

Article

Hyperspectral Prediction of Variety, SPAD Value, and Water Content of Oilseed Rape Leaves Using an Improved WGAN-GP

Qinfeng Zhang ¹, Shenghui Shen ¹, Guoyi Yu ¹, Biyao Jin ¹, Junwei Sun ¹, Lupeng Li ² and Chu Zhang ^{1,*} ¹ School of Information Engineering, Huzhou Normal University, Huzhou 313000, China² Institute of Crop Science, Huzhou Academy of Agricultural Sciences, Huzhou 313000, China

* Correspondence: chuzh@zjhu.edu.cn

Abstract

Rapid and non-destructive identification of oilseed rape varieties and prediction of leaf Soil Plant Analysis Development (SPAD) value and water content are important for variety evaluation and plant-status monitoring. However, hyperspectral prediction is constrained by small sample size, costly reference measurements, and acquisition-batch differences. Samples were collected over four consecutive days. Day 1 samples were used for training and Wasserstein generative adversarial network with gradient penalty (WGAN-GP) generation. For Days 2–4, 20% of the samples from each day were combined to form a selection-validation set, while the remaining 80% were retained separately as Test sets 1–3. The improved WGAN-GP integrated principal component analysis–Gaussian mixture model (PCA–GMM)-based class assignment, a partial least squares regression (PLSR) consistency loss, physicochemical distribution control, and generated-sample selection. Under single-task modeling, the improved framework enhanced all three tasks across the test sets. On Test set 3, variety accuracy increased from 0.4672 to 0.6100, SPAD root mean square error of prediction (RMSEP) decreased from 6.2201 to 4.6303, and the water-content test-set correlation coefficient (r_p) increased from 0.6803 to 0.7842. The improved multi-task model also enhanced all three tasks. These findings support the framework within the investigated three-variety, four-day leaf setting.

Keywords: oilseed rape; hyperspectral imaging; WGAN-GP; SPAD; water content

1. Introduction

Oilseed rape (*Brassica napus* L.) is a Brassica crop with high economic value, serving as a vegetable, ornamental plant, and important source of edible oil [1]. With the development of modern breeding technologies, oilseed rape breeding has made substantial progress. Breeding targets are ultimately reflected in phenotypic traits, such as ornamental color, disease and insect resistance, yield, and quality-related indicators [2]. Accurate and efficient phenotypic evaluation is therefore essential for variety selection and breeding improvement [3]. In addition, high-throughput phenotypic monitoring during plant growth can support more precise crop management.

With advances in sensing technologies, the high-throughput acquisition and analysis of plant phenotypes using different sensors have become an important research direction in plant phenotyping [4–7]. Hyperspectral imaging, which integrates spectral and imaging information, can simultaneously capture spectral responses and spatial features of target samples. Combined with machine learning and deep learning methods, hyperspectral



Academic Editor: Surya Kant

Received: 31 July 2026

Revised: 24 August 2026

Accepted: 24 August 2026

Published: 26 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

imaging has been widely used in plant variety identification, physiological and biochemical trait detection, and biotic and abiotic stress phenotyping [8–10].

Hyperspectral images are three-dimensional data cubes with high dimensionality and large data volume, requiring suitable data processing and modeling strategies [11,12]. Previous studies have commonly used machine learning and deep learning methods to analyze spectral data extracted from hyperspectral images for plant phenotypic detection [12–14]. However, most spectral modeling studies have focused on single-task prediction. Even when multiple phenotypic traits are measured from the same samples, separate models are usually established for different tasks [15,16]. Since visible/near-infrared spectra contain integrated information related to multiple sample properties, a single spectrum can potentially support several prediction tasks. Multi-task learning provides an effective strategy to use shared spectral information for the simultaneous prediction of multiple traits, and has been applied in several visible/near-infrared spectral studies [17,18]. For plants, one spectral profile may correspond to multiple phenotypic characteristics, making multi-task learning suitable for simultaneous phenotypic detection.

Although hyperspectral imaging greatly improves phenotyping efficiency compared to manual measurement, the sample size in practical plant phenotyping studies is often limited by sample preparation and low-throughput reference measurements. Moreover, plant growth is dynamic and affected by genotype, individual variation, environmental conditions, acquisition batches, and instrument responses [18,19]. These factors may weaken the generalization ability of models built from a single batch of samples [20]. Despite such differences, visible/near-infrared spectral responses related to plant physicochemical traits may still share certain common patterns across species, growth stages, and acquisition batches.

With the rapid development of artificial intelligence, generative models have been widely used to synthesize data in different modalities, including text, images, and videos. Generative adversarial networks (GANs) are representative generative models consisting of a generator and a discriminator. The generator produces synthetic data, while the discriminator distinguishes real samples from generated ones. Through adversarial training, the generator gradually learns to produce samples that are difficult for the discriminator to distinguish. Several improved GAN models have been developed, including DCGAN, CGAN, ACGAN, WGAN, and WGAN-GP [21,22]. In spectral analysis, GAN-based methods have also been introduced for spectral data generation, and improved variants have been proposed to generate higher-quality synthetic spectra for different research purposes [23–25]. Among these methods, WGAN-GP has attracted attention because of its stable training process, gradient-penalty-based mitigation of mode collapse, and robustness to hyperparameter settings. It has also been explored in spectral data augmentation.

Most GAN-based spectral data generation studies focus on a single classification or regression task, and the generated data are mainly used for single-task modeling. However, synthetic data generated for one task may not be directly applicable to other tasks measured from the same samples. For plant phenotyping, one spectrum usually corresponds to several phenotypic traits. Therefore, generating data that can support simultaneous multi-trait detection is important for spectral-based plant phenotyping. Although a few studies have reported multi-task spectral data generation [26,27], research involving both classification and regression targets remains limited. Limited attention has been given to constructing a unified spectral sample that contains one discrete class target and multiple continuous physiological targets and can be directly used for simultaneous classification and regression within a single multi-output model. In this study, oilseed rape variety, water content, and SPAD value were predicted simultaneously, involving one classification task and two regression tasks.

Existing GAN-based spectral augmentation studies have demonstrated the feasibility of generating synthetic data from available datasets. However, generated samples often remain strongly dependent on the distribution of the original training data, and their diversity and usefulness under different acquisition conditions remain insufficiently examined. Therefore, generating diverse samples that can effectively complement the original training distribution remains an important issue.

In this study, hyperspectral imaging was combined with WGAN-GP to detect oilseed rape leaf variety, water content, and SPAD value. Hyperspectral images of three oilseed rape varieties were collected continuously over four days, and water content and SPAD values were measured. The contribution of this study is a spectral augmentation framework for multi-phenotypic prediction using data collected over four consecutive days. The main objectives were as follows: (1) to establish partial least squares-based classification and regression models using Day 1 samples as the training set and to evaluate their performance on Test sets 1–3 constructed from the samples collected on Days 2–4; (2) to establish a unified spectral generation and Principal Component Analysis–Gaussian Mixture Model (PCA-GMM) soft-labeling strategy for single-task classification, which enables the generation of spectra for all three varieties via a single WGAN-GP model trained on first-day samples; (3) to construct a WGAN-GP framework for single-task regression, a fixed partial least squares regression (PLSR) consistency constraint is incorporated into regression-oriented spectral generation to maintain consistent mapping between generated spectra and physicochemical labels, while an auxiliary distribution constraint is introduced to expand phenotypic coverage; (4) to establish a multi-task spectral generation pipeline to generate unified paired samples integrated with spectra, SPAD values, leaf water content and variety labels for joint classification and regression tasks.

2. Materials and Methods

2.1. Sample Preparation and Sampling Design

The oilseed rape samples consisted of three varieties: Zheyongyou 50, OM43-3, and OM42-3. The plants were sown in November 2025 in the experimental field of the Huzhou Academy of Agricultural Sciences, Zhejiang Province, China, and leaves were collected at the seedling stage in January 2026. Each variety was grown in three separate experimental plots. Each plot covered approximately 20 m² and contained approximately 650 plants. The plots of the three varieties were located adjacent to one another, and all plants were subjected to the same routine field-management practices.

Leaf sampling was conducted over four consecutive days. On each day, intact and healthy leaves were randomly selected from the available plants, without imposing a fixed leaf orientation during subsequent imaging. One leaf was collected from each plant, and no plant was sampled more than once during the four-day experiment. Therefore, the leaves collected on different days originated from different plants. Immediately after detachment, each leaf was placed in an insulated box containing ice packs and transported to the laboratory. The same collection, transportation, imaging, and measurement procedures were applied on all four days. Hyperspectral image acquisition, fresh weight measurement, and SPAD measurement were performed on the same day as the sample collection. The experiment was conducted over four consecutive days, and the daily sample numbers are listed in Table 1.

Table 1. Record table of the number of original samples collected.

	Day 1	Day 2	Day 3	Day 4	Total
Class 0	130	120	120	118	488
Class 1	122	128	122	118	490
Class 2	124	120	122	116	482
Total	376	368	364	352	1460

After hyperspectral image acquisition, the SPAD value and water content of each leaf were measured to establish the corresponding phenotypic reference values.

2.2. Hyperspectral Image Acquisition and Spectral Preprocessing

Hyperspectral images were acquired using an FX17 near-infrared hyperspectral camera (Specim, Spectral Imaging Ltd., Oulu, Finland) equipped with an OLET17.5 lens. The nominal spectral range of the camera was 900–1700 nm, and its spectral resolution was 8 nm. The effective detector pixel size was 18.7 µm. Image acquisition was controlled using Lumo Scanner 2020 software.

Illumination was provided by two symmetrically arranged groups of halogen lamps positioned on both sides of the camera. Each group consisted of three 35 W lamps. The distance between each lamp group and the camera was 20 cm, the vertical distance between the lower edge of the lamps and the moving platform was 22 cm, and the illumination angle was approximately 45°. The camera-to-sample distance was fixed at 31.4 cm. The camera exposure time was set to 8.26 ms, and the platform scanning speed was 24.7 mm s^{−1} to minimize geometric distortion.

Before image acquisition, the camera and halogen lamps were switched on and allowed to stabilize for 15 min. During imaging, the leaves were randomly oriented on the moving platform and arranged without contact. Because the entire leaf was included in the hyperspectral image, no fixed leaf-placement direction was imposed.

To improve image quality and reduce the effects of dark current and illumination variation, the raw hyperspectral images were corrected using white and dark reference images. The corrected image was calculated in Equation (1).

$$R = \frac{I - D}{W - D} \quad (1)$$

where R is the corrected image, I is the raw image, D is the dark reference image, and W is the white reference image.

After hyperspectral image acquisition, spectral data were extracted for each oilseed rape leaf. Background pixels were removed using threshold segmentation, and connected-component labeling was then applied to identify individual leaf regions. Pixels with adjacent positions and gray-level characteristics corresponding to leaf samples were grouped to construct the region of interest (ROI) for each sample. The average spectrum of all pixels within each ROI was calculated to represent the spectral information of the corresponding leaf and used as the raw spectrum for subsequent analysis.

