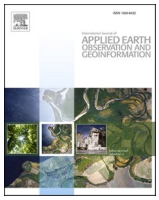
International Journal of Applied Earth Observation and Geoinformation 127 (2024) 103681

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Multimodal deep fusion model based on Transformer and multi-layer residuals for assessing the competitiveness of weeds in farmland ecosystems

Zhaoxia Lou [b], Longzhe Quan [a,*], Deng Sun [b], Fulin Xia [c], Hailong Li [b], Zhiming Guo [a]

[a] College of Engineering, Anhui Agricultural University, Anhui 230036, China
[b] College of Engineering, Northeast Agricultural University, Harbin 150030, China
[c] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China

ARTICLE INFO

Keywords:
Weed competition monitoring
UAV remote sensing
Multimodal data fusion
Deep learning
Multilayer residual fusion
Transformer Encoder

ABSTRACT

Weed competitiveness monitoring is crucial for field management at specific locations. Recent research in the fusion of multimodal data from unmanned aerial vehicles (UAVs) has propelled this advancement. However, these studies merely stack extracted features equivalently, neglecting the full utilization of fused information. This study utilizes hyperspectral and LiDAR data collected by UAVs to proposes a multimodal deep fusion model (MulDFNet) using Transformer and multi-layer residuals. It utilizes a comprehensive competitive index (CCI-A) based on multidimensional phenotypes of maize to assess the competitiveness of weeds in farmland ecosystems. To validate the effectiveness of this model, a series of ablation studies were conducted involving different modalities data, with/without the Transformer Encoder (TE) modules, and different fusion modules (shallow residual fusion module, deep feature fusion module). Additionally, a comparison was made with early/late stacking fusion models, traditional machine learning models, and deep learning models from relevant studies. The results indicate that the multimodal deep fusion model utilizing HSI, VI, and CHM data achieved a predictive effect of $R^2 = 0.903$ (RMSE = 0.078). Notably, the best performance was observed during the five-leaf stage. The combination of shallow and deep fusion modules demonstrated better predictive performance compared to a single fusion module. The positive impact of the TE module on model performance is evident, as its multi-head attention mechanism aids in better capturing the relationships and importance between feature maps and competition indices, thereby enhancing the model's predictive capability. In weed competition prediction, the multimodal deep fusion model proposed in this study has demonstrated significantly better predictive performance compared to early/late stacking fusion models and other machine learning models (RF, SVR, PLS, DNN-F2 and Multi-channel CNN). Overall, the multimodal deep fusion model developed in this study demonstrates outstanding performance in assessing weed competitiveness and can predict the competitive intensity of weeds in maize across various growth stages on a broad scale.

1. Introduction

The competitive behavior of weeds, as they vie for crucial resources essential for maize growth, adversely impacts the normal growth and development of maize (Bada et al., 2022). Nonetheless, maintaining a moderate level of weeds in farmland contributes to preserving ecological balance, reducing the risk of soil loss, enhancing nutrient circulation, and naturally controlling pests, among other multiple benefits (Scavo and Mauromicale, 2020). The intensity of competition is not constant among crops and weeds, it exhibits dynamic changes as plants gradually grow (Fang et al., 2018). Due to the dynamic nature of competition, it is necessary to study weed competition during multiple growth stages of

Abbreviations: UAV, Unmanned aerial vehicle; CCI-A, Comprehensive competition indices; RCI, Relative competitive intensity; CI, Competition intensity index; CB, Competitive balance index; HSI, Hyperspectral imaging; LiDAR, Light Detection and Ranging; GSD, Ground sampling distance; DOM, Domain images; PH, Plant height; ST, Stalk thickness; N, Nitrogen elements; P, Phosphorus elements; RF, Random Forest; PLS, Partial Least Squares; OSAVI, Optimized soil adjusted vegetation index; CHM, Canopy height model; VI, Vegetation indices; PCA, Principal component analysis; PC1, The first principal component; SPA, Successive Projections Algorithm; TE, Transformer Encoder; MSA, Multi-Head Self-Attention layer; MLP, Multi-Layer Perceptron; LN, Layer Normalization; SA, Self-Attention; R^2, Coefficient of determination; RMSE, Root mean squared error; SVR, Support Vector Regression.

* Corresponding author.
E-mail address: quanlongzhe@163.com (L. Quan).

https://doi.org/10.1016/j.jag.2024.103681
Received 18 October 2023; Received in revised form 10 January 2024; Accepted 22 January 2024
Available online 29 January 2024

maize. Thereby achieving scientific and ecological weed management, all while preserving weed diversity.

Currently, widely used indicators for quantifying the intensity of competition include Relative Competition Intensity (RCI), Competitive Balance Index (CB), and Competition Intensity Index (CI) (Weigelt and Jolliffe, 2003). However, their representation of competition is not as comprehensive and profound due to their reliance on single data metric (Lazzaro et al., 2019; Rasmussen and Nielsen, 2020; Vajari, 2021). To overcome these limitations, Lou and her team introduced the Comprehensive Competition Index (CCI), a multidimensional assessment metric that incorporates various aspects such as plant structure and biochemical factors (Quan et al., 2023).

Unmanned Aerial Vehicles (UAVs) that carry small sensors acquire phenotypic information about crops in agricultural environments to monitor crop growth conditions (Pipatsitee et al., 2022). For instance, utilizing Hyperspectral Imaging (HSI) technology to collect canopy spectral information allows for the assessment of plant responses to stressors such as drought, pests, diseases, and thermal stress (Kuswidiyanto et al., 2022; Zovko et al., 2019). Furthermore, employing Light Detection and Ranging (LiDAR) technology enables the evaluation of parameters including crop density, canopy height, above-ground biomass and nitrogen uptake by utilizing ground elevation information (Bates et al., 2021; Eitel et al., 2014; Hütt et al., 2023). However, due to the distinct technical characteristics and data features of different sensors, data from a single sensor is challenging to fully depict the actual ground information (Li et al., 2022). To overcome this issue, multimodal remote sensing methods are gradually becoming more prevalent. In multimodal approaches, when fusing together various modes data from the same region or target, each modality's information can complement each other, fully leveraging their strengths while mitigating their limitations (Li et al., 2022). This integration provides more comprehensive and multi-layered surface information for agricultural monitoring, enabling the formulation of more robust and reliable field management decisions. Multimodal remote sensing methods have been applied in various aspects, including crop monitoring and classification, land cover mapping, yield estimation, and detection of plant stress caused by climate change (Karmakar et al., 2024; Maimaitijiang et al., 2020).

For multimodal remote sensing fusion methods, in the early stages of research, there was a predominant use of traditional machine learning techniques, such as SVR and RF (Almeida et al., 2021). As research has advanced and technology continuously developed, deep learning fusion techniques have gradually become the focus of investigation and have been widely applied in this field (Maimaitijiang et al., 2020; Nguyen et al., 2023). However, despite the significant achievements of deep learning, it mostly employs single-level feature fusion methods, in which early or late fusion modes are used at different positions within the neural network to integrate features for various modalities (Wang et al., 2022). Regarding early fusion approaches, they involve integrating information obtained from different sensors at relatively shallow positions within the model architecture. As for late fusion approaches, they involve integrating information from different sensors at deeper levels within the model architecture. However, with these single-level feature fusion approaches, they merely stack and combine the extracted features equivalently, neglecting the full exploitation of the fused information (Wang et al., 2021; Zhou et al., 2021). In order to optimize the feature fusion process of multimodal remote sensing data, researchers have employed various methods. For example, they use cross-channel reconstruction modules for feature fusion (Wu et al., 2022), adopt interleaved perception CNNs to integrate heterogeneous information (Zhang et al., 2022), incorporate the Squeeze-and-Excitation module for adaptive feature fusion (Feng et al., 2019), and design self-guided and cross-guided attention modules for multi-branch feature fusion (Dong et al., 2022). Through this series of improvements, the feature fusion of multimodal data is further strengthened. These studies showcase the potential of deep learning to learn multiple modal features in data fusion applications, thereby overcoming the limitations of learning features solely from a single modality. However, in the context of agricultural monitoring, a specific field or task, there exists a requirement for relevant improvements in current deep learning fusion models to enhance their adaptability. Kong et al. proposed a cross-level fusion strategy based on feature-level Soft-VLAD aggregation and decision-level gaussian probability fusion. This strategy expands the advantages of multi-granularity feature complementarity, utilizing a vast amount of fine-grained information to achieve precise recognition of crop species (Kong et al., 2021). Ma et al. devised a dynamic fusion module with adaptive modality attention adjustment, effectively utilizing multimodal canopy information. This approach significantly enhances the model's precision and its ability to adapt well to various wheat varieties (Ma et al., 2023). These studies indicate that such tailored improvements can enhance the performance and adaptability of the model in agricultural monitoring. Additionally, they demonstrated the feasibility of using a deep learning-based multimodal remote sensing fusion model to describe complex farmland weed competition relationships.

Therefore, this study aims to develop a multimodal deep fusion model (MulDFNet) based on Transformer and multi-layer residual to assess the competition of weeds in farmland ecosystems. The proposed framework employs a three-branch structure comprising HSI, VI, and CHM to acquire spatial, spectral, and elevation characteristics for vegetation canopy. Subsequently, through the Transformer's multi-head attention module, the specific features of each modality are highlighted, intensifying the focus on crucial features. Following this, we introduce a multi-layer residual fusion module to perform shallow level feature fusion for different branches. This fusion module comprehensively considers the interrelation between different modal features, while also balancing subtle feature disparities and more abstract semantic information. Finally, by employing a multi-level hierarchical deep fusion approach, the shallow characteristics are integrated to build a deep characteristics integration module, enhancing the comprehensive understanding and assessment ability of farmland weed competition.

2. Materials and methods

2.1. Field experimental design

From May to September 2021, we conducted experiments in Harbin, China, specifically at the Agricultural Demonstration Base of Northeast Agricultural University (latitude 45°45′, longitude 126°54′). Within this region, the presence of nutrient-rich black soil stands out as a prominent feature. The annual average rainfall ranging from 400 to 600 mm, ensuring a water supply for crops throughout the cycle. In conjunction with this, the annual average effective accumulated temperature of 2800 °C, contributing to an environment conducive to robust plant growth and development.

Seeding took place on May 6, 2021, within the experimental region, utilizing the maize variety Xianghe88. Addition experiments were carried out to assess the competitive interactions between maize and weeds (Swanton et al., 2015). We conducted field weed count surveys to classify weed density into five distinct treatments, denoted as Levels 1 to 5 (also identified as N0 to N4). The weed densities for these treatments are as follows: 0 plants/m^2, 20 plants/m^2, 40 plants/m^2, 80 plants/m^2, and 160 plants/m^2. Three replications were carried out for each treatment, resulting in a total of 15 plots, each covering 3×15 m^2. Neighboring plots were set up with protective strips between them, and weed seedlings within the experimental plots were thinned to mimic the density in the various treatment areas. Specifics of the experimental design were expounded upon in the study of Lou and her team (Quan et al., 2023).

