agriculture Lab, School of Life Sciences, Technical University of Munich, 85354 Freising, Germany