To reduce noise at the spectral edges, the first and last 12 bands were removed, and 200 middle bands ranging from 943 to 1657 nm were retained. To improve spectral stability and modeling performance, the raw spectra were preprocessed using seven-point moving average smoothing and standard normal variate (SNV) transformation [28]. The moving average smoothing was used to reduce random noise caused by the imaging system and environmental disturbance while preserving the overall spectral profile. SNV was applied to each individual spectrum to reduce the effects of surface variation, light scattering, and baseline drift, thereby improving spectral comparability among samples [28].

Accordingly, the spectral curves presented in the subsequent analysis represent SNV-transformed spectral values rather than physical reflectance. Figure 1 shows the overall workflow of spectral extraction and preprocessing.

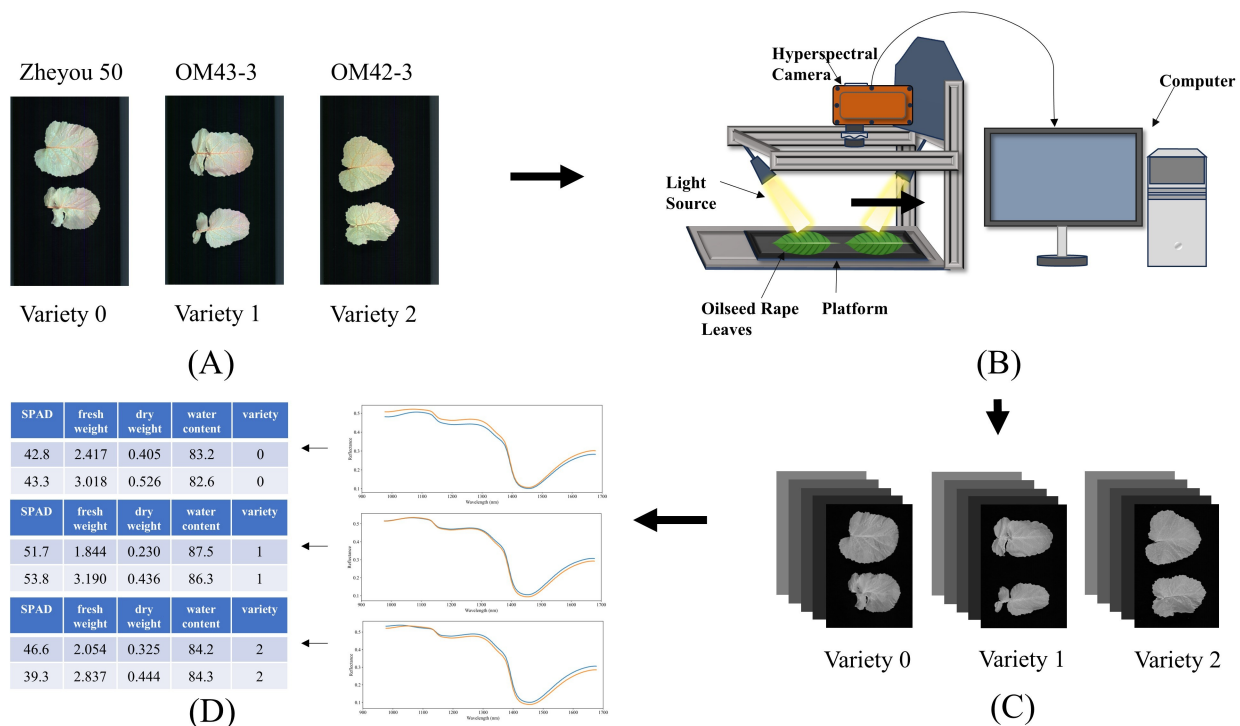


Figure 1. (A) Pseudo-color image (formed by grayscale images at wavelengths 1078 nm, 1340 nm and 1606 nm) in the hyperspectral images, (B) Hyperspectral imaging system, (C) Hyperspectral cube image, (D) Spectral curve of the sample and its label.

2.3. Reference Measurements

Variety label and physiological traits were measured immediately after hyperspectral image acquisition. All measurements were conducted under light-avoiding conditions following standard procedures. The label values of all samples collected each day were measured on the same day, and sample IDs were kept consistent throughout the process to ensure one-to-one correspondence between spectral data and physicochemical traits. The variety names were provided by Huzhou Agricultural Science Research Institute. Zheyu 50, OM43-3, and OM42-3 were encoded as class labels 0, 1, and 2, respectively.

After the samples arrived at the laboratory, hyperspectral images were acquired first. The fresh mass of each leaf was then measured immediately after image acquisition, with an interval of less than 5 s between the completion of imaging and weighing. The SPAD value was subsequently measured at five evenly distributed positions from the leaf tip to the base using an SPAD-502 chlorophyll meter (Konica Minolta, Osaka, Japan), and the mean of the five readings was used as the SPAD reference value [17]. Sample identifiers were maintained throughout image acquisition and reference measurement to ensure one-to-one correspondence among each spectrum, fresh mass, SPAD value, water content, and variety label.

Fresh and dry masses were measured using a calibrated Sartorius BSA series balance (Sartorius, Göttingen, Germany) with a capacity of 220 g, a readability of 1 mg, and a specified repeatability of 1 mg. Before formal measurements, a repeatability test was conducted using six leaves. Each leaf was removed from and repositioned on the balance and weighed three consecutive times. No variation was observed among the three displayed readings at the balance readability. Based on this preliminary test, one fresh-mass and one dry-mass

reading were recorded for each sample in the formal experiment. Following fresh-mass and SPAD measurements, the leaves were dried in an oven at 75 °C for 8 h, after which their dry masses were recorded using the same balance. The drying conditions were selected based on a previous oilseed rape leaf study, in which a preliminary experiment was conducted by weighing leaves after 2 h of drying and then at 30 min intervals. In that experiment, no further mass change was observed after 4 h at 75 °C. The 8 h drying period used in the present study therefore exceeded the previously validated minimum drying duration and was adopted to ensure sufficient dehydration of the samples [29]. Leaf water content (WC) was calculated in Equation (2).

$$WC = \frac{FW - DW}{FW} \times 100\% \quad (2)$$

where WC is the leaf water content, FW is the fresh weight, and DW is the dry weight. This method estimates water content based on the proportion of water loss relative to fresh weight and can effectively reflect the water status of oilseed rape leaves.

Finally, each leaf sample was assigned a complete dataset consisting of the reflectance spectrum, SPAD value, water content, and class label. The SPAD value represented the relative chlorophyll level, water content represented the leaf water status, and the class label identified the oilseed rape variety. SPAD and water content were used as continuous target variables for single-task regression and multi-task modeling, while class labels were used as discrete targets for classification and multi-task modeling.

2.4. Outlier Sample Removal

To improve dataset reliability and reduce the influence of abnormal samples on subsequent modeling and generated data quality, outlier removal was performed before model construction. First, outliers in SPAD and water content were removed using boxplot analysis based on the interquartile range [30]. Then, principal component analysis (PCA) was applied to extract the main spectral variation. PCA can compress high-dimensional and highly collinear spectral variables into a few principal components while retaining the major variance information, making it useful for identifying samples that deviate from the overall distribution [31]. Based on the PCA results, Hotelling's T^2 statistic was used to evaluate the deviation in each sample in the principal component space, and spectral outliers exceeding the 95% confidence limit were removed [32,33]. These criteria were applied as fixed statistical quality-control rules and were not selected, adjusted, or optimized according to subsequent model performance, generated-sample screening results, or the performance of the selection-validation and test sets. The purpose of this procedure was therefore to remove potentially abnormal measurements arising from experimental or measurement variability rather than to improve the prediction results of a particular model. Similar pre-modeling quality-control procedures for measured trait or spectral outliers have also been adopted in previous spectroscopy and chemometric studies [34–36].

After two rounds of physicochemical outlier screening and PCA-based spectral outlier removal, the descriptive statistics of the remaining samples for each day are shown in Table 2. As shown in Table 2, the physicochemical traits of oilseed rape leaves changed with acquisition time. The mean SPAD value decreased continuously from 50.57 on Day 1 to 45.93 on Day 4, showing a clear temporal trend. In contrast, the mean water content remained relatively stable at approximately 86%, with low within-day variation, although slight differences were still observed among acquisition batches.

Table 2. Descriptive statistics for sample measurement (WC stands for water content).

	Sample Number	Range (SPAD)	Mean (SPAD)	Standard Deviation (SPAD)	Range (WC)	Mean (WC)	Standard Deviation (WC)
Day 1	344	38.8–65.9	50.57	5.54	83.6–89.4	86.38	1.04
Day 2	347	37.1–62.7	49.21	5.01	82.5–89.0	86.13	1.25
Day 3	341	34.6–57.9	46.73	4.92	82.4–89.4	86.10	1.26
Day 4	324	35.6–57.6	45.93	4.42	83.5–88.9	86.06	0.93

2.5. Improved WGAN-GP Algorithm

In this study, WGAN-GP was used as the basic generative framework. According to the characteristics of oilseed rape leaf spectra and phenotypic prediction tasks, task-specific data generation strategies were developed for single-task classification, single-task regression, and multi-task prediction, providing augmented training samples for subsequent spectral modeling.

2.5.1. Improved WGAN-GP for Single-Task Data Generation

The single-task phenotypic detection in this study included oilseed rape variety classification and regression prediction of SPAD and water content. For single-task classification, the objective was to augment spectral samples for different varieties and improve the classification model. Previous studies often generated class-specific samples by training separate models for each class [23]. In this study, all three classes were used together for WGAN-GP-based spectral generation to reduce model complexity, and a PCA-GMM-based soft labeling strategy was introduced to assign class labels to generated spectra. Specifically, a PCA-GMM mapping model was first built using real spectra. PCA was used to project the spectra into a low-dimensional feature space while retaining the main variance information and reducing spectral redundancy [31]. Gaussian mixture model (GMM) clustering was then performed, and the class proportion within each cluster was used to establish a cluster–class probability mapping. For each generated spectrum, the sample was projected into the same PCA space, and its posterior probabilities of belonging to different GMM clusters were calculated. The final class probability was obtained by weighting the cluster posterior probability with the corresponding class probability within each cluster. The equation is shown in Equation (3).

$$P(c|x_g) = \sum_{k=1}^K P(k|x_g)P(c|k) \quad (3)$$

where K is the number of GMM clusters, $P(k|x_g)$ is the probability that the generated sample belongs to the k -th cluster, and $P(c|k)$ is the proportion of class c within that cluster. The class with the highest probability was assigned as the final label of the generated sample. Compared to hard class mapping, this soft labeling strategy better represents the uncertainty of samples near class boundaries and reduces the risk of incorrect label assignment in subsequent classification modeling [37]. In this way, the generated samples formed spectrum–class label pairs for single-task classification.