2.2. Data acquisition

2.2.1. Unmanned aerial vehicle data collection

The LR1601-IRIS Lidar sensors (LICA Inc., Beijing, China) and Pika L

Hyperspectral sensor (Resonon Inc., Bozeman, MA, USA) are integrated into the DJI M600 Pro UAV (DJI Inc., Shenzhen, China). Data collection, involving both spectral and lidar, occurred in a single flight campaign. Flight parameters included an altitude of 30 m, speed of 1.0 m/s, and overlap rate of 35 %. Refer to Fig. 1a for the UAV system, Table 1 for sensor parameters, and Fig. 1c for the collection date and stage.

2.2.2. Farmland data collection

The collected farmland data include multiple growth parameters at different developmental stages of maize (Fig. 1d). These parameters encompass both structural attributes (plant height and stalk thickness) and details regarding nutritional content (Nitrogen and Phosphorus elements). Ground measurements were conducted during five pivotal growth stages, namely the three-leaf stage, five-leaf stage, jointing stage, trumpet stage, and flowering stage. The phenotypic changes in maize plants from the three-leaf stage to the flowering stage provide a visual representation of the competition effects. Following the flowering stage, the phenotypic parameters do not vary significantly, and the nutrients begin to gradually transition towards the seeds. For precise information on the collection dates and stages, refer to Fig. 1c.

For each treatment plot, we conducted a measurement of structural parameters by randomly selecting ten maize plants as samples. Using a telescopic ruler and vernier caliper, we measured the natural plant height (PH) and stalk thickness (ST) at the base of these plants (Chukwudi et al., 2021). Furthermore, from each experimental plot, three maize leaves were randomly taken as samples for the determination of nitrogen (N) and phosphorus (P) content. These samples were shredded, blended, and their tissue fluids were extracted, followed by the addition of chemical reagents. Finally, we employed a colorimetric method to determine the nitrogen and phosphorus content in the maize leaf samples (Watt et al., 2020).

2.3. Calculation of CCI-A for multiple growth stages

Through previous research, Lou et al. have demonstrated the feasibility of using the multi-dimensional growth parameters of maize (PH, ST, N, P) for PCA mapping and establishing a Comprehensive Competition Index (CCI-A) to represent the competition intensity of weeds at different stages (Lou et al., 2022; Quan et al., 2023). The calculation of CCI-A is depicted in Eq. (1), with the input being the mapping of the first principal component (PC1) of the multidimensional growth parameters of maize (Damalas and Koutroubas, 2022). Table S1 displays the CCI-A values of per treatment plot across various growth stages.

$$CCI\text{-}A = \left(P_{weed\text{-}free} - P_{weedy}\right) / P_{weed\text{-}free} \quad (1)$$

where: CCI-A signifies comprehensive competition indices, $P_{weed\text{-}free}$ denotes crop performance under weed-free conditions, and P_{weedy} signifies crop performance under weedy conditions.

2.4. Hyperspectral data processing

The spectral images of the study domain (DOM) at different stages were acquired using remote sensing image processing software, with a ground sampling distance (GSD) of 10 cm/pixel for the hyperspectral images.

Research has shown that soil background impacts the overall spectral response of the canopy. To more accurately capture and analyze vegetation characteristics, it is necessary to remove the soil background and extract canopy images. Applying OSAVI improves the distinguishability of vegetation and soil regions in remote sensing images. By selecting a suitable threshold, the soil background can be effectively removed from the DOM image (Chen and Wang, 2022).

Hyperspectral data typically encompass multiple bands, each reflecting distinct spectral features and physical processes. Vegetation indices (VIs) extract comprehensive spectral information by combining and calculating data from different bands, allowing for the assessment of vegetation status. Comprehensive spectral information provides a more comprehensive perspective, allowing us to better understand information about vegetation's photosynthesis, nutritional status, growth stages, and vegetation types. With the advancement of remote sensing technology, numerous VIs have been developed to estimate crop phenotypic parameters. In this study, 29 VIs (Table S2) with strong correlations to vegetation structural indicators, physiological and biochemical features, and crop yield parameters were selected for the research (Almeida et al., 2021; Du et al., 2018; Liu et al., 2021).

This study utilized 150 consecutive bands of hyperspectral data and 29 vegetation indices as canopy spectral features. However, the high similarity among these features leads to information redundancy, which could potentially impact the accuracy and timeliness of subsequent analyses. To address this issue, we computed the average spectrum of all bands and VIs for each sample and employed the successive projections algorithm (SPA) to select hyperspectral bands and vegetation indices relevant to weed competition (Xia et al., 2023). This algorithm can identify informative variables and minimize collinear effects among variables to the greatest extent. Previous research has applied SPA to choose effective features for various agricultural tasks, including yield prediction, pest and disease recognition, growth monitoring, and soil analysis, among others. The sensitive features chosen by the SPA algorithm are assessed based on their RMSE.

2.5. Lidar data processing

We obtained images of canopy structure across the study area by reconstructing the point cloud using ENVI-Lidar software.

We further extracted valuable canopy structure data, which encompassed both the digital surface model (DSM) and digital elevation model (DEM). In addition to ground elevation, the DSM also includes additional surface characteristics beyond the DEM. Meanwhile, the canopy height model (CHM) is derived by parsing the discrepancy between DSM and DEM, and the resolution mirrors that of the aforementioned models. Functioning as a representation of vertical vegetation height relative to the ground, the CHM offers insights into both the horizontal and vertical canopy distribution (Alonzo et al., 2020). Both DSM and DEM have a GSD of 25 cm/pixel.

2.6. Dataset construction for multimodal data

The model dataset was constructed by segmenting preprocessed time-series remote sensing data of HSI, VI, and CHM into image patches. Because of the variation in resolution of the different modal data, the image patch dimensions for HSI and VI were set at 10 × 10 pixels, whereas the image patch size for CHM was 4 × 4 pixels. The multimodal dataset contains HSI, VI, and CHM image patches and the corresponding CCI-A labels. The remote sensing data of each modality (HSI, VI, CHM) from the five periods were subdivided into 1855 image patches. Subsequently, using a random partitioning approach, they were allocated to the training and testing sets in an 8:2 ratio.

3. Modeling framework

3.1. Overall framework

Fig. 2 displays the overall framework of the proposed multimodal deep fusion model (MulDFNet) based on Transformer and multi-layer residual. The framework consists of five stages: multimodal feature extraction stage, Transformer encoding stage, shallow feature fusion stage, deep feature fusion stage, and regression prediction stage.

3.2. Multilayer CNN feature extraction backbone

In order to extract features from HSI, VI, CHM data, we employ a

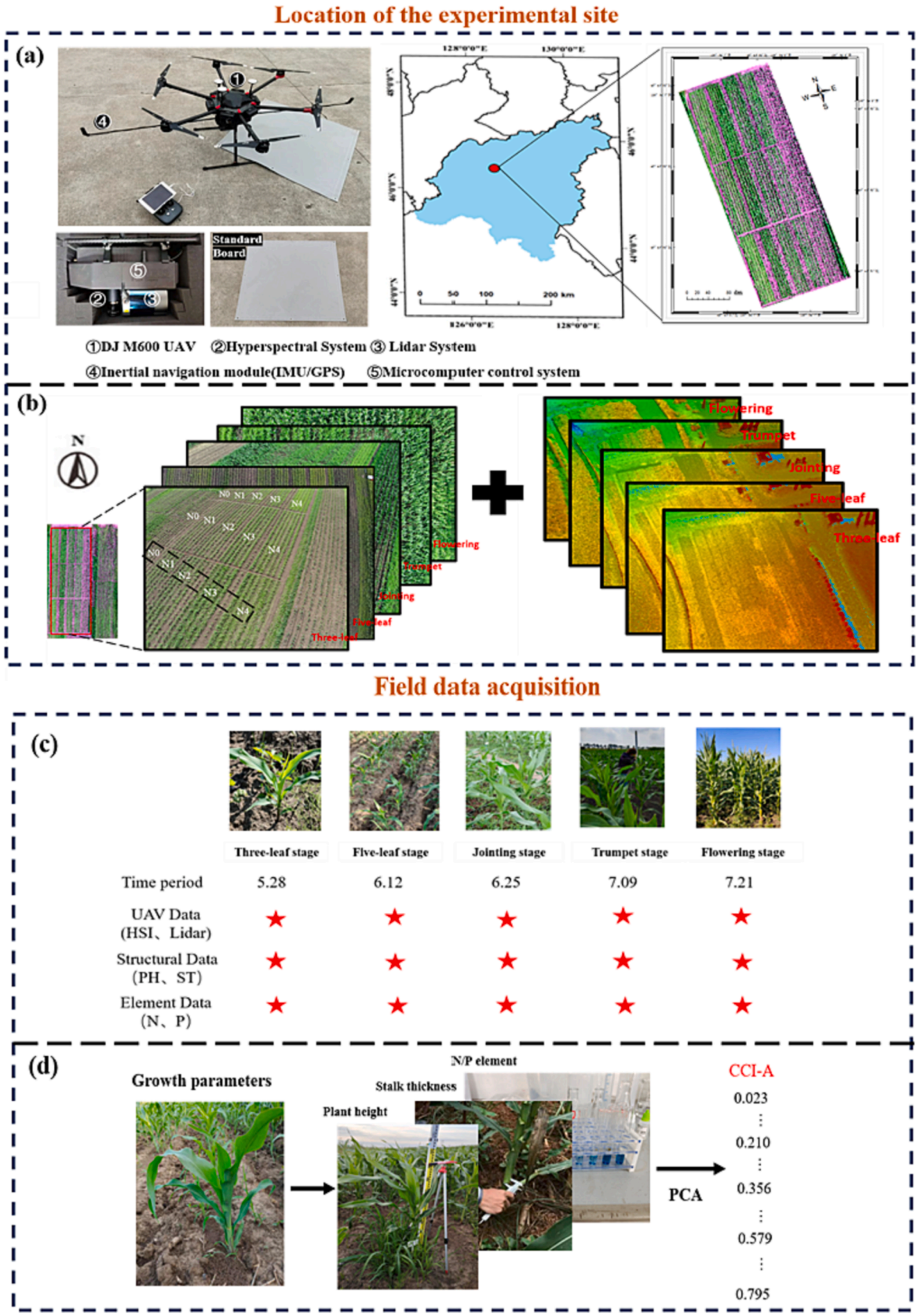


Fig. 1. Field experimental design. (a) Experimental site; (b) UAV image acquisition; (c) Data collection time for each growth stage; (d) Farmland data collection and CCI-A index creation.

Table 1
Main technical parameters of spectral and lidar sensors.

Pika L Hyperspectral		LR1601 Lidar	
Spectral range	388–1030 nm	Pulse repetition frequency	5–20 Hz
Number of spectral channels	150	Laser wavelength	905 nm
Field of view angle	17.6∘	Returns per pulse	2
Focal length	17 mm	Point density	200–800 points/m^2

three-branch convolutional neural network as the foundational architecture of the fusion model. The detailed network structure for feature extraction is depicted in Fig. 3.

The HSI and VI network branches utilize a combination of 3D and 2D convolutions to obtain spatial and spectral features of the vegetation canopy. The HSI and VI branches consist of 3D convolutional layers, 3D normalization layers, 2D convolutional layers, 2D normalization layers, and activation layers. The ultimate output HSI and VI feature is denoted as T_{HSI} and T_{VI}.

The network architecture of the CHM branch is similar to that of the HSI and VI. The CHM data undergoes 2D convolution to extract structural features of the vegetation canopy. This branch consists of 2D convolutional layers, 2D normalization layers, and activation layers. The ultimate output CHM features are denoted as T_{CHM}. Taking into account the differences in data dimensions and sizes between HSI, VI, and CHM, we adjusted the number of convolutional modules and kernel sizes to obtain feature maps of the same size.

3.3. Transformer Encoder

To use the feature maps generated by the multilayer CNN feature extraction backbone as input, it's necessary to divide them into N equally sized patches for feature serialization. This is done to capture the relationships between elements in the sequence. The module mainly

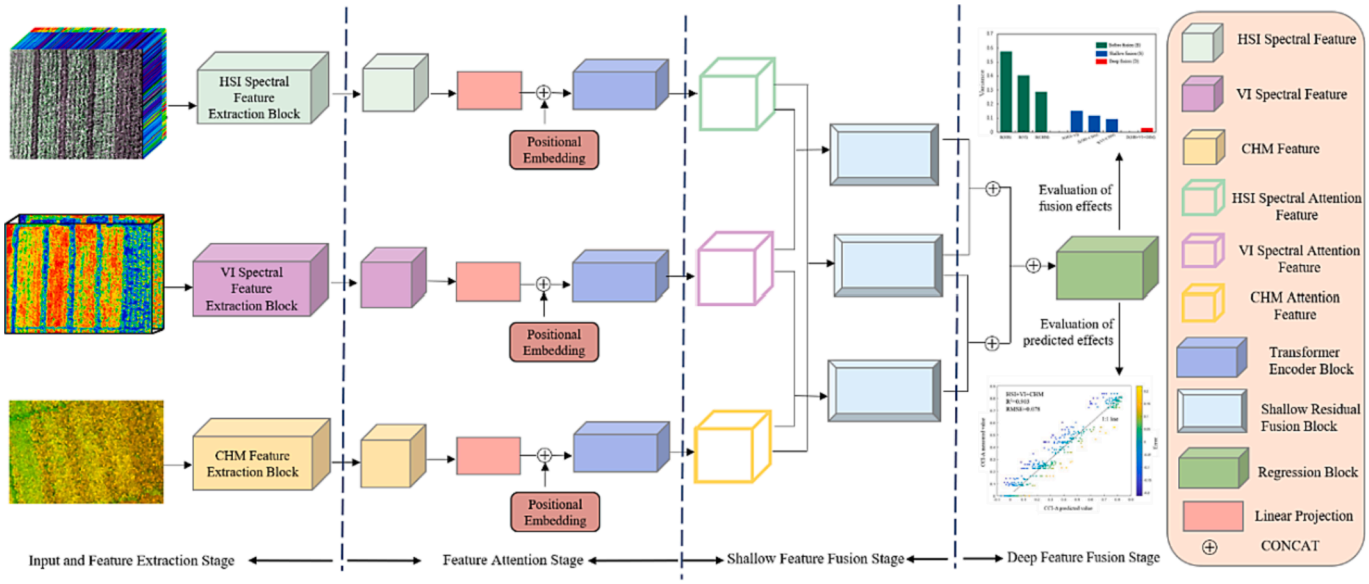


Fig. 2. Overall architecture of the $MulDFNet_{(HSI\text{-}VI\text{-}CHM)}$ model.

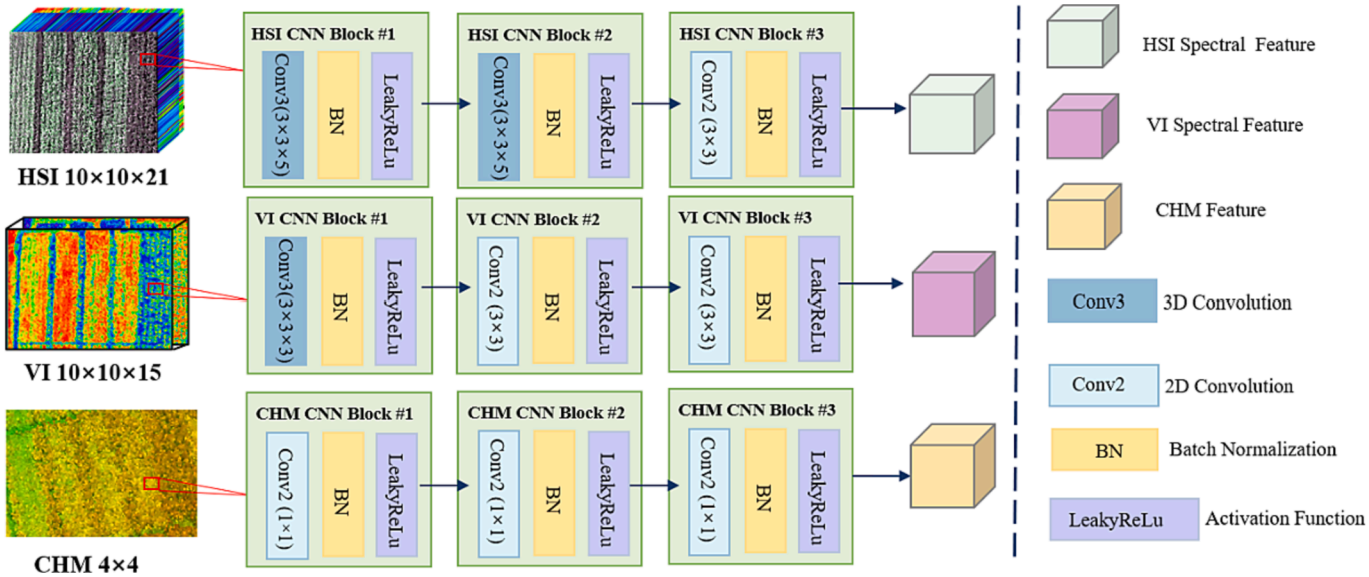


Fig. 3. The multimodal data feature extraction backbone of the fusion model.

includes two subcomponents.

In its first component, patches are mapped to a D-dimensional embedding space using linear transformations. This linear transformation is implemented by a trainable matrix, where D is a latent hyperparameter that affects the model's parameter number. To introduce the positional information of elements into the model, a position embedding is added to each patch. The final input sequence is:

$$Z_0 = \left[x_p^1 E; x_p^2 E; \cdots; x_p^N E\right] + E_{pos} \tag{2}$$

where: x_p^i represents the i-th patch, E is the input embedding matrix, E_{pos} represents the positional embedding matrix.

The second crucial subpart is the Transformer Encoder (TE), which has the capability to capture intricate relationships among elements within a sequence. The encoding segment comprises multiple consecutive Transformer Encoders, with the output from the former encoder becoming the input of the next. Each encoder comprises Multi-Head Self-Attention layer (MSA), Multi-Layer Perceptron (MLP), and two Layer Normalization (LN) layers (Qing et al., 2021). Residual connections are established before the MSA and MLP layers (denoted as "⊕" in the diagram).

TE performs well due to its core MSA mechanism, as shown in Fig. 4. The Self-Attention (SA) mechanism within it has the capacity to dynamically adjust the weighting of each element within the input sequence, thus highlighting crucial information. Specifically, the encoded sequence information is converted to the Q, K, and V matrices by three independent linear transformations. These linear transformations are achieved through three trainable weight matrices, W_Q, W_K, and W_V. Attention scores are computed through the dot product between Q and K, and then the softmax function is applied to compute the weights of these scores. The formula for SA is as follows:

$$SA = Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)V \tag{3}$$

The MSA comprises multiple sets of weight matrices for transforming Q, K, and V. The identical computational procedure is applied to calculate multi-head attention values. The results from every attention head are then linked together. The process is represented with the following formula:

$$MSA(Q, K, V) = Concat(SA_1, SA_2, \cdots, SA_h)W \tag{4}$$

In the subsequent step, the weight matrices obtained in the prior phase are introduced into the MLP layer. The MLP comprises two fully connected layers, and between these two layers, the Gaussian Error Linear Unit (GELU) is applied as a non-linear activation function. The formula for this process is as follows:

$$MLP(x) = GELU(xW_1 + b_1)W_2 + b_2 \tag{5}$$

The output form of the feature information after being encoded by the encoder is N sets of D-dimensional matrices, where the input size (T_{in}) is identical to the output size (F_{out}). In order to input this information into the next encoder, it needs to be concatenated into a matrix and its shape be reorganized.

3.4. Shallow fusion module for multilayer residuals

Inspired by the workings of ResNet, we devised a shallow fusion module with multiple layers of residual connections to achieve the interaction and fusion of intermediate feature maps from diverse sensor information. Compared to directly training fusion layers to match the required underlying mapping, we enable the fusion layer to fit residual mappings. This enables the fusion layer to better capture subtle feature differences and correlations, thereby enhancing the model's expressiveness.

As depicted in Fig. 5, we have developed a shallow fusion module with three layers of residual connections, enabling shallow fusion through pairwise cross-combination of intermediate feature maps from three distinct modal data. Each underlying fusion mapping is respectively represented as $h(F^{HSI}, F^{VI})$, $h(F^{HSI}, F^{CHM})$, and $h(F^{VI}, F^{CHM})$, where F^{HSI}, F^{VI}, and F^{CHM} are feature maps generated from the encoder modules. Taking the F^{HSI} and F^{VI} fusion in the first layer as an example, we first concatenate F^{HSI} and F^{VI} along the depth dimension. Then, the channel dimension of the features is decreased using a 1×1 convolutional layer and the features are fused using a 3×3 convolutional layer. Applying a skip-connection F^{HSI} and F^{VI} enables the training of the aforementioned non-linear fusion layer to fit the mapping in equation

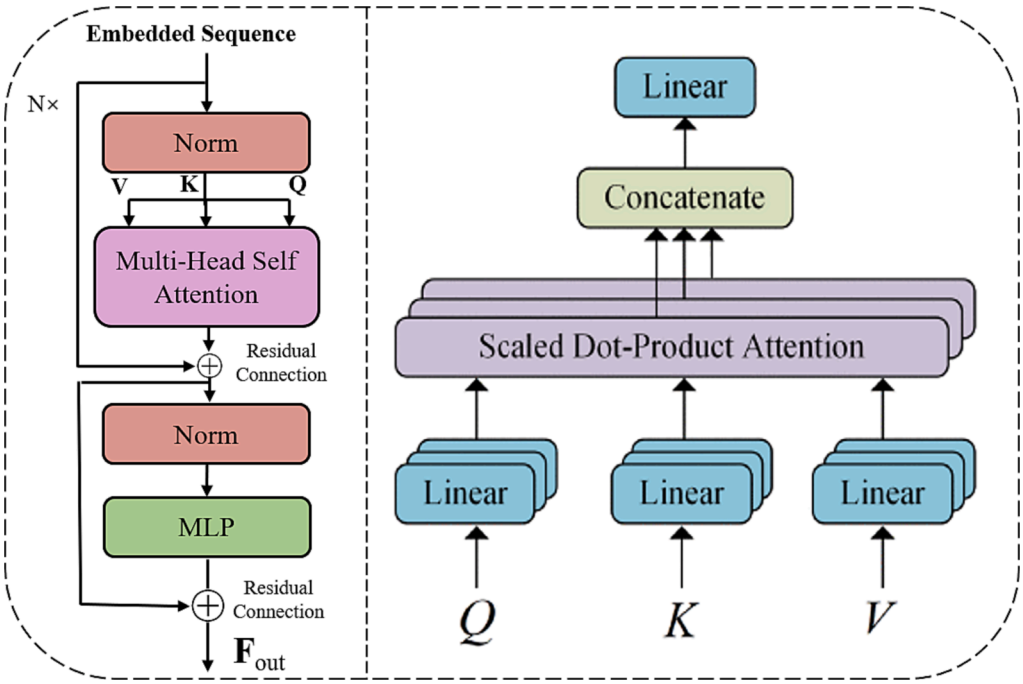


Fig. 4. Graphical illustration of the Transformer Encoder and MSA module.