Based on the spectral generation strategy for classification, an improved WGAN-GP was further developed for single-task regression to simultaneously generate oilseed rape leaf spectra and their corresponding continuous physicochemical values. Unlike classification-oriented generation, regression-oriented generation needs to ensure not only that the generated spectra are close to the real spectral distribution, but also that the generated spectra and labels maintain a reasonable predictive relationship. Therefore, the spectrum and one physicochemical indicator were jointly modeled in WGAN-GP, and

PLSR consistency and physicochemical distribution constraints were introduced into the generator loss to improve the usability of generated samples. Each real sample consisted of 200 spectral variables and one continuous target value. The target was the SPAD value for SPAD prediction and water content for water content prediction. Since both targets were continuous variables, the same regression generation framework was used, with only the normalization parameters adjusted according to the physical range of each indicator.

The discriminator took either real or generated joint vectors as input. For the generator, the adversarial loss alone can encourage generated samples to approach the real data distribution, but it cannot fully ensure a reasonable correspondence between generated spectra and generated physicochemical values. Therefore, a pretrained PLSR model was introduced as an auxiliary constraint. The PLSR model was first trained using real spectra and their corresponding physicochemical values, with the optimal number of latent variables determined by cross-validation. During WGAN-GP training, the PLSR parameters were fixed. The generated spectra were fed into the PLSR model to predict physicochemical values, and a PLSR consistency loss was constructed to constrain the agreement between generated spectra and labels. This reduced the occurrence of samples with realistic spectra but mismatched labels. The PLSR consistency loss was defined in Equation (4).

$$L_{PLSR} = \text{MSE}(R(\tilde{x}), \tilde{y}) \quad (4)$$

where $(R(\cdot))$ denotes the pretrained PLSR regressor, $\tilde{x}$ is the generated spectrum, and $\tilde{y}$ is the generated physicochemical value.

In addition, a physicochemical distribution constraint (PDC) was designed to improve label controllability and reduce excessive dependence on the range of a single training batch. This constraint controlled the mean and standard deviation in generated labels within each batch, allowing the generated samples to remain close to the training distribution while expanding in a predefined direction. The loss was defined in Equation (5).

$$L_{PDC} = (\mu_g - \mu_t)^2 + (\sigma_g - \sigma_t)^2 \quad (5)$$

where μ_g and σ_g are the mean and standard deviation in generated physicochemical values in the current batch, and μ_t and σ_t are the predefined target mean and standard deviation. The physicochemical variables were mapped to the normalized interval $[-1, 1]$ using predefined physical ranges of 34–66 for SPAD and 80–90 for water content. Accordingly, generated target values were mapped back to these physical ranges after inverse normalization. Within this normalized space, the reference target mean and standard deviation used in the PDC penalty were set to $\mu_t = -0.80$ and $\sigma_t = 0.80$, respectively, for both SPAD and water content. These values served as soft distributional reference targets rather than statistical quantities that the generated samples were required to satisfy exactly. Specifically, the negative target mean encouraged the generated labels toward the lower part of the predefined physical range, whereas the relatively large target standard deviation encouraged a broader label distribution.

Thus, the total generator loss for single-task regression consisted of three terms: the WGAN adversarial loss, the PLSR consistency loss, and the physicochemical distribution constraint. The loss was defined in Equation (6).

$$L_G = w_{adv}L_{adv} + w_{PLSR}L_{PLSR} + w_{PDC}L_{PDC} \quad (6)$$

where $L_{adv} = -\mathbb{E}_{z \sim p_z(z)} [D(G(z))]$, and w_{adv} , w_{PLSR} , and w_{PDC} are the weights of the three loss terms, which were set to 0.8, 0.1, and 0.1, respectively. A higher weight was assigned to the adversarial loss because the generator first needed to learn the overall

distribution of real spectra. The PLSR consistency and distribution constraints were used as auxiliary terms with smaller weights to avoid over-restricting the generator and reducing sample diversity.

After training, random noise was input into the generator to obtain synthetic joint samples in the normalized space. The generated spectra and physicochemical values were then inverse-normalized to the original scale. The improved WGAN-GP could generate spectra close to the real distribution while strengthening the correspondence between spectra and physicochemical values through the PLSR consistency constraint, thereby providing more physically reasonable and predictive samples for single-task regression modeling.

Figure 2 illustrates the training process of the improved WGAN-GP for single-task data generation.

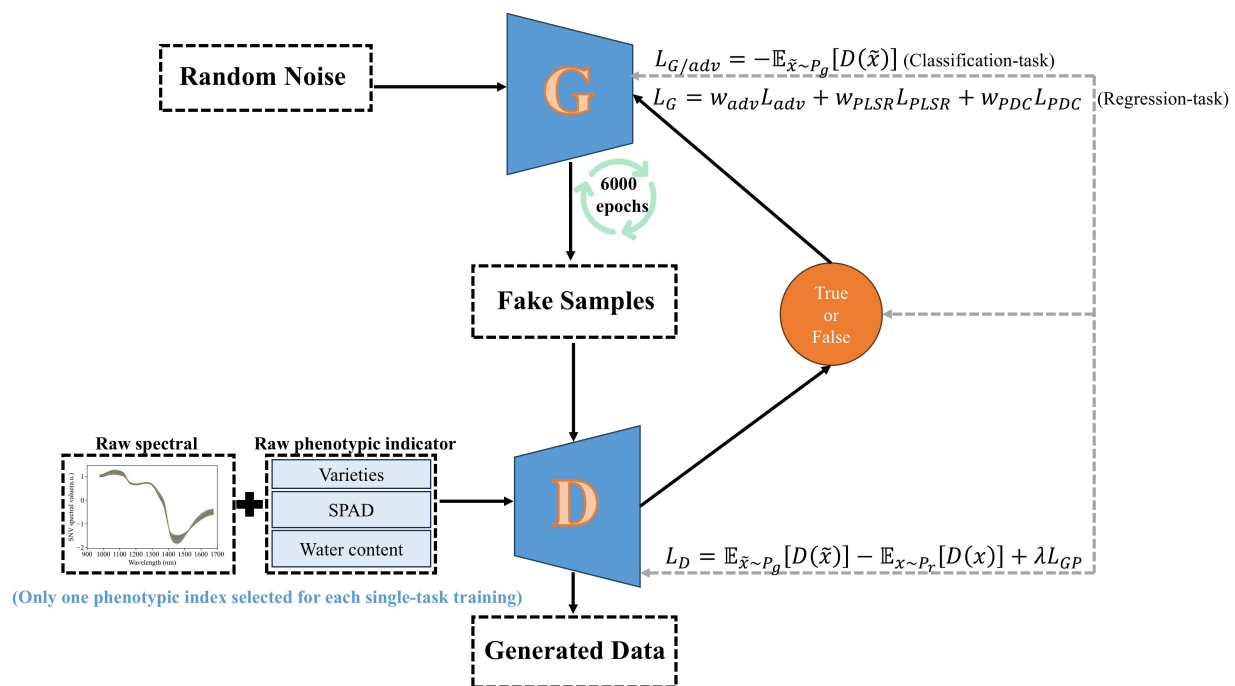


Figure 2. Schematic diagram of the WGAN-GP training framework for single-task learning.

2.5.2. Improved WGAN-GP for Multi-Task Data Generation

In single-task spectral data generation, the model mainly focuses on the relationship between spectra and one physicochemical indicator. The SPAD value and water content are continuous physiological variables, whereas variety is a discrete genetic category. Although these targets have different biological meanings, they may share spectral information because genotype and leaf physiological status can jointly influence near-infrared reflectance. Therefore, a multi-task framework was developed to use the common spectral representation for simultaneous variety classification and SPAD and water-content prediction.

For multi-task generation, each real sample consisted of 200 spectral variables, two continuous physicochemical indicators, and one class label. Since the class label was discrete, directly generating it as a continuous output could lead to ambiguous class assignments. Therefore, only the spectra, SPAD values, and water content were used as continuous joint vectors during WGAN-GP training. The spectral variables were normalized to $[-1, 1]$ using Min–Max normalization, while SPAD and water content were normalized according to their physical ranges to reduce the influence of scale differences on model training.

In the network structure, the generator took random noise z as input and produced a 202-dimensional joint vector through one-dimensional transposed convolution. The first 200 dimensions represented the generated spectrum, and the last two dimensions

represented the corresponding SPAD value and water content. The discriminator took real or generated joint vectors as input, extracted joint features using one-dimensional convolution, and output a Wasserstein score. Compared to WGAN-GP that generates spectra only, this structure enabled the model to learn the relationship between spectral features and continuous phenotypic traits within the same generation process, thereby improving multi-task consistency.

To strengthen the consistency between generated spectra and the two continuous traits, a multi-output PLSR (Multi-PLSR) model was introduced as an auxiliary constraint. The PLSR consistency losses for SPAD and water content were calculated separately. In addition, physicochemical distribution constraints were designed to expand the phenotypic space of the original training data within a reasonable range. These constraints acted on the generated SPAD and water content labels by controlling their means and standard deviations, allowing the generated samples to maintain physical plausibility while covering a broader distribution.

Based on this design, the PLSR consistency loss and physicochemical distribution constraint were combined for each continuous task to construct the SPAD loss and water content loss. The PLSR consistency loss was assigned the main weight to ensure the predictive agreement between generated spectra and labels, while the distribution constraint was assigned a smaller weight to moderately regulate the label distribution and avoid excessive concentration within the original training range.

Finally, the total generator loss consisted of three terms: the WGAN adversarial loss, the SPAD combined loss, and the water content combined loss. The loss was defined in Equation (7).

$$L_G = w_{adv}L_{adv} + w_{SPAD}L_{SPAD} + w_{WC}L_{WC} \quad (7)$$

where $L_{adv} = -E [D(G(z))]$, $D(\cdot)$ is the Wasserstein score output by the discriminator, and $G(z)$ is the generated joint sample. The SPAD and WC losses were calculated separately using the same PLSR consistency and physicochemical distribution terms defined in Equations (4) and (5). Consistent with single-task regression generation, a higher weight was assigned to w_{adv} to ensure that the generated spectra remained close to real spectra in shape and distribution.

Since class labels are discrete variables, directly generating them as continuous outputs in WGAN-GP may lead to unclear class boundaries and ambiguous class assignments. Therefore, class labels were not directly output by the generator. Instead, after spectra and continuous physicochemical values were generated, class labels were assigned using the same PCA-GMM soft mapping strategy as in single-task classification.