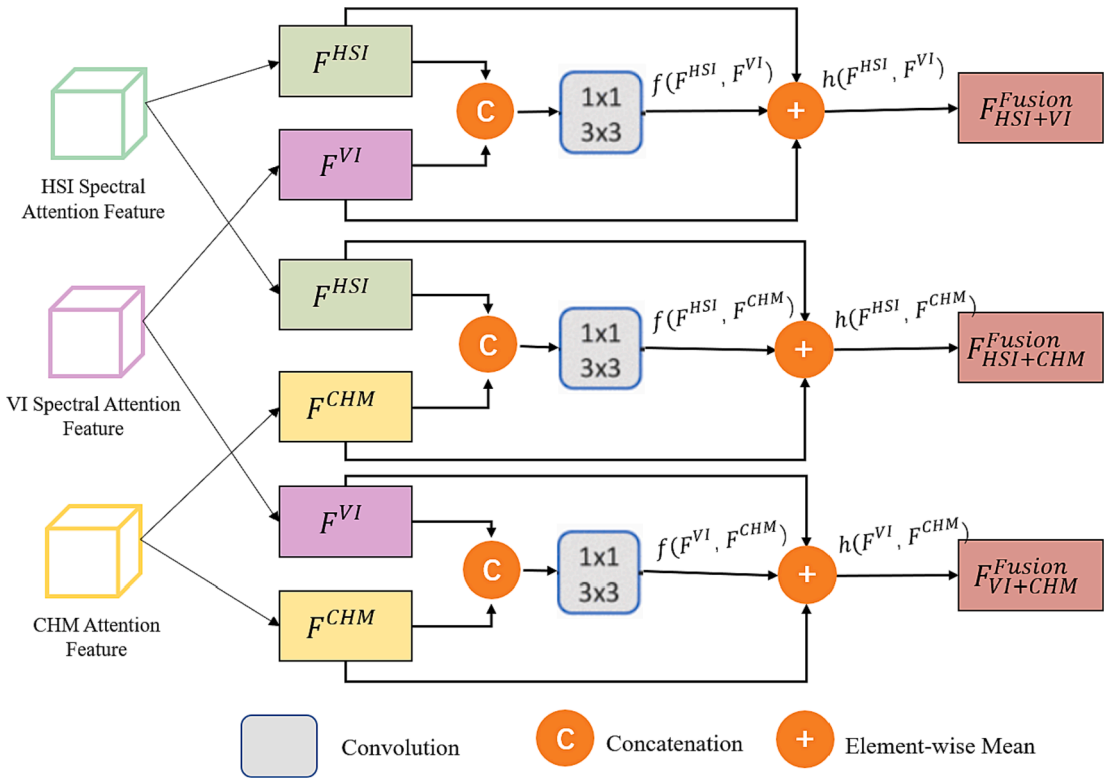


Fig. 5. The architecture of the shallow fusion module for multilayer residuals.

(6).

$$f(F^{HSI}, F^{VI}) = h(F^{HSI}, F^{VI}) - F^{HSI} - F^{VI} \tag{6}$$

3.5. Deep feature fusion module

After the fusion of multiple layers of residuals, we get various shallow fusion features, including HSI and VI fusion feature F^{Fusion}_{HSI+VI}, HSI and CHM fusion feature $F^{Fusion}_{HSI+CHM}$, and VI and CHM fusion feature F^{Fusion}_{VI+CHM}. Next, the obtained shallow fusion features are fed back into the three-branch convolutional neural network to acquire features S^{Fusion}_{HSI+VI}, $S^{Fusion}_{HSI+CHM}$, and S^{Fusion}_{VI+CHM}, which are then used for deep feature fusion (Fig. 6). We define the process of deep fusion as:

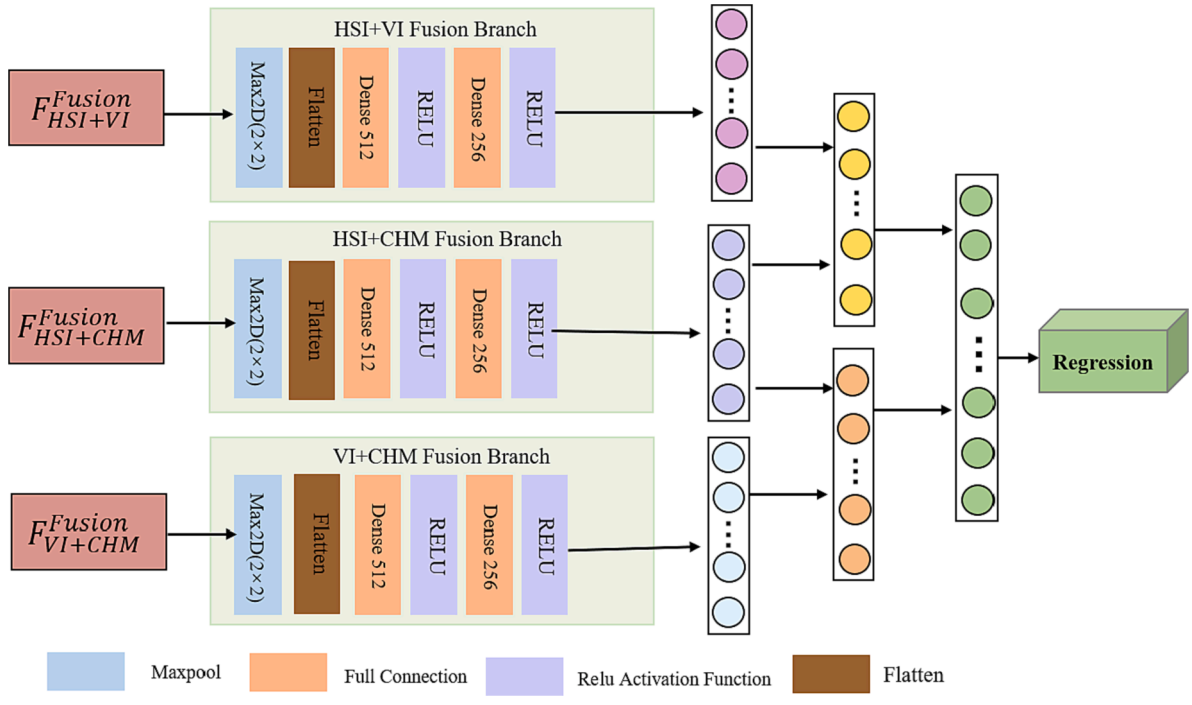


Fig. 6. The architecture of the deep feature fusion module.

$$S_M = f(W \bullet (S^{Fusion}_{HSI+VI} \| S^{Fusion}_{HSI+CHM} \| S^{Fusion}_{VI+CHM}) + b) \quad (7)$$

where: || denotes concatenation, W represents weights, and b represents the bias of the fully connected. Subsequently, the ultimately obtained deep fusion feature S_M was input to the Regression Module for weed competition index regression.

3.6. Network training

We use Mean Squared Error (MSE) as the loss function for the regression model, designed to impose a strong penalty on large errors, thereby enhancing the model's robustness to outliers. MSE measures the model's performance by calculating the average of the squared differences between predicted values and true values, offering both intuitiveness and ease of optimization. During the training process, we employ the Adam optimization algorithm and adjust optimization using a learning rate of 0.001 to ensure the convergence speed and stability of the training. Additionally, we incorporate techniques such as batch normalization and dropout regularization to further enhance the model's generalization capabilities. The training was performed for 150 epochs, employing a batch size of 64. R^2 and RMSE were computed to serve as assessment metrics for gauging the efficacy of the model.

The model training was carried out on a Windows 10 computer equipped with an Intel Core i5-10300H CPU, 16 GB of RAM, and an NVIDIA GeForce RTX 1650 GPU. The model programs have been developed using Python 3.7.4, while the network architecture has been built using PyTorch 1.9.0 + cu102.

4. Results and discussion

4.1. Spectral band and vegetation index selection

4.1.1. Spectral band selection

This study employs the SPA algorithm for spectral band selection. The best number of spectral variables was identified according to the predicted RMSE values. Fig. 7a illustrates the variation of RMSE values as the quantity of spectral bands. When the quantity of bands is 21, the minimum RMSE is 0.141. The distribution of the chosen 21 sensitive spectral bands is broad, covering multiple spectral regions, including the blue-violet, green, red, and near-infrared. The specific bands are 425 nm, 434 nm, 438 nm, 442 nm, 483 nm, 516 nm, 529 nm, 558 nm, 626 nm, 677 nm, 694 nm, 716 nm, 724 nm, 733 nm, 755 nm, 759 nm, 763 nm, 922 nm, 927 nm, 1003 nm, 1008 nm.

4.1.2. Vegetation index selection

We investigated the correlation between vegetation indices and weed competition indices, as depicted in Fig. 8. The analysis indicates a pronounced correlation among various vegetation indices, and certain vegetation indices display a strong correlation with competition indices, such as GNDVI, NDVI, SR2, and VOG2. These vegetation indices include the expression of parameters such as color, pigment, substance, structure, etc. This indicates that weed competition influences these

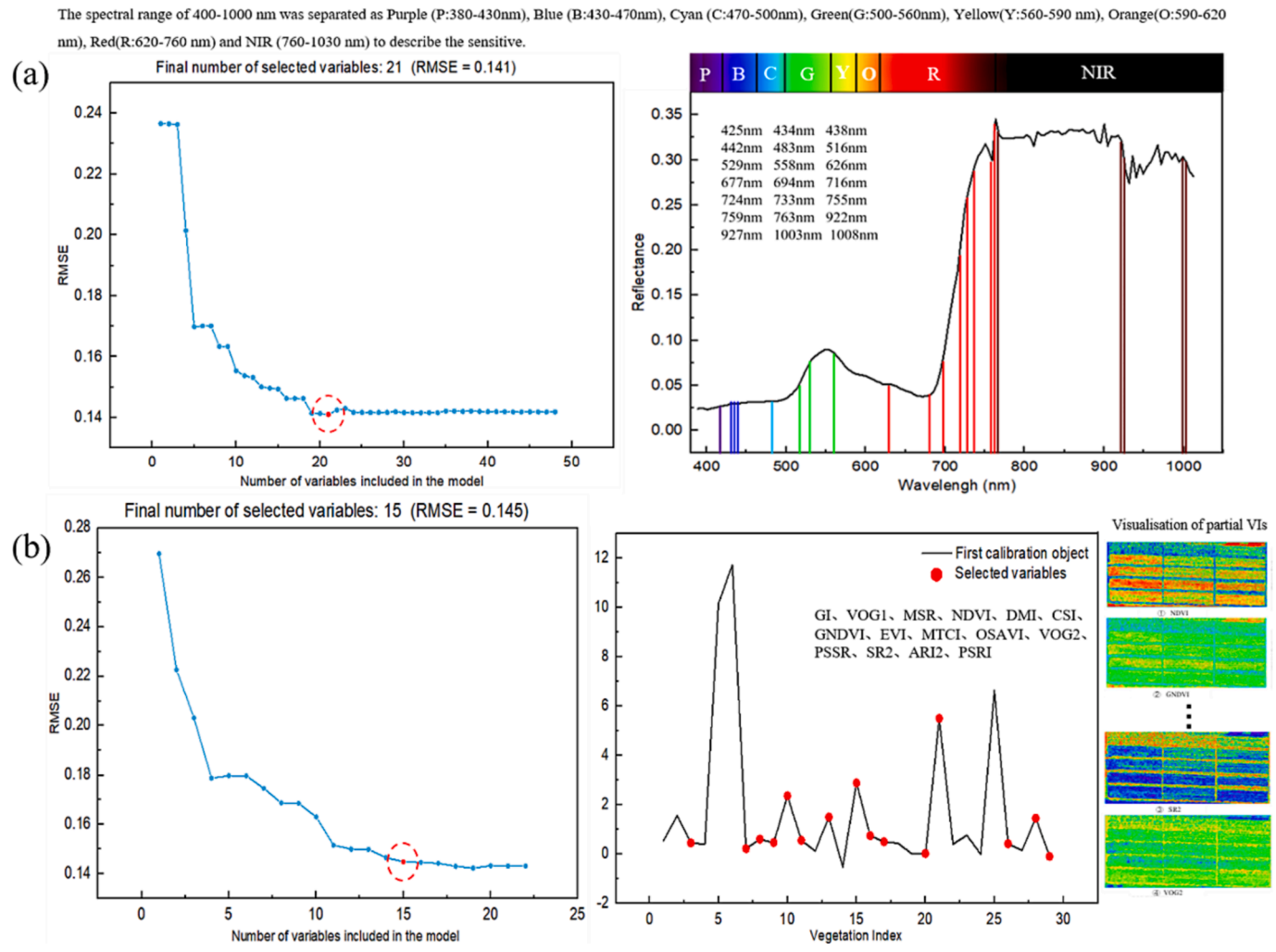


Fig. 7. Spectral band and vegetation index screening. (a) Sensitive spectral band selection; (b) Vegetation index selection.

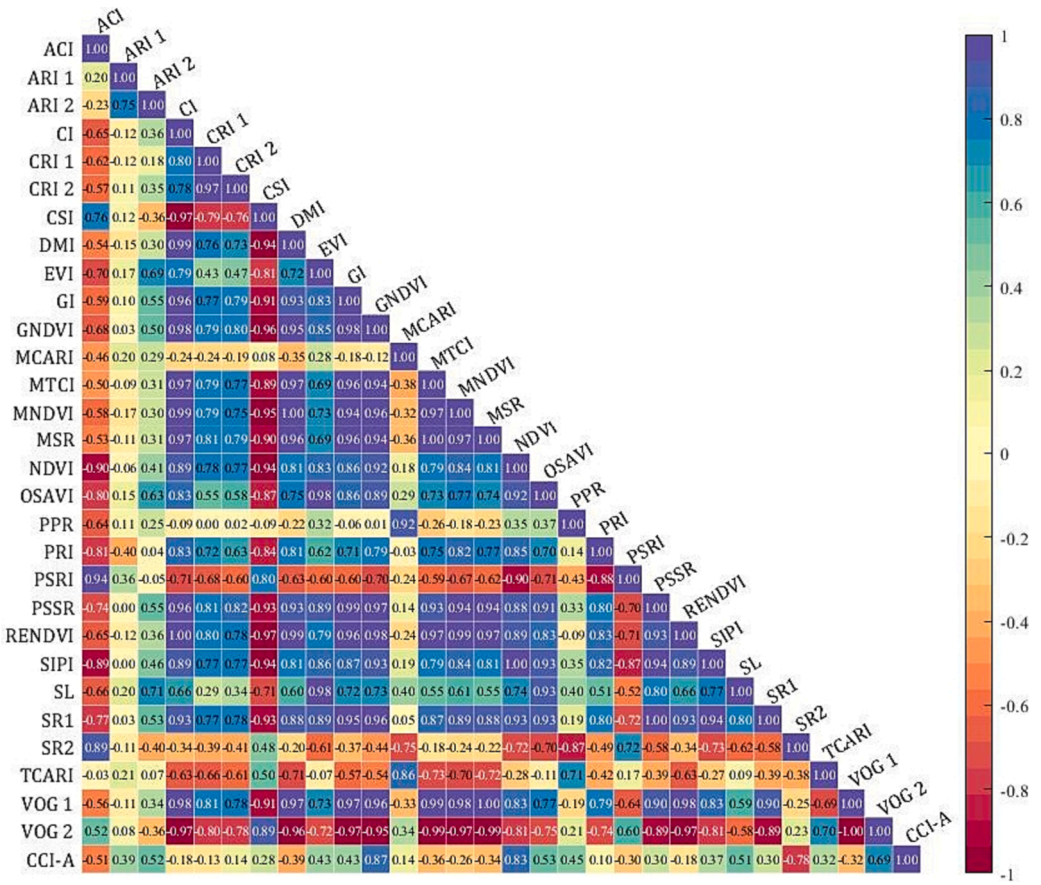


Fig. 8. Correlation between vegetation indices and weed competition indices.

parameters in the canopy. Therefore, the utilization of comprehensive spectral information from vegetation indices holds promising feasibility in quantifying weed competition intensity.

Through the preceding correlation analysis of the 29 vegetation indices, we observed significant correlations and a substantial amount of redundant information among them. Consequently, this study employs the SPA algorithm to select vegetation indices, thereby eliminating redundancy and enhancing model stability and predictability. Fig. 7b illustrates the variation of RMSE values with different numbers of vegetation indices. When the number of vegetation indices is 15, the minimum RMSE is 0.145. These 15 vegetation indices include GI, VOG1, MSR, NDVI, DMI, CSI, GNDVI, EVI, MTCI, OSAVI, VOG2, PSSR, SR2, ARI2, PSRI.

4.2. Analysis of the multimodal deep fusion model's performance

4.2.1. MulDFNet$_{(HSI\text{-}VI\text{-}CHM)}$ model prediction effect

The performance evaluation of the MulDFNet model, containing 37 features (21 spectral bands, 15 VIs, and 1 CHM), was conducted. The regression accuracy of the training set's CCI-A indices is $R^2 = 0.968$ (RMSE = 0.043). For the test set, the regression accuracy of the CCI-A indices is $R^2 = 0.903$ (RMSE = 0.078), with R^2 values for five periods being 0.201, 0.902, 0.891, 0.878, and 0.851, as shown in Fig. 9. The regression effect of the model is poorest during the three-leaf stage. This is because at this phase, weed competition is only in its initial stage, and the distinction in vegetation canopy within various competition levels is not yet pronounced. The model's inability to extract sufficient features expressing competition is causing a relatively poor regression effect. In the five-leaf stage, the model exhibits the best regression effect. During this stage, significant differences in competition arise, leading to adaptive changes in the vegetation canopy layers across various levels of competition. This results in the model's regression effect being notably superior to that of the three-leaf stage. At the jointing, trumpet, and flowering stages, the regression effect showed a slight decrease, but the R^2 values remain above 0.85. Overall, the multimodal deep fusion model demonstrates good consistency between the predicted CCI-A and the actually calculated CCI-A during the last four stages.

4.2.2. MulDFNet $_{(HSI\text{-}VI\text{-}CHM)}$ model fusion data analysis

By employing data visualization techniques, the similarity and dissimilarity of feature data before and after fusion are compared. In the process of data fusion, preserving the structural characteristics of the data is particularly crucial. The preservation of structural features refers to the ability of fused data to retain the essential structural attributes of the original data, maintaining similarity or consistency in certain aspects with the original data. Through the data frequency graph in Fig. 10b, it is observed that the HSI, VI, and CHM feature data tend to follow a normal distribution in terms of data structure before and after fusion. The shallow and deep fusion modules have preserved and optimized the original data structure. Variance is a statistical metric that measures the stability of data. Through the data distribution and variance in Fig. 10a, it is observed that compared to the original data, the data distribution after both shallow and deep fusion is more stable, with the variance gradually decreasing. This indicates that data fusion might have facilitated the sharing or overlapping of feature information from HSI, VI, and CHM, resulting in the fused data becoming more stable with lesser variations. This implies that the fusion modules possibly reduced information redundancy and duplication to some extent, making the data more compact. The convolution within the fusion module performs weighted fusion of data from different branches, amplifying the influence of certain modal data, thereby reducing the overall variance. Overall, data fusion maintains the structural characteristics of HSI, VI,