Overall, the proposed multi-task improved WGAN-GP generated not only spectra but also joint data reflecting the relationships among spectra, SPAD, water content, and variety. WGAN-GP was used to maintain the overall distributional similarity between real and generated spectra, while PLSR consistency and physicochemical distribution constraints improved the reliability of generated continuous labels. Class labels were further obtained through PCA-GMM soft mapping. The final generated data format was “200-dimensional spectrum + SPAD + water content + class label”, which could be directly used for training multi-task prediction models. This strategy was designed to alleviate the limitation of small training datasets and improve the utility of generated samples for subsequent phenotypic prediction.

Figure 3 shows the training process of the improved WGAN-GP for multi-task data generation.

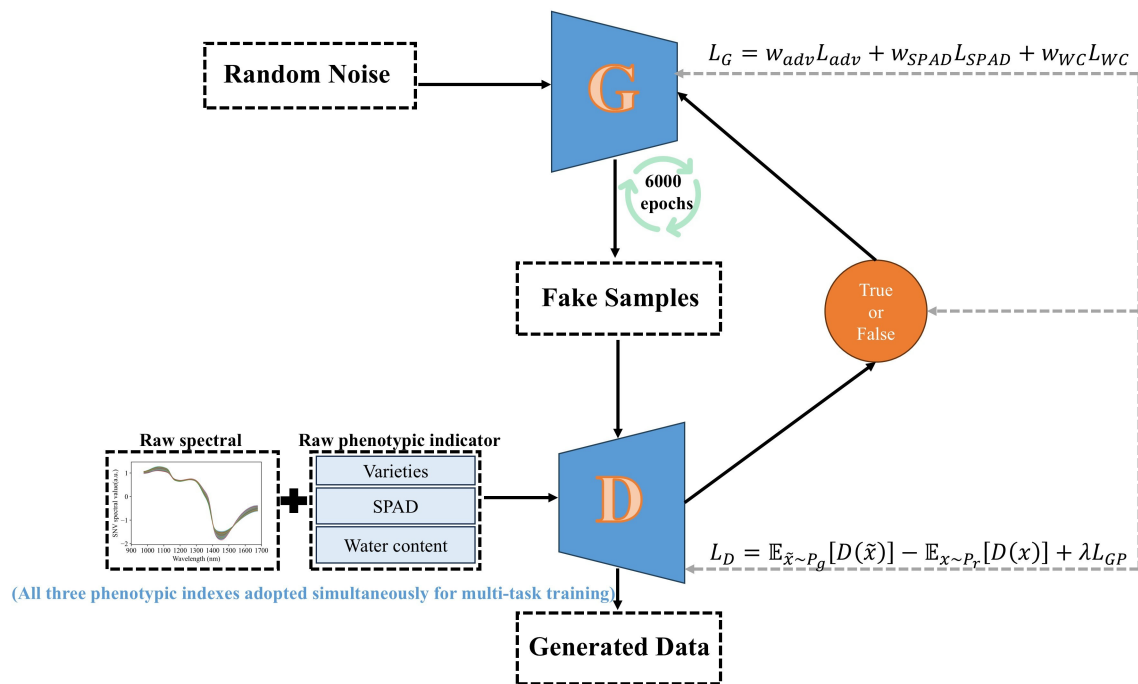


Figure 3. Schematic diagram of the WGAN-GP training framework for multi-task learning.

2.6. Dataset Partition and Generated-Sample Selection

The samples collected on Day 1 were used as the training set and as the real samples for training the WGAN-GP models. After outlier removal, the samples collected on Days 2, 3, and 4 were separately divided into a selection-validation subset and a test subset reserved for final evaluation at a ratio of 20:80.

For each acquisition day, dataset partitioning was performed within each variety using the Kennard–Stone algorithm based on the Euclidean distances between standardized spectral variables [38]. This procedure ensured that the 20% selection subset covered the major spectral variations in the corresponding day while maintaining adequate representation of all three varieties. The SPAD values and water-content measurements were not used when calculating the Kennard–Stone distances. The three day-specific 20% subsets were combined to form a selection-validation set containing 202 samples, while the remaining 278, 273, and 259 samples constituted Test sets 1, 2, and 3, respectively. No samples were shared between the selection-validation set and any of the three test sets. For each task, 5000 candidate samples were first generated by the trained WGAN-GP and evaluated sequentially according to their generation order. Starting from the first candidate, one generated sample was temporarily added to the current training set, which consisted of the original Day 1 samples and all previously accepted generated samples. A partial least squares-based surrogate model was then rebuilt and evaluated on the selection-validation set. For classification, PLS-DA accuracy was used as the evaluation criterion, whereas for regression, the PLSR correlation coefficient and RMSE were jointly considered. If the inclusion of a candidate resulted in deterioration of the corresponding selection-validation performance, the candidate was rejected and the next candidate was evaluated; otherwise, it was retained and included in the training set for the subsequent iteration. No additional empirical numerical threshold was applied during this process. The candidates were evaluated sequentially until 344 generated samples had been retained, which was equal to the number of original Day 1 training samples. All reported test results were calculated using Test sets 1–3 retained from Days 2–4.

2.7. Phenotypic Analysis Modeling Method

2.7.1. Partial Least Squares Discriminant Analysis

PLS-DA was used for single-task classification of oilseed rape leaf samples. PLS-DA extends partial least squares regression to classification by converting discrete class labels into a numerical response matrix and extracting latent variables that maximize the covariance between the spectral matrix and the class response matrix [39]. This method is suitable for high-dimensional spectral data because it reduces spectral redundancy while retaining discriminative information for classification. In this study, the optimal number of latent variables was selected by grid search with 10-fold cross-validation on the training set, and the search range was set from 1 to 20.

2.7.2. Partial Least Squares Regression

PLSR was used to build single-task regression models for quantitative prediction of leaf SPAD value and water content. As a classical multivariate calibration method in chemometrics, PLSR is well-suited for spectral analysis, where data often show high dimensionality, strong multicollinearity, and noise interference [40]. In this study, the optimal number of latent variables was selected by grid search with 10-fold cross-validation on the training set, with the search range set from 1 to 20.

2.7.3. Multi-PLS for Multi-Task Prediction

To simultaneously predict class label, SPAD value, and leaf water content from the same spectral input, a multi-output partial least squares model was further established. SPAD value and water content were arranged as a two-column continuous response matrix and modeled using multi-response PLSR. The variety labels 0, 1, and 2 were transformed into three-dimensional one-hot vectors, i.e., [1, 0, 0], [0, 1, 0], and [0, 0, 1], respectively. During prediction, the classification model generated three continuous outputs corresponding to the three varieties, and the class associated with the maximum output value was assigned as the final predicted variety label. This approach can capture common spectral information related to different traits and improve modeling efficiency when several correlated targets need to be predicted simultaneously [41]. In this study, no additional manual scaling was applied to the response variables before PLS modeling; the spectral variables and response variables were internally centered and scaled during PLS fitting. The optimal number of latent variables was selected by grid search with 10-fold cross-validation on the training set, with the search range set from 1 to 20.

2.8. Performance Evaluation Metrics

To evaluate model performance and the quality of generated data for both the original and augmented training sets, classification accuracy (Acc.), correlation coefficient (r), and root mean square error (RMSE) were used as the main evaluation metrics.

For classification, Acc. was used to assess the ability of the model to identify different oilseed rape varieties. It was calculated in Equation (8).

$$\text{Acc.} = \frac{N_{\text{correct}}}{N} \quad (8)$$

where N_{correct} is the number of correctly classified samples and N is the total number of samples. A higher Acc indicates better overall classification performance.

Precision was used to evaluate the reliability of positive predictions for each class. It represents the proportion of correctly predicted positive samples among all samples predicted as positive. It was calculated in Equation (9).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

where TP is the number of true positive samples and FP is the number of false positive samples. A higher Precision indicates fewer false-positive classifications.

Recall was used to evaluate the ability of the model to correctly identify samples belonging to each class. It represents the proportion of correctly identified positive samples among all actual positive samples. It was calculated in Equation (10).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

where FN is the number of false negative samples. A higher Recall indicates that a larger proportion of samples from the corresponding class is correctly recognized.

The F1-score was used to provide a balanced evaluation of Precision and Recall. It is defined as their harmonic mean. It was calculated in Equation (10).

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

For regression, r was used to evaluate the strength of the linear association between measured and predicted values. It was calculated in Equation (11).

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (12)$$

where y_i and $\hat{y}_i$ are the measured and predicted values of the i -th sample, respectively; $\bar{y}$ and $\bar{\hat{y}}$ are their corresponding mean values; and n is the number of samples. An r value closer to 1 indicates a stronger positive linear association between measured and predicted values and a better ability of the model to capture variation in the reference measurements.

RMSE was used to quantify the average prediction error. It was calculated in Equation (12).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

where the variables are defined as above. A lower RMSE indicates a smaller average prediction error and greater predictive accuracy. Because SPAD and water content have different units and value ranges, r and RMSE were calculated separately for the two regression tasks. For the training, selection-validation and test sets, they were denoted as r_c , RMSEC, r_v , RMSEV, r_p , and RMSEP, respectively.

2.9. Software and Hardware

All data processing and model development were performed on computer systems equipped with 32 GB RAM, an NVIDIA GeForce RTX 3060 GPU, and Intel i7-11700 CPU. The computational environment utilized Python (v3.9), with PyCharm IDEs (2023.1) as the programming platforms.

3. Results

3.1. Spectral Analysis

Figure 4 shows the mean SNV-transformed spectral curves of the four acquisition batches from Day 1 to Day 4. Overall, the oilseed rape leaf samples showed similar spectral trends within 943–1657 nm, indicating consistent basic reflectance characteristics

in the near-infrared region. However, differences in SNV-transformed spectral values were still observed among batches, which may be related to variations in leaf water content, chlorophyll level, internal tissue structure, and other physiological traits.

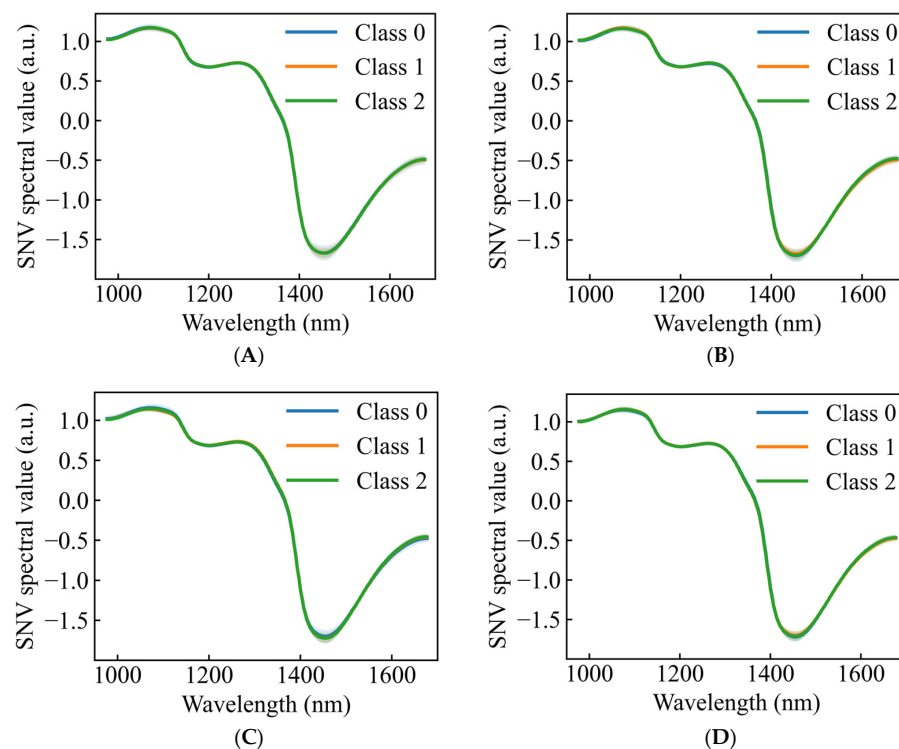


Figure 4. Mean spectra with standard deviation in each wavelength of four batches of spectra (A) Training set (B) Test set 1 (C) Test set 2 (D) Test set 3.

In terms of spectral shape, the SNV-transformed spectra showed relatively high spectral values around 1000–1100 nm, followed by a gradual decrease after 1150 nm and a distinct absorption valley around 1400–1500 nm. In the near-infrared region, leaf spectra are mainly affected by water, organic compounds, and cellular structure. The overtone and combination absorptions of O–H, C–H, and N–H bonds are important contributors to NIR spectral features [42]. The fluctuations observed at 943–1050 nm may be associated with higher-order overtone absorption of O–H bonds in water, reflecting changes in leaf water status. The spectra were relatively smooth within 1100–1300 nm, but this region still contained information related to C–H and O–H vibrations, which may be associated with differences in dry matter, cellular structure, and organic composition [29].

3.2. Generated Data Analysis

3.2.1. Observed Generated Spectral Curve

According to Section 2.6, generated samples that markedly weakened the original spectral–label relationship or reduced prediction performance on the selection-validation set were excluded. The number of retained generated samples was set equal to the number of real training samples; therefore, 344 generated samples were retained for each task. Since conventional WGAN-GP usually generates class-specific samples by training separate models for different classes, this study did not adopt class-wise WGAN-GP generation for single-task classification in order to maintain consistency with the single-task regression and multi-task experiments. Instead, only the improved WGAN-GP was used for single-task classification data generation. The spectral curves of all Day 1 real samples, as well as the spectra generated by WGAN-GP and the improved WGAN-GP for single-task and

multi-task data generation, are shown in Figure S1 in the Supplementary Materials. As shown in Figure S1, the generated spectra maintained similar curve shapes and trends to the real spectra, although some differences and fluctuations in spectral intensity were observed.

Figure 5 shows the mean spectral curves of the selected samples generated by the improved WGAN-GP for single-task classification and multi-task data generation. The spectral curves generated for single-task regression showed similar shapes and trends and are therefore not shown. As shown in Figure 5, the generated and real spectra were highly consistent in overall distribution and variation trend. Meanwhile, subtle but distinguishable differences among different classes were retained, providing useful spectral information and data support for subsequent modeling analysis.

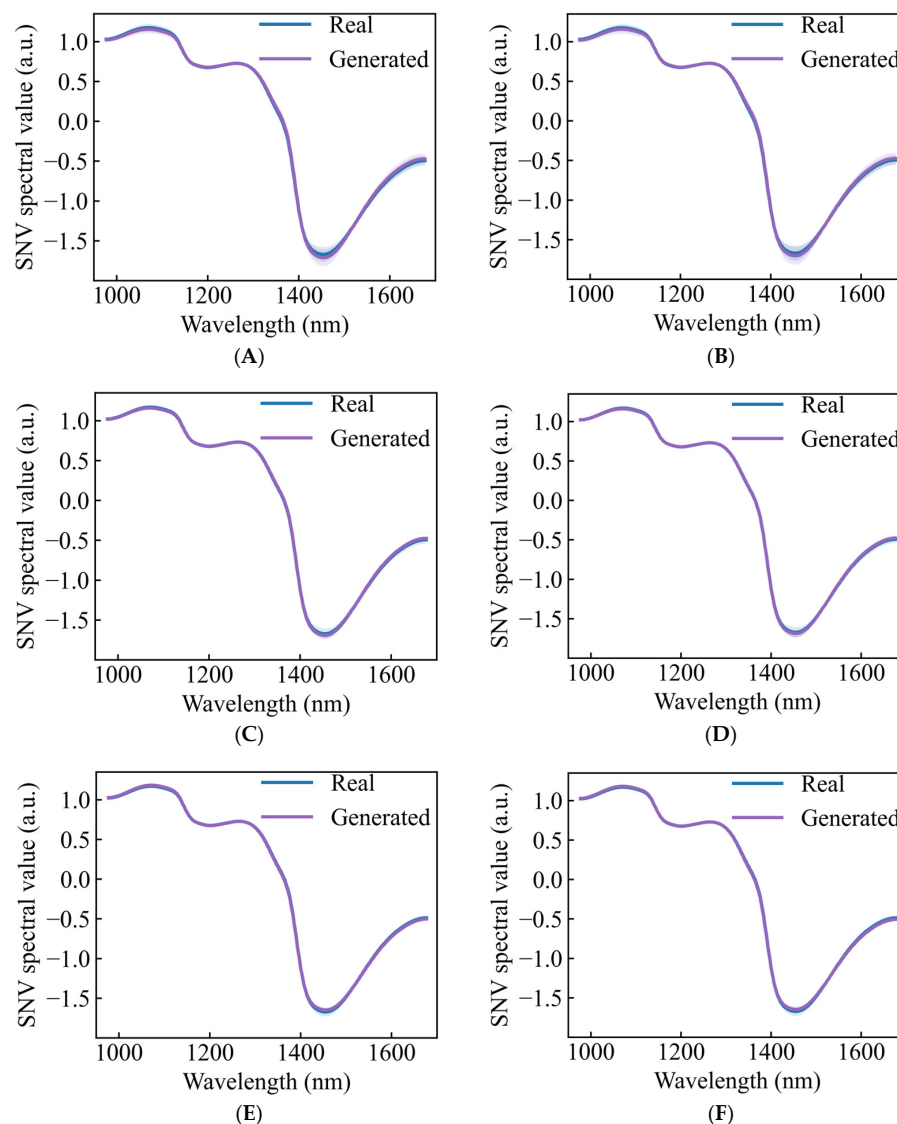
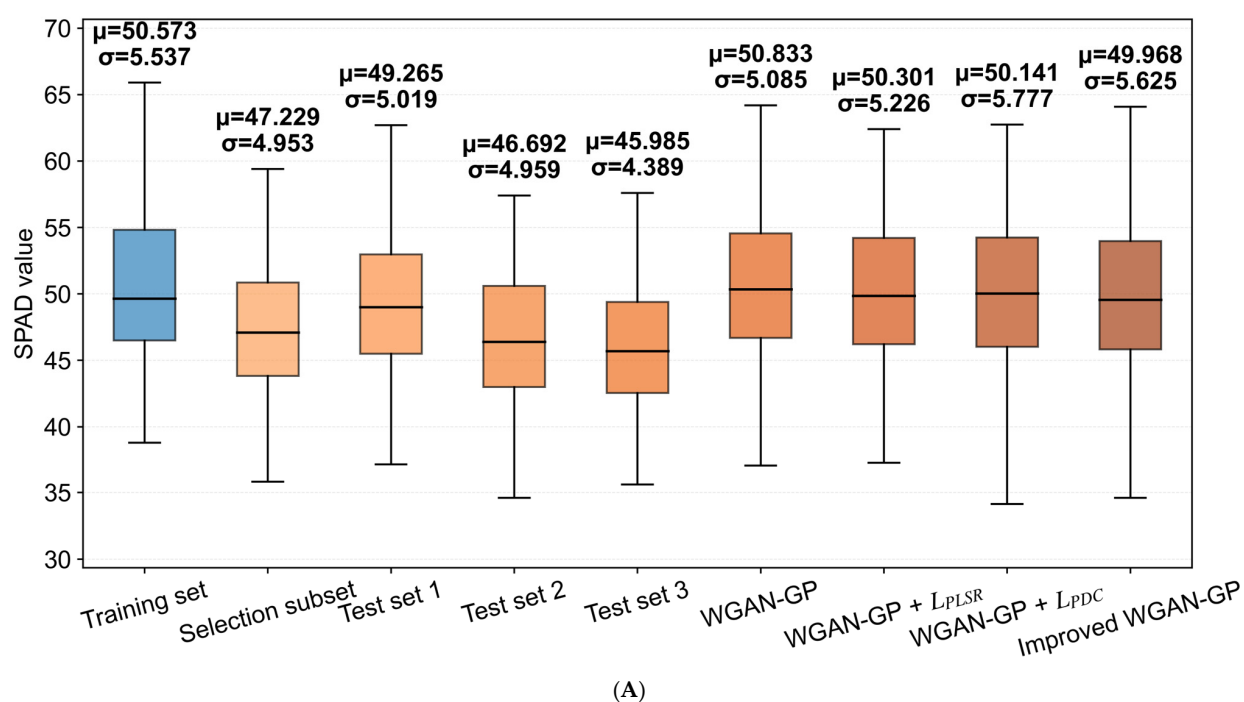


Figure 5. Mean spectra with standard deviation in each wavelength between real and generated data. (A) Class 0 of Single-task (classification) (B) Class 0 of Multi-task (C) Class 1 of Single-task (classification) (D) Class 1 of Multi-task (E) Class 2 of Single-task (classification) (F) Class 2 of Multi-task.