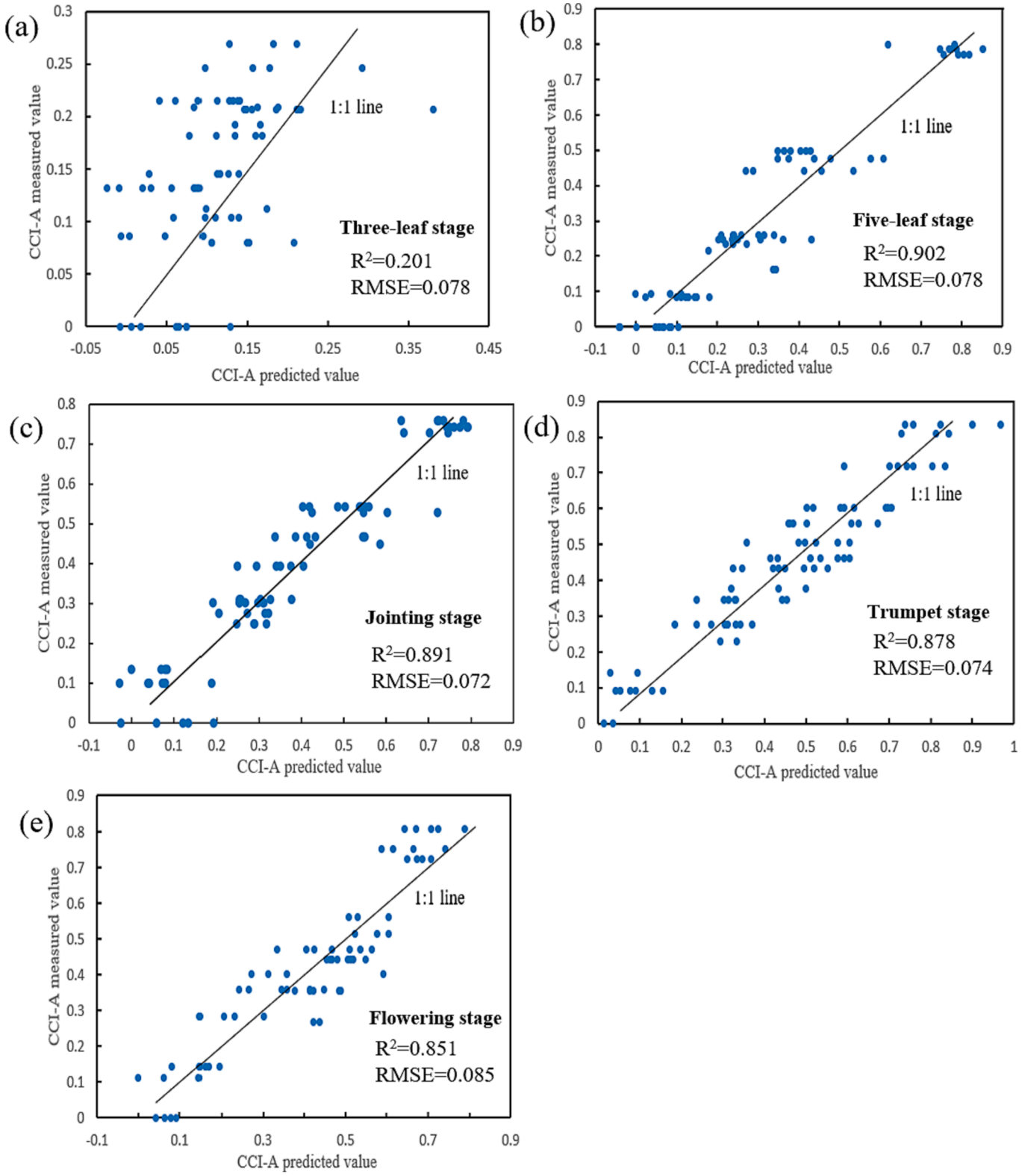


Fig. 9. Predictive effectiveness of multimodal deep fusion models at various phases. (a) three-leaf stage; (b) five-leaf stage; (c) jointing stage; (d) trumpet stage; (e) flowering stage.

and CHM data while also enhancing data quality.

4.3. Ablation study of the multimodal deep fusion model

4.3.1. Ablation analysis of different modal data

Taking into account the influence of different modal data on model's performance, seven sets of cross-combination experiments were

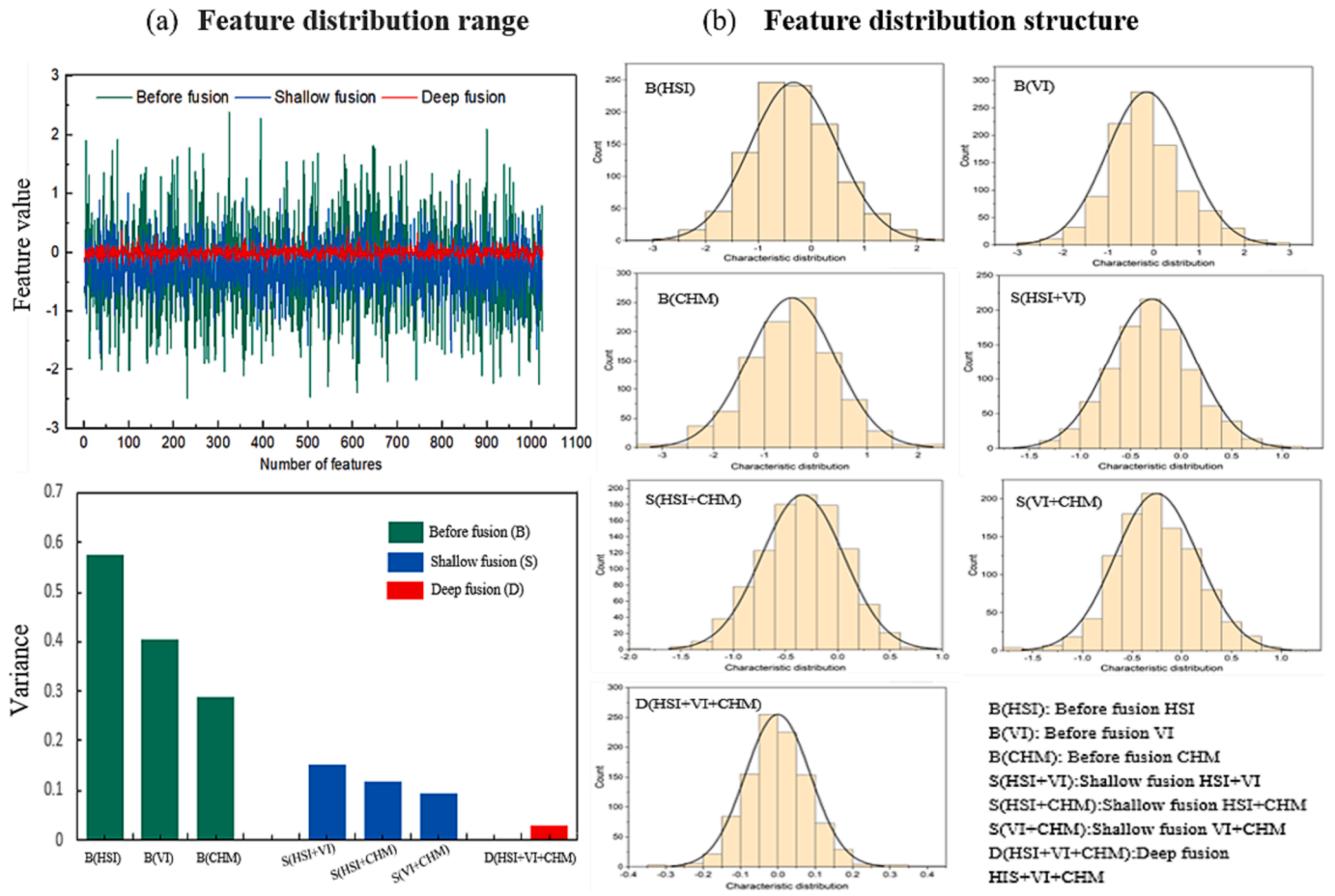


Fig. 10. Analysis of similarity and dissimilarity of feature data before and after fusion. (a) Feature distribution range; (b) Feature distribution structure.

conducted for HSI, VI, and CHM data using single-modal, two-modal, and three-modal data input approaches. The experimental outcomes are depicted in Table 2. The results indicate that the single-modal data model has an average R^2 of 0.705, the two-modal data model has an average R^2 of 0.821, and the three-modal data model demonstrates optimal predictive performance with an R^2 of 0.903 (RMSE = 0.078). For HSI data, both single-modal and multi-modal data models have achieved favorable results, fully showcasing the superior performance of hyperspectral in plant stress assessment. For CHM data, the effectiveness of the single-modal CHM data model is the weakest, as it can only represent certain structural features of the canopy. However, the fusion of CHM and HSI produces the best results in the two-modal data model with an R^2 of 0.856. For VI data, it reflects the combined information from multiple spectral bands, which leads to its performance being better than CHM but slightly lower than HSI. However, the fusion of VI and HSI generates asymptotic saturation issues in spectral features, leading to a slightly less effective outcome compared to the fusion of HSI and CHM.

Overall, in weed competition prediction, the fusion of multiple modal data leads to better results and outperforms any single modal data of HSI, VI, or CHM. These findings are consistent with prior research, demonstrating that the integration of canopy spectral, vegetation indices, and structural information offers distinctive and supplementary insights for plant stress evaluation (Fei et al., 2023; Sun et al., 2022).

Table 2
Precision analysis of different modal data models.

Feature type	Feature num	R^2	RMSE
HSI + VI + CHM	21 + 15 + 1	0.903	0.078
HSI + CHM	21 + 1	0.856	0.092
HSI + VI	21 + 15	0.811	0.106
VI + CHM	15 + 1	0.798	0.112
HSI	21	0.779	0.114
CHM	1	0.616	0.150
VI	15	0.722	0.130

4.3.2. *Ablation analysis of the Transformer Encoder module*

TE is a key component in the proposed fusion framework, which is critical to improving the quality of the fused data. Fig. 11 illustrates the comparison of predictive performance with/without the TE module in different modal data models. Among various modal data combination models, the HSI, VI, and CHM fusion model with TE module attained the highest R^2 value of 0.903. This indicates that the TE module effectively utilized the synergistic information from these diverse inputs, resulting in a significant enhancement of accuracy in vegetation analysis. Furthermore, even when certain features are excluded, the TE module continues to have a positive impact on the model. For instance, when the TE module is integrated, models such as HSI + CHM, HSI + VI, and VI + CHM exhibit significant improvements in R^2 values, with 0.856, 0.811, and 0.798, respectively. Interestingly, even single-modal data models for HSI, VI, and CHM benefit from the integration of the TE module. When comparing the R^2 values of models without the TE module, it further emphasizes the contribution of the TE module to enhancing model performance. In all modal data combination scenarios, the models lacking the TE module attain R^2 values lower than those of the models enhanced with the TE module.

This observation signifies that in weed competition assessment, the TE module has a positive impact on the model's performance. The multi-head attention mechanism within the TE module aids the model in better capturing the relationships and significance between remote

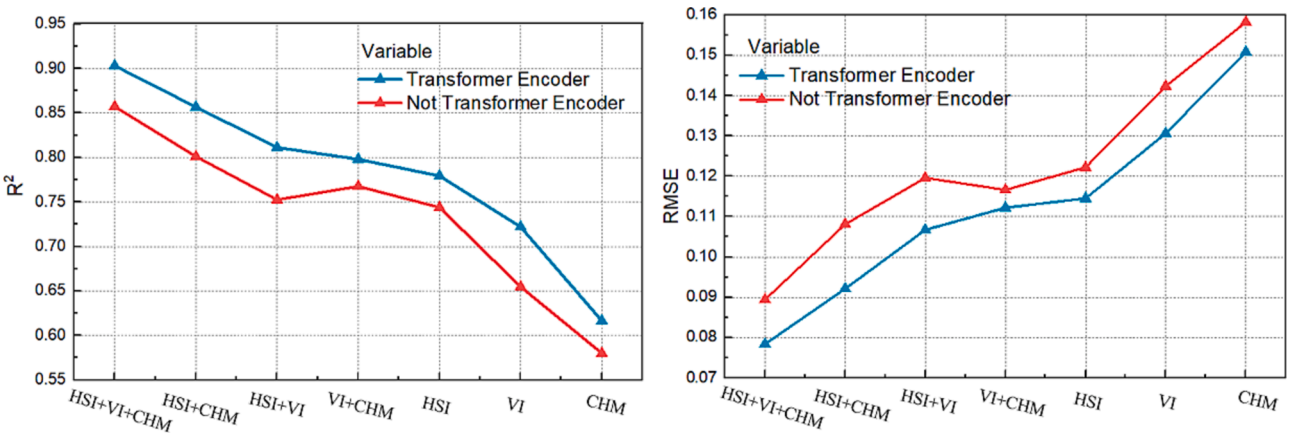


Fig. 11. The impact of the Transformer Encoder module with/without on the model's performance.

sensing imagery and competition indices, thereby enhancing predictive performance.