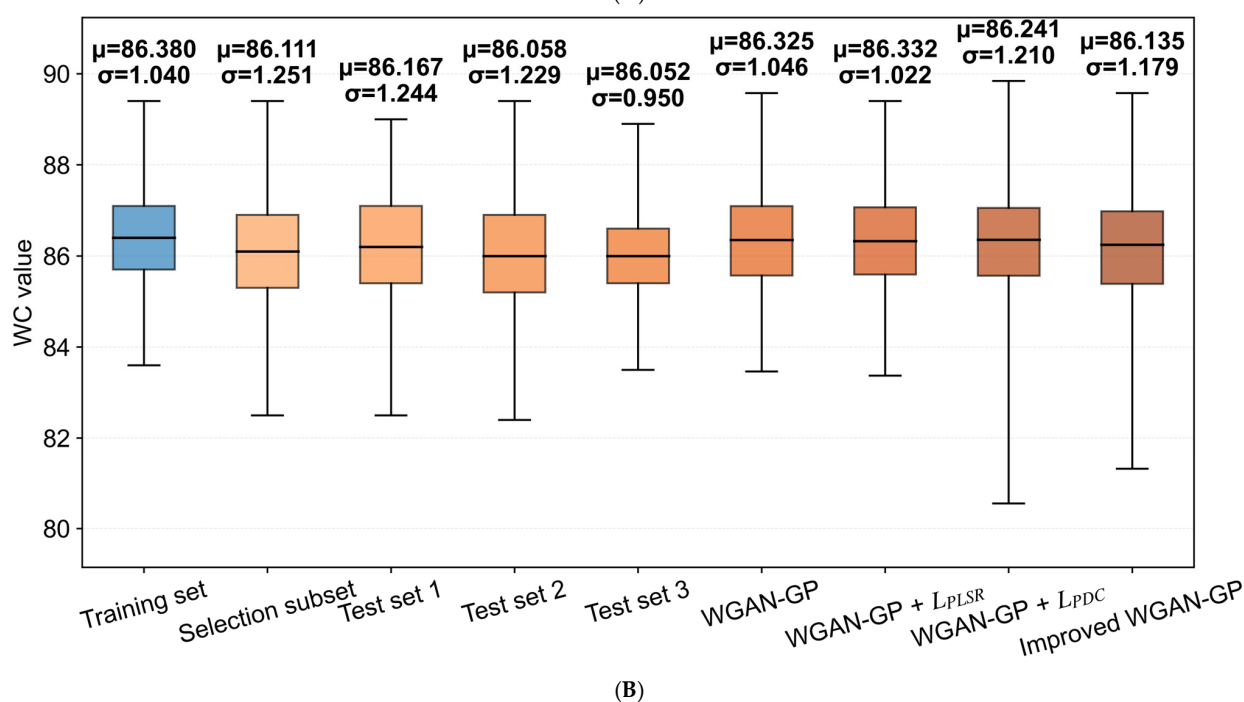
3.2.2. Box Plot Results

Figure 6 compares the distributions of SPAD and water-content labels in the training set, selection set, test sets, and generated datasets. Taking water content as an example, the basic WGAN-GP and the variant with the PLSR consistency loss largely reproduced the

mean and dispersion of the Day 1 training data. After introducing the physicochemical distribution constraint, the generated water-content values showed a moderately lower mean and greater dispersion, resulting in a broader distribution that showed greater coverage of the values observed in the selection set and Test sets 1–3. The complete improved WGAN-GP retained this broader coverage while avoiding excessive deviation from the real-data range. Similar patterns were observed for SPAD generation and in both the single-task and multi-task settings. The PLSR consistency loss mainly preserved the relationship between generated spectra and labels, whereas the physicochemical distribution constraint adjusted the location and spread of the generated labels. Their combination therefore enabled the improved WGAN-GP to maintain physically plausible distributions while extending the coverage of the original training data.



(A)



(B)

Figure 6. Cont.

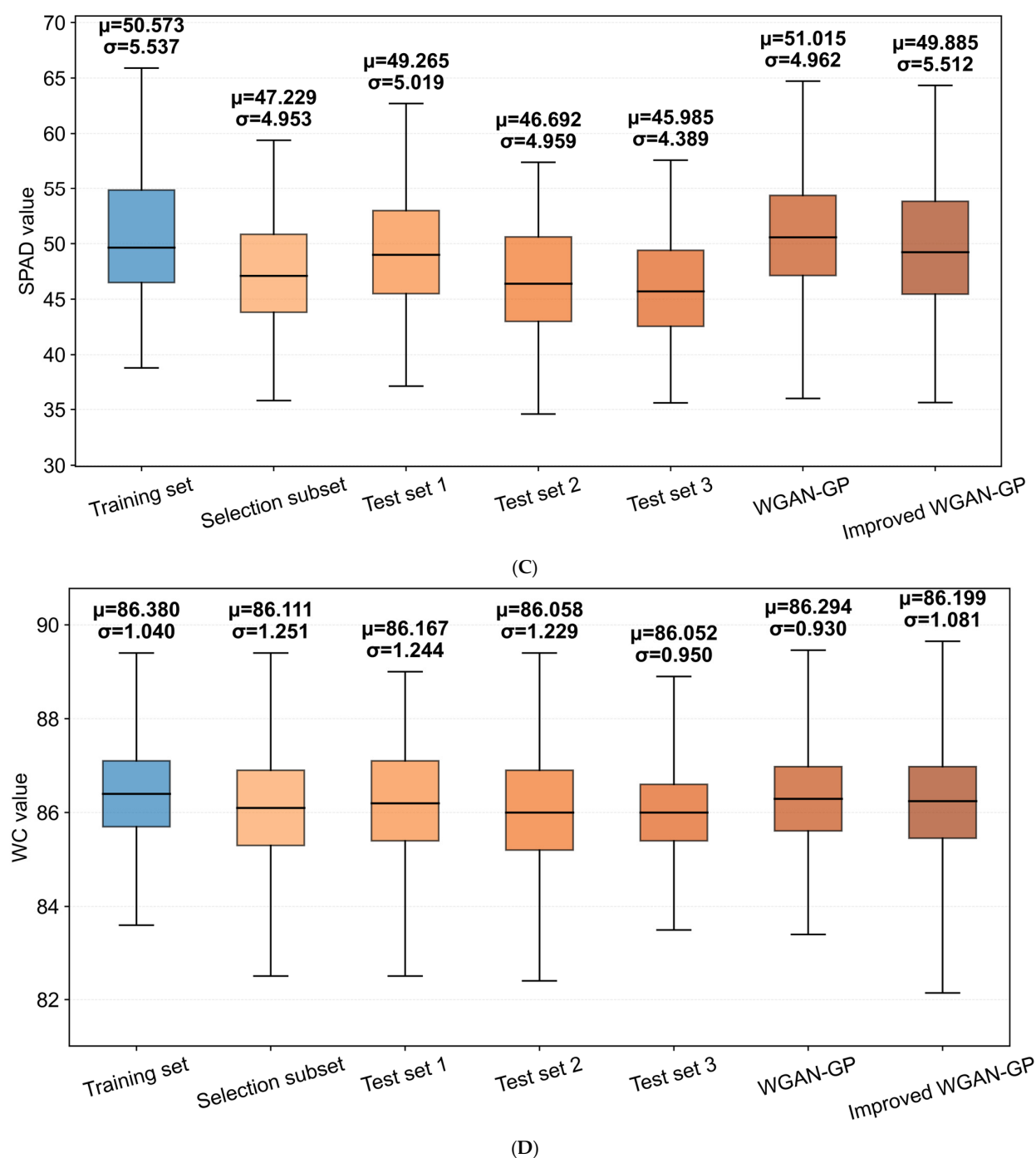


Figure 6. Box plot analysis of chemical values for real and generated data. (A) Single-task for SPAD (B) Single-task for WC (C) Multi-task for SPAD (D) Multi-task for WC.

3.2.3. T-SNE Evaluation Results

Figure 7 presents the t-SNE visualization of the selected generated spectra. The generated samples showed partial local overlap with the corresponding real samples, while some distribution shifts and differences in sample density were also observed. These results provide a qualitative visualization of the local relationship between the real and generated spectra [43].

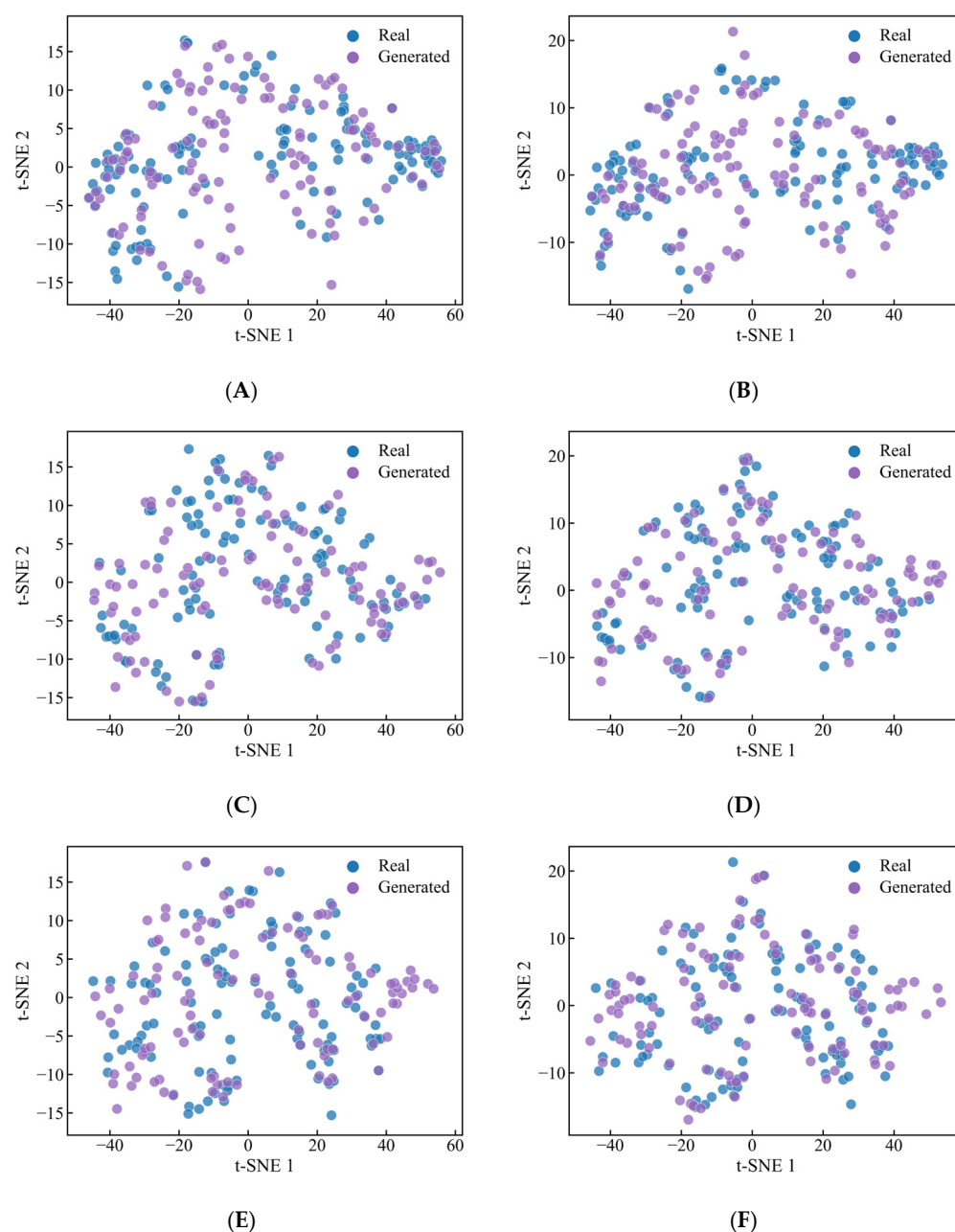


Figure 7. T-SNE result analysis chart. (A) Single-task (classification) Class 0 (B) Multi-task Class 0 (C) Single-task (classification) Class 1 (D) Multi-task Class 1 (E) Single-task (classification) Class 2 (F) Multi-task Class 2.

3.3. Prediction Performance and Component Analysis

All final prediction results reported in Tables 3 and 4 were calculated using the three 80% test subsets retained from Days 2–4. The combined 20% selection-validation set was used only for generated-sample screening and is reported in the tables to transparently describe the model-development process. The three test sets were not involved in generated-sample screening or model optimization. Nevertheless, because the selection-validation set and Test sets 1–3 were derived from the same acquisition days, these results should be interpreted as evaluation within the present four-day experimental design rather than as independent generalization to completely unseen acquisition days or batches.

Table 3. Results of single-task modeling for oilseed rape leaves under different single-task WGAN-GP data augmentation strategies.

Model	Generate Method	Training Set			Selection Subset			Test Set 1			Test Set 2			Test Set 3		
		Acc.	r _c	RMSEC	Acc.	r _v	RMSEV	Acc.	r _p	RMSEP	Acc.	r _p	RMSEP	Acc.	r _p	RMSEP
PLS-DA (Variety)	None *	0.9360	/	/	0.5792	/	/	0.6475	/	/	0.6630	/	/	0.4672	/	/
	Improved WGAN-GP	0.9317	/	/	0.6089	/	/	0.6978	/	/	0.6740	/	/	0.6100	/	/
PLSR (SPAD)	None	/	0.8250	3.1286	/	0.5733	5.5786	/	0.6301	4.3530	/	0.6918	5.3218	/	0.5936	6.2201
	WGAN-GP	/	0.8470	2.8258	/	0.6871	4.9012	/	0.6712	4.2395	/	0.6907	5.5145	/	0.5927	5.4852
	WGAN-GP + PLSR	/	0.8708	2.7203	/	0.7017	4.5946	/	0.6757	4.1290	/	0.7123	5.2061	/	0.6063	5.0378
	Improved WGAN-GP	/	0.8940	2.5851	/	0.7031	4.4551	/	0.6893	3.9527	/	0.7371	4.9229	/	0.6316	4.6303
PLSR (WC)	None	/	0.8938	0.4663	/	0.6882	0.9523	/	0.6723	0.9671	/	0.7875	0.7655	/	0.6803	0.7802
	WGAN-GP	/	0.8963	0.4667	/	0.7280	0.9132	/	0.6947	0.9685	/	0.7640	0.8086	/	0.7092	0.8182
	WGAN-GP + PLSR	/	0.9154	0.4062	/	0.7358	0.6699	/	0.7050	0.9509	/	0.7817	0.7750	/	0.7416	0.6714
	Improved WGAN-GP	/	0.9198	0.4262	/	0.7865	0.7872	/	0.7627	0.8324	/	0.7880	0.7602	/	0.7842	0.6028

*: None represents modeling with raw original data. Acc. stands for classification accuracy. r_c, r_v, r_p are the correlation coefficients for the training set, selection subset and test set. RMSEC, RMSEV, RMSEP are the RMSE values for the training set, selection subset and test set. WC stands for water content.

Table 4. Results of multi-task modeling for oilseed rape leaves under different multi-task WGAN-GP data augmentation strategies.

Model	Generate Method	Training Set			Selection Subset			Test Set 1			Test Set 2			Test Set 3		
		Acc.	r _c	RMSEC	Acc.	r _v	RMSEV	Acc.	r _p	RMSEP	Acc.	r _p	RMSEP	Acc.	r _p	RMSEP
Multi-PLS (Variety)	None *	0.9128	/	/	0.5594	/	/	0.6187	/	/	0.6447	/	/	0.4942	/	/
	WGAN-GP	0.9099	/	/	0.6040	/	/	0.6151	/	/	0.6703	/	/	0.5174	/	/
	Improved WGAN-GP	0.9070	/	/	0.6188	/	/	0.7122	/	/	0.7033	/	/	0.6178	/	/
Multi-PLS (SPAD)	None	/	0.8256	3.1237	/	0.5699	5.7891	/	0.6358	4.3595	/	0.6905	5.5253	/	0.5947	6.5296
	WGAN-GP	/	0.8139	2.8799	/	0.6486	5.0228	/	0.6767	4.4028	/	0.6939	5.4889	/	0.5859	5.6688
	Improved WGAN-GP	/	0.8756	2.5808	/	0.6625	4.6030	/	0.6569	4.0653	/	0.7162	4.9536	/	0.6126	4.5985
Multi-PLS (WC)	None	/	0.8951	0.4635	/	0.6727	0.9944	/	0.6873	0.9324	/	0.7712	0.7955	/	0.6527	0.8383
	WGAN-GP	/	0.8898	0.4500	/	0.7479	0.8355	/	0.6797	0.9568	/	0.7681	0.7879	/	0.7301	0.6685
	Improved WGAN-GP	/	0.9130	0.4308	/	0.7654	0.8277	/	0.7224	0.9279	/	0.7726	0.7836	/	0.7376	0.6720

*: None represents modeling with raw original data. Acc. stands for classification accuracy. r_c, r_v, r_p are the correlation coefficients for the training set, selection subset and test set. RMSEC, RMSEV, RMSEP are the RMSE values for the training set, selection subset and test set. WC stands for water content.

3.3.1. Results of Variety Classification

When only the original Day 1 samples were used, the single-task PLS-DA model achieved a training accuracy of 0.9360, whereas the accuracies on Test sets 1–3 were 0.6475, 0.6630, and 0.4672, respectively. The substantially lower accuracies on the three test sets, particularly Test set 3, indicate that spectral differences among the daily acquisition batches weakened variety classification performance. After augmentation with samples generated by the improved WGAN-GP, the single-task classification accuracies increased to 0.6978, 0.6740, and 0.6100 on Test sets 1–3, respectively. The accuracy of all test sets was improved, and the most remarkable improvement was observed in Test Set 3, with an accuracy increase of 0.1428.

A similar pattern was observed for multi-task classification. Without data augmentation, the multi-PLS model achieved accuracies of 0.6187, 0.6447, and 0.4942 on the three test sets. The basic WGAN-GP only yielded limited and inconsistent accuracy improvements across different Test sets. In contrast, the improved WGAN-GP increased the accuracies to 0.7122, 0.7033, and 0.6178, respectively. These results indicate that the complete augmentation framework produced more consistent classification improvements across the three test sets than the basic WGAN-GP.

Class-specific evaluation further demonstrated that data augmentation improved the overall variety classification performance in both single-task and multi-task prediction, although the magnitude of improvement varied among classes and test sets. For single-task classification, the detailed Precision, Recall, and F1-score values for Classes 0–2 on Test sets 1–3 are presented in Table S1, and the corresponding normalized confusion matrices before and after improved WGAN-GP augmentation are shown in Figure S2. For multi-task classification, the class-specific results under no augmentation, WGAN-GP augmentation, and improved WGAN-GP augmentation are provided in Table S2, with the corresponding normalized confusion matrices shown in Figure S3; overall, the improved WGAN-GP yielded better class-level classification performance across the three test sets, with particularly evident improvements for several difficult-to-classify cases.

3.3.2. Results of SPAD Prediction

For single-task SPAD prediction, the model trained only with the original Day 1 samples obtained correlation coefficients of 0.6301, 0.6918, and 0.5936 and RMSEP values of 4.3530, 5.3218, and 6.2201 on Test sets 1–3, respectively. The basic WGAN-GP produced uneven changes among the test sets. Although the RMSEP values decreased on Test sets 1 and 3, the correlation coefficient and RMSEP on Test set 2 did not improve compared with the non-augmented model. Compared to the basic WGAN-GP, adding the PLSR consistency loss increased the correlation coefficients and decreased RMSEP values. With the complete improved WGAN-GP framework, the correlation coefficients further increased to 0.6893, 0.7371, and 0.6316, while the RMSEP values decreased to 3.9527, 4.9229, and 4.6303. Compared to the non-augmented model, the complete framework reduced the RMSEP on all three test sets, with the largest absolute reduction observed on Test set 3, from 6.2201 to 4.6303.

In multi-task SPAD prediction, the non-augmented model obtained correlation coefficients of 0.6358, 0.6905, and 0.5947 and RMSEP values of 4.3595, 5.5253, and 6.5296. The improved WGAN-GP reduced the RMSEP values to 4.0653, 4.9536, and 4.5985, respectively, and increased the corresponding correlation coefficients to 0.6569, 0.7162, and 0.6126. The improvement was again most evident for Test set 3.

3.3.3. Results of WC Prediction

For single-task water-content prediction, the original PLSR model achieved correlation coefficients of 0.6723, 0.7875, and 0.6803 and RMSEP values of 0.9671, 0.7655, and 0.7802 on Test sets 1–3, respectively. The effect of the basic WGAN-GP was inconsistent. It increased the correlation coefficient on Test sets 1 and 3, but the RMSEP values were not consistently reduced. After the PLSR consistency loss was added, the correlation coefficients increased from 0.6947, 0.7640, and 0.7092 for the basic WGAN-GP to 0.7050, 0.7817, and 0.7416, respectively. The corresponding RMSEP values decreased from 0.9685, 0.8086, and 0.8182 to 0.9509, 0.7750, and 0.6714. This result again demonstrates that the PLSR consistency loss improved the reliability of the generated spectrum–label pairs. The complete improved WGAN-GP achieved the best overall test performance, with correlation coefficients of 0.7627, 0.7880, and 0.7842 and RMSEP values of 0.8324, 0.7602, and 0.6028. Compared with the non-augmented model, the largest improvement was obtained on Test set 3, where the correlation coefficient increased from 0.6803 to 0.7842 and the RMSEP decreased from 0.7802 to 0.6028.

For multi-task water-content prediction, the improved WGAN-GP obtained correlation coefficients of 0.7224, 0.7726, and 0.7376 and RMSEP values of 0.9279, 0.7836, and 0.6720 on Test sets 1–3, respectively. These results were generally better than those of the non-augmented and basic WGAN-GP models, although the magnitude of improvement differed among acquisition days.

Overall, the results show that basic adversarial generation alone did not guarantee improvement across all test sets. The PLSR consistency loss contributed to improved regression performance by strengthening spectrum–label correspondence. Meanwhile, the complete improved WGAN-GP provided further overall improvements across the three test sets.