4.3.3. Ablation analysis of the fusion module

To fully demonstrate the validity of the two fusion modules in the proposed model framework, we conducted ablation experiments on them. The experiments involved four types of combined data and explored three variations of the model framework with different fusion module combinations: deep fusion module only, shallow fusion module only, and a combination of both shallow and deep fusion modules. The experimental results are shown in Fig. 12. The outcomes indicate that the model based on the deep fusion module has an average R^2 of 0.721, while the model based on the shallow fusion module has an average R^2 of 0.733. Regardless of whether it's three-modal or two-modal data input, the performance of models with single deep or shallow fusion modules is relatively poor, with R^2 ranging from 0.78 to 0.68. The model with the combination of shallow and deep fusion modules proposed in this study achieved an average R^2 of 0.842, which is an improvement of 0.121 compared to the single deep fusion module and an improvement of 0.109 compared to the single shallow fusion module. Whether it is a three-modal or two-modal data input, this model shows significant improvement. Overall, the results of ablation experiments with different fusion modules further validate the validity of the proposed model, especially the combined shallow and deep fusion module.

4.4. Comparative analysis of different model

4.4.1. Comparative analysis of different fusion strategies

To prove the superiority of the proposed multimodal deep fusion model (MulDFNet). We compare its prediction effects with a multi-branch convolutional model using early/late stacked fusion. Fig. 13 illustrates the model framework of the early/late stacked fusion network. Fig. 14 illustrates the comparison of predictive effects among different fusion models. The results indicate that the MulDFNet fusion model shows a clear advantage in R^2 and RMSE metrics, especially achieving the best predictive effect of $R^2 = 0.903$ for the three-modal fusion of HSI, VI, and CHM. Compared to the early stacked fusion strategy, this model exhibited an average R^2 improvement of 0.176 for two-modal fusion and an R^2 improvement of 0.179 for three-modal fusion. In comparison with the late stacked fusion strategy, this model demonstrated an average R^2 improvement of 0.167 for two-modal fusion and an R^2 improvement of 0.171 for three-modal fusion. Overall, whether it's two-modal or three-modal data fusion, the multimodal deep fusion model (MulDFNet) achieved favorable predictive effects compared to the early/late stacked fusion.

4.4.2. Comparative analysis of MulDFNet model and other models

To comprehensively assess the effectiveness of the MulDFNet fusion model, we referenced two deep learning models previously employed in studies and applied them to the weed competition prediction task in our research. These two prior studies include the Multi-channel

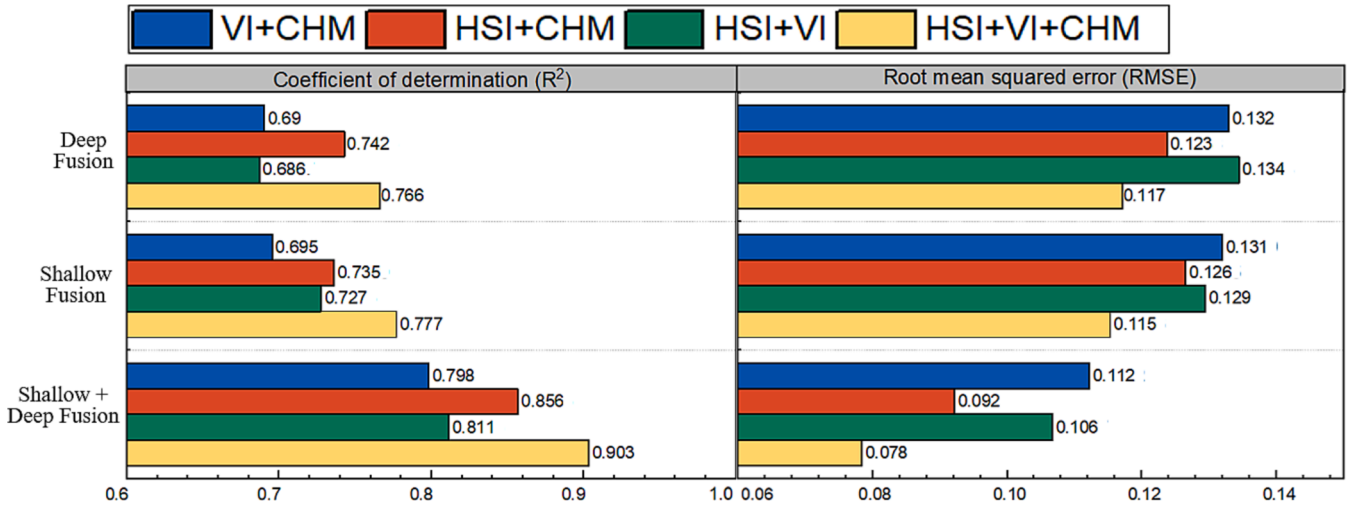


Fig. 12. The effect of different fusion modules on the model's accuracy.

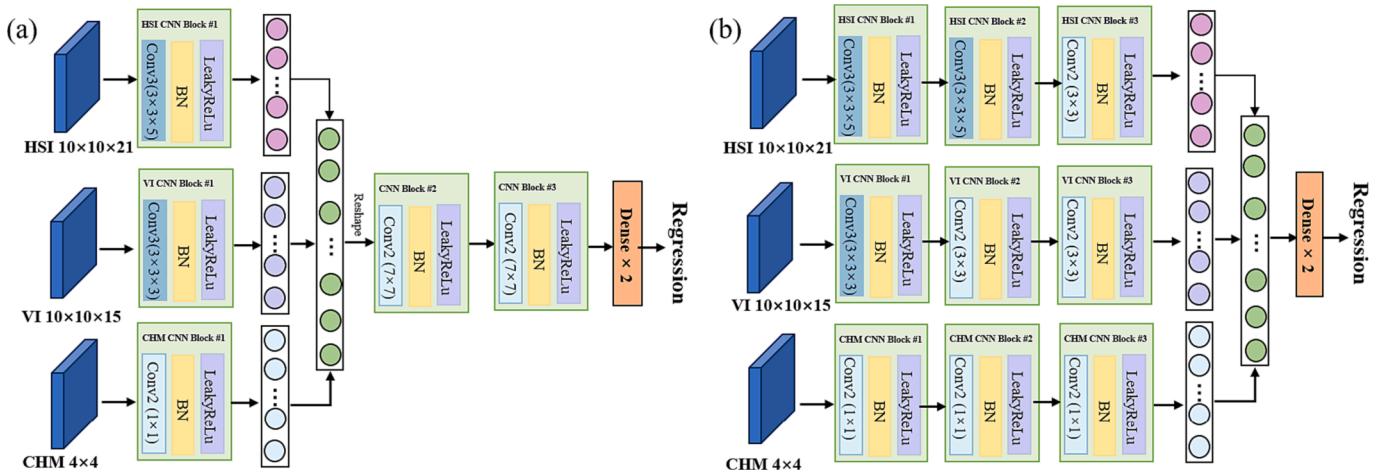


Fig. 13. Network architectures of conventional fusion strategies. (a) Early stacked fusion; (b) Late stacked fusion.

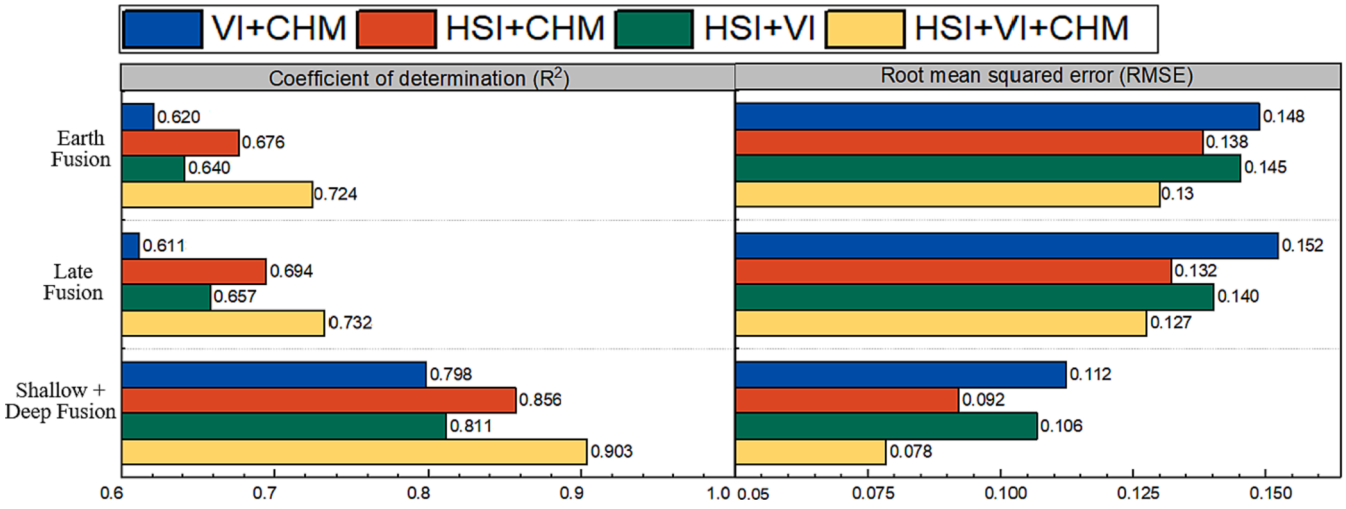


Fig. 14. Comparison of the effects of models with different fusion strategies.

convolutional neural network model established by Nguyen et al., focusing on the fusion of hyperspectral, thermal, and LiDAR remote sensing data for predicting high-throughput phenotypes of corn (Nguyen et al., 2023). The second study is the intermediate-level feature fusion model based DNN(DNN-F2) developed by Maimaitijiang et al., targeting the fusion of multi-modal data, including canopy spectra, structure, thermal, and texture information for predicting soybean yield (Maimaitijiang et al., 2020). Additionally, we conducted comparative analyses with three traditional machine learning models: RF, SVR, and PLS. These models have demonstrated good performance in previous studies (Almeida et al., 2021). The results are presented in Table S3.

In the combination of HSI+VI+CHM data, the MulDFNet model demonstrated outstanding performance, achieving an R^2 of 0.903 in prediction. In comparison, the R^2 values for the DNN-F2 feature fusion model and the Multi-channel CNN model are 0.832 and 0.815, respectively. The R^2 values for the traditional machine learning models RF, SVR, and PLS are 0.845, 0.825, and 0.800, respectively. This further highlights the exceptional performance of MulDFNet model in the fusion of multimodal data. In other data combinations, the MulDFNet model continues to maintain a relatively high predictive performance, with an average R^2 of 0.821. In contrast, the average R^2 values for the DNN-F2 feature fusion model and the Multi-channel CNN model are 0.737 and 0.743, respectively. Among the traditional machine learning models, the RF model performs the best with an average R^2 of 0.783, followed by SVR with an average R^2 of 0.757, and PLS shows relatively poorer performance with an average R^2 of 0.718. Notably, under single-modal data conditions, the performance of the MulDFNet model still maintains a relatively high level compared to other models. This clearly demonstrates the superiority of the MulDFNet model and its adaptability to different modal data.

Although the Multi-channel CNN model and DNN-F2 feature fusion model established by Nguyen et al. and Maimaitijiang et al. demonstrate accurate and robust performance in predicting soybean and maize crop phenotypes. However, for the assessment of weed competition in this study, the MulDFNet model showcases unique advantages in the deep fusion of multi-modal remote sensing data, delivering more comprehensive and precise predictions of weed competition and offering enhanced decision support for agricultural field management.

The multimodal deep fusion model developed in this study performs well in the assessment of weed competition. This method has substantial reference value for research in areas such as agricultural field management, crop growth monitoring, pest and disease prediction. However, due to the highly specialized nature of hyperspectral and LiDAR technologies, there is still some distance to cover for the model's practical application in farmland. Our future research will still require the development of new model approaches to adapt to a wider range of agricultural domains.

5. Conclusion

This study developed a multimodal deep fusion model based on Transformers and multi-layer residuals (MulDFNet), and utilized the Comprehensive Competition Index (CCI-A) derived from multidimensional maize phenotypic data to evaluate the competitiveness of weed in farmland ecosystems. The results indicate that the 21 hyperspectral sensitive bands and 15 vegetation indices obtained using the SPA algorithm, along with canopy height data, possess significant potential for weed competition assessment. The multimodal deep fusion model that using HSI, VI, and CHM data has achieved optimal effects in predicting weed competition, with an R^2 value of 0.903 (RMSE = 0.078). Furthermore, the fused data not only retained the structural characteristics of HSI, VI, and CHM data but also improved data quality. The fusion of multiple modal data leads to better predictive effect and outperforms any individual modality data of HSI, VI, or CHM. In weed competition prediction, our designed multimodal deep fusion model achieved significantly better predictive performance compared to the early/late-stage stacked fusion models and other machine learning models. To summarize, our research demonstrates the validity of the developed multimodal deep fusion model in weed competition assessment. This holds promising prospects for advancing weed management and precision agriculture practices.

CRediT authorship contribution statement

Zhaoxia Lou: Writing – original draft, Methodology. **Longzhe Quan:** Writing – review & editing. **Deng Sun:** Investigation, Data curation. **Fulin Xia:** Investigation, Data curation. **Hailong Li:** Methodology. **Zhiming Guo:** Visualization, Formal analysis.

Declaration of competing 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.

Data availability

The authors do not have permission to share data.

Acknowledgments

The authors thank the National Natural Science Foundation of China (32271998 and 52075092) and Anhui Provincial University Research Program (2023AH040138) for providing financial support for the research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2024.103681.

References

Almeida, D.R.A.d., et al., 2021. Monitoring restored tropical forest diversity and structure through UAV-borne hyperspectral and lidar fusion. Remote Sens. Environ. 264, 112582 https://doi.org/10.1016/j.rse.2021.112582.

Alonzo, M., et al., 2020. Mapping tall shrub biomass in Alaska at landscape scale using structure-from-motion photogrammetry and lidar. Remote Sens. Environ. 245, 111841 https://doi.org/10.1016/j.rse.2020.111841.

Bada, M.R., et al., 2022. Evaluation of weed management practices on weed dynamics and yield of maize (Zea mays L.). Crop. Res. 57, 330–334. https://doi.org/10.31830/2454-1761.2022.CR-879.

Bates, J.S., et al., 2021. Estimating Canopy Density Parameters Time-Series for Winter Wheat Using UAS Mounted LiDAR. Remote Sens. (Basel) 13. https://doi.org/10.3390/rs13040710.

Chen, P., Wang, F., 2022. Effect of crop spectra purification on plant nitrogen concentration estimations performed using high-spatial-resolution images obtained with unmanned aerial vehicles. Field Crop Res 288, 108708. https://doi.org/10.1016/j.fcr.2022.108708.

Chukwudi, U.P., et al., 2021. Influence of heat stress, variations in soil type, and soil amendment on the growth of three drought-tolerant maize varieties. Agronomy 11, 1485. https://doi.org/10.3390/agronomy11081485.

Damalas, C.A., Koutroubas, S.D., 2022. Weed competition effects on growth and yield of spring-sown white lupine. Horticulturae. 8 https://doi.org/10.3390/horticulturae8050430.

Dong, W., et al., 2022. Multibranch feature fusion network with self- and cross-guided attention for hyperspectral and LiDAR classification. IEEE Trans. Geosci. Remote Sens. 60, 1–12. https://doi.org/10.1109/TGRS.2022.3179737.

Du, L., et al., 2018. Application of spectral indices and reflectance spectrum on leaf nitrogen content analysis derived from hyperspectral LiDAR data. Opt. Laser Technol. 107, 372–379. https://doi.org/10.1016/j.optlastec.2018.06.019.

Eitel, J.U.H., et al., 2014. LiDAR based biomass and crop nitrogen estimates for rapid, non-destructive assessment of wheat nitrogen status. Field Crop Res. 159, 21–32. https://doi.org/10.1016/j.fcr.2014.01.008.

Fang, F., et al., 2018. Population dynamics of black-grass Alopecurus myosuroides in wheat fields and its effect on wheat yield components. J. Plant Protect. 2, 340–346.

Fei, S., et al., 2023. UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat. Precis. Agric. 24, 187–212. https://doi.org/10.1007/s11119-022-09938-8.

Feng, Q., et al., 2019. Multisource hyperspectral and LiDAR data fusion for urban land-use mapping based on a modified two-branch convolutional neural network. ISPRS Int. J. Geo Inf. 8 https://doi.org/10.3390/ijgi8010028.

Hütt, C., et al., 2023. UAV LiDAR Metrics for Monitoring Crop Height, Biomass and Nitrogen Uptake: A Case Study on a Winter Wheat Field Trial. PFG – Journal of Photogrammetry. Remote Sensing and Geoinformation Science. 91, 65–76. https://doi.org/10.1007/s41064-022-00228-6.

Karmakar, P., et al., 2024. Crop monitoring by multimodal remote sensing: a review. Remote Sens. Appl.: Soc. Environ. 33, 101093 https://doi.org/10.1016/j.rsase.2023.101093.

Kong, J., et al., 2021. Multi-stream hybrid architecture based on cross-level fusion strategy for fine-grained crop species recognition in precision agriculture. Comput. Electron. Agric. 185, 106134 https://doi.org/10.1016/j.compag.2021.106134.

Kuswidiyanto, L.W., et al., 2022. Plant disease diagnosis using deep learning based on aerial hyperspectral images: a review. Remote Sens. (Basel) 14. https://doi.org/10.3390/rs14236031.

Lazzaro, M., et al., 2019. Unraveling diversity in wheat competitive ability traits can improve integrated weed management. Agron. Sustain. Dev. 39, 6. https://doi.org/10.1007/s13593-018-0551-1.

Li, J., et al., 2022. Deep learning in multimodal remote sensing data fusion: a comprehensive review. Int. J. Appl. Earth Obs. Geoinf. 112, 102926 https://doi.org/10.1016/j.jag.2022.102926.

Liu, N., et al., 2021. Hyperspectral imagery to monitor crop nutrient status within and across growing seasons. Remote Sens. Environ. 255, 112303 https://doi.org/10.1016/j.rse.2021.112303.

Lou, Z., et al., 2022. Hyperspectral remote sensing to assess weed competitiveness in maize farmland ecosystems. Sci. Total Environ. 844, 157071 https://doi.org/10.1016/j.scitotenv.2022.157071.

Ma, J., et al., 2023. Field-scale yield prediction of winter wheat under different irrigation regimes based on dynamic fusion of multimodal UAV imagery. Int. J. Appl. Earth Obs. Geoinf. 118, 103292 https://doi.org/10.1016/j.jag.2023.103292.

Maimaitijiang, M., et al., 2020. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sens. Environ. 237, 111599 https://doi.org/10.1016/j.rse.2019.111599.

Nguyen, C., et al., 2023. UAV multisensory data fusion and multi-task deep learning for high-throughput maize phenotyping. Sensors 23. https://doi.org/10.3390/s23041827.

Pipatsitee, P., et al., 2022. Effectiveness of vegetation indices and UAV-multispectral imageries in assessing the response of hybrid maize (Zea mays L.) to water deficit stress under field environment. Environ. Monit. Assess. 195, 128. https://doi.org/10.1007/s10661-022-10766-6.

Qing, Y., et al., 2021. Improved transformer net for hyperspectral image classification. Remote Sens. (Basel) 13. https://doi.org/10.3390/rs13112216.

Quan, L., et al., 2023. Multimodal remote sensing application for weed competition time series analysis in maize farmland ecosystems. J. Environ. Manage. 344, 118376 https://doi.org/10.1016/j.jenvman.2023.118376.

Rasmussen, J., Nielsen, J., 2020. A novel approach to estimating the competitive ability of Cirsium arvense in cereals using unmanned aerial vehicle imagery. Weed Res. 60, 150–160. https://doi.org/10.1111/wre.12402.

Scavo, A., Mauromicale, G., 2020. Integrated Weed Management in Herbaceous Field Crops. Agronomy 10. https://doi.org/10.3390/agronomy10040466.

Sun, Z., et al., 2022. Simultaneous prediction of wheat yield and grain protein content using multitask deep learning from time-series proximal sensing. Plant Phenomics. 2022, 1–13. https://doi.org/10.34133/2022/9757948.

Swanton, C.J., et al., 2015. Experimental methods for crop-weed competition studies. Weed Sci. 63, 2–11. https://doi.org/10.1614/WS-D-13-00062.1.

Vajari, K.A., 2021. Assessing the intra-specific competition and its relation with tree structure in a beech forest (Fagus orientalis Lipsky). Rev. Bras. Bot. 44, 957–961. https://doi.org/10.1007/s40415-021-00752-6.

Wang, F., et al., 2021. Combining spectral and textural information in UAV hyperspectral images to estimate rice grain yield. Int. J. Appl. Earth Obs. Geoinf. 102, 102397 https://doi.org/10.1016/j.jag.2021.102397.

Wang, X., et al., 2022. Multi-attentive hierarchical dense fusion net for fusion classification of hyperspectral and LiDAR data. Information Fusion. 82, 1–18. https://doi.org/10.1016/j.inffus.2021.12.008.
Watt, M.S., et al., 2020. Using hyperspectral plant traits linked to photosynthetic efficiency to assess N and P partition. ISPRS J. Photogramm. Remote Sens. 169, 406–420.
Weigelt, A., Jolliffe, P., 2003. Indices of plant competition. J. Ecol. 91, 707–720.
Wu, X., et al., 2022. Convolutional Neural Networks for Multimodal Remote Sensing Data Classification. IEEE Trans. Geosci. Remote Sens. 60, 1–10. https://doi.org/10.1109/TGRS.2021.3124913.
Xia, F., et al., 2023. Weed resistance assessment through airborne multimodal data fusion and deep learning: A novel approach towards sustainable agriculture. Int. J. Appl. Earth Obs. Geoinf. 120, 103352 https://doi.org/10.1016/j.jag.2023.103352.
Zhang, M., et al., 2022. Information fusion for classification of hyperspectral and LiDAR data using IP-CNN. IEEE Trans. Geosci. Remote Sens. 60, 1–12. https://doi.org/10.1109/TGRS.2021.3093334.
Zhou, J., et al., 2021. Yield estimation of soybean breeding lines under drought stress using unmanned aerial vehicle-based imagery and convolutional neural network. Biosyst. Eng. 204, 90–103. https://doi.org/10.1016/j.biosystemseng.2021.01.017.
Zovko, M., et al., 2019. Hyperspectral remote sensing of grapevine drought stress. Precis. Agric. 20, 335–347. https://doi.org/10.1007/s11119-019-09640-2.