4. Discussion

Although this study improved WGAN-GP for single-task and multi-task phenotypic data augmentation of oilseed rape leaves, several limitations remain. First, the quality of the generated data strongly depends on the quality and representativeness of the original training samples. If the original samples do not adequately cover the variation ranges of SPAD, water content, class labels, and spectral features, the generator may produce less reliable samples in sparsely represented regions. Consequently, some generated samples with broader physicochemical ranges were removed during screening. Second, the original models showed different prediction performance among the test sets collected on different days, indicating that acquisition timing, individual plant variation, and physiological status affected spectral consistency. Generated samples provided useful supplementary information but could not completely eliminate these acquisition-related differences. Third, generated-sample screening relied on the selection-validation set constructed from Days 2–4. Although the selection-validation set and Test sets 1–3 contained no overlapping samples, they originated from the same acquisition days. Therefore, the present results should be interpreted within this specific experimental design and should not be considered independent validation on completely unseen acquisition days or batches. Fourth, the PDC target ranges were empirically specified with reference to the overall physicochemical variation observed during the experiment and physiologically reasonable ranges reported in related studies. Therefore, these settings should be regarded as empirical distribution-control parameters rather than universally applicable physiological thresholds, and future studies should define such parameters using training data or external prior knowledge independently of the evaluation data. Fifth, SPAD and water content showed relatively weak correlation in this study, increasing the difficulty of generating samples with consistent

spectrum–SPAD–water content–class relationships. In addition, the influence of different generated-sample quantities was not investigated, and the number of selected generated samples was fixed to equal the number of real training samples. And it should also be noted that the initial data-quality-control procedure was performed before dataset partitioning. Although the exclusion criteria were used solely for measurement-quality control and were not tuned according to subsequent model performance, the use of physicochemical measurements at this stage may have influenced the composition of the subsequently formed subsets. The reported evaluation should be interpreted as applying to the quality-controlled sample population used in this study.

Future work can be improved in several aspects. First, the sample size should be expanded to include more varieties, growth stages, stress conditions, and acquisition dates, thereby improving the coverage of the original data distribution. Second, more controllable generative models, such as conditional WGAN-GP, continuous conditional GANs, and diffusion models, should be explored to enable targeted spectral generation under specified SPAD, water content, or class conditions. Third, transfer learning and calibration-transfer strategies can be combined with data augmentation to reduce spectral differences among acquisition periods. Fourth, a more comprehensive evaluation system for generated samples should be established by jointly considering spectral similarity, label consistency, distribution distance, and downstream prediction performance. These improvements may help build more robust hyperspectral phenotyping models and provide technical support for oilseed rape leaf variety identification and physiological-trait prediction under complex field and multi-date monitoring conditions. And further studies should evaluate the method using independently collected datasets covering additional varieties, growth stages, seasons, locations, and crop species before broader transferability can be concluded. In addition, future studies should further evaluate this issue by defining exclusion criteria exclusively from the training data or by conducting sensitivity analyses without outlier removal.

5. Conclusions

This study developed an improved WGAN-GP-based data augmentation framework for hyperspectral prediction of detached oilseed rape leaves collected over four consecutive days in both single-task and multi-task settings, covering variety identification, SPAD prediction, and water-content prediction. For the classification task, the spectra of different varieties were generated using a single WGAN-GP model, followed by probabilistic class assignment through PCA-GMM. This design reduced the number of class-specific generative models and avoided repeated training. For the regression tasks, a PLSR consistency loss was introduced to strengthen the correspondence between the generated spectra and their physicochemical labels, thereby improving the reliability of synthetic samples for predictive modeling. In addition, a physicochemical distribution constraint loss was used to regulate the mean and dispersion of the generated labels and to broaden the coverage of the continuous target variables.

Through the combined effects of these constraints, the complete improved WGAN-GP provided overall performance improvements across the three test sets compared with the basic WGAN-GP. The generated-sample screening procedure further removed synthetic samples that adversely affected model performance, and the selected augmented training data improved prediction performance across the three test sets. The multi-task framework enabled the same set of generated samples to support one classification task and two regression tasks simultaneously. Although the single-task regression models performed better for some individual metrics, the multi-task approach provided greater computational integration for unified data augmentation and simultaneous prediction.

The findings of this study are limited to detached-leaf data from three oilseed rape varieties collected at one experimental site during one growing season over four consecutive days. Moreover, because the selection-validation and test subsets were derived from the same Days 2–4, the present evaluation does not constitute independent validation on completely unseen acquisition periods. Further external validation involving additional varieties, growth stages, locations, seasons, and independently collected batches is needed to more fully assess the generalizability and practical scope of the proposed method.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/agriculture16171825/s1>, Figure S1: (A) Spectral curves of all samples on Day 1. (B) Spectral curves generated by the improved WGAN-GP for single-task classification. (C) Spectral curves generated by WGAN-GP for single-task SPAD regression. (D) Spectral curves generated by the improved WGAN-GP for single-task SPAD regression. (E) Spectral curves generated by WGAN-GP for single-task WC regression. (F) Spectral curves generated by the improved WGAN-GP for single-task WC regression. (G) Spectral curves generated by WGAN-GP for multi-task prediction. (H) Spectral curves generated by the improved WGAN-GP for multi-task prediction; Figure S2: Normalized confusion matrices of single-task variety classification before and after improved WGAN-GP augmentation. (A–C) Classification results obtained using the original training data on Test sets 1–3, respectively; (D–F) classification results obtained using the training data augmented by the improved WGAN-GP on Test sets 1–3, respectively; Figure S3: Normalized confusion matrices of multi-task variety classification after WGAN-GP and improved WGAN-GP augmentation. (A–C) Classification results obtained using the training data augmented by the WGAN-GP on Test sets 1–3, respectively; (D–F) classification results obtained using the training data augmented by the improved WGAN-GP on Test sets 1–3, respectively; Table S1: Class-specific classification performance of the single-task PLS-DA model before and after improved WGAN-GP augmentation; Table S2: Class-specific classification performance of the multi-task PLS model under different WGAN-GP augmentation strategies.

Author Contributions: Q.Z.: Methodology, Software, Investigation, Data curation, Formal analysis, Visualization, Writing—original draft, Writing—review and editing. S.S.: Investigation, Formal analysis, Writing—review and editing. G.Y.: Investigation, Formal analysis, Writing—review and editing. B.J.: Investigation, Formal analysis, Writing—review and editing. J.S.: Investigation, Formal analysis, Writing—review and editing. L.L.: Resources, Conceptualization, Funding Acquisition, Writing—review and editing. C.Z.: Conceptualization, Visualization, Methodology, Validation, Project administration, Supervision, Funding acquisition, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Huzhou Public Welfare Project (Grant Number: 2024GZ31).

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Conflicts 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

1. Abbadi, A.; Leckband, G. Rapeseed breeding for oil content, quality, and sustainability. *Eur. J. Lipid Sci. Technol.* **2011**, *113*, 1198–1206. [\[CrossRef\]](#)
2. Liersch, A.; Bocianowski, J.; Nowosad, K.; Mikolajczyk, K.; Spasibionek, S.; Wielebski, F.; Matuszczak, M.; Szala, L.; Cegielska-Taras, T.; Sosnowska, K.; et al. Effect of Genotype x Environment Interaction for Seed Traits in Winter Oilseed Rape (*Brassica napus* L.). *Agriculture* **2020**, *10*, 607. [\[CrossRef\]](#)
3. Mason, A.S.; Snowdon, R.J. Oilseed rape: Learning about ancient and recent polyploid evolution from a recent crop species. *Plant Biol.* **2016**, *18*, 883–892. [\[CrossRef\]](#)
4. Li, Y.W.; Wu, X.Y.; Xu, W.Z.; Sun, Y.D.; Wang, Y.; Li, G.J.; Xu, P. High-Throughput physiology-based stress response phenotyping: Advantages, applications and prospective in horticultural plants. *Hortic. Plant J.* **2021**, *7*, 181–187. [\[CrossRef\]](#)

5. Xiao, Q.L.; Bai, X.L.; Zhang, C.; He, Y. Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review. *J. Adv. Res.* **2022**, *35*, 215–230. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Yang, W.N.; Feng, H.; Zhang, X.H.; Zhang, J.; Doonan, J.H.; Batchelor, W.D.; Xiong, L.Z.; Yan, J.B. Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Mol. Plant* **2020**, *13*, 187–214. [\[CrossRef\]](#)
7. Song, P.; Wang, J.L.; Guo, X.Y.; Yang, W.N.; Zhao, C.J. High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *Crop J.* **2021**, *9*, 633–645. [\[CrossRef\]](#)
8. Dmitriev, P.A.; Alexandrov, A.A.; Poltavskiy, A.V.; Dmitrieva, A.A.; Kozlovsky, B.L. Identifying maple species using an artificial neural network and synthetic spectral profiles on hyperspectral images of their leaves. *Int. J. Remote Sens.* **2025**, *46*, 6826–6846. [\[CrossRef\]](#)
9. Raghav, M.; Dubey, A.; Singh, J. Hyperspectral Imaging for Detection and Classification of Plant Primary and Secondary Metabolites: A Review. *Phytochem. Anal.* **2026**, *37*, 4–27. [\[CrossRef\]](#)
10. Ulas, F.; Lahlali, R.; Laasli, S.E.; Aasim, M.; Dababat, A.; Sameeullah, M.; Imren, M. Artificial intelligence applications in abiotic and biotic plant stress management: A comprehensive bibliometric and literature review. *Physiol. Mol. Plant Pathol.* **2026**, *142*, 103098. [\[CrossRef\]](#)
11. Das, M.; Yeo, W.S.; Saptoro, A. A review of machine learning in hyperspectral imaging for food safety. *Vib. Spectrosc.* **2025**, *139*, 103828. [\[CrossRef\]](#)
12. Khan, A.; Vibhute, A.D.; Mali, S.; Patil, C.H. A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. *Ecol. Inf.* **2022**, *69*, 101678. [\[CrossRef\]](#)
13. Wang, C.Y.; Liu, B.H.; Liu, L.P.; Zhu, Y.J.; Hou, J.L.; Liu, P.; Li, X. A review of deep learning used in the hyperspectral image analysis for agriculture. *Artif. Intell. Rev.* **2021**, *54*, 5205–5253. [\[CrossRef\]](#)
14. Paulus, S.; Mahlein, A.K. Technical workflows for hyperspectral plant image assessment and processing on the greenhouse and laboratory scale. *GigaScience* **2020**, *9*, giaa090. [\[CrossRef\]](#)
15. Hume, R.; Marschner, P.; Mason, S.; Schilling, R.K.; Liu, H.J.; Jewell, N.; Brien, C.J.; Mosley, L.M. Assessing wheat responses to soil acidification and liming using hyperspectral imaging. *Plant Soil* **2025**, *511*, 925–943. [\[CrossRef\]](#)
16. Pandey, P.; Ge, Y.F.; Stoerger, V.; Schnable, J.C. High Throughput In vivo Analysis of Plant Leaf Chemical Properties Using Hyperspectral Imaging. *Front. Plant Sci.* **2017**, *8*, 1348. [\[CrossRef\]](#)
17. Zhai, Y.N.; Wang, J.; Zhou, L.; Zhang, X.C.; Ren, Y.; Qi, H.N.; Zhang, C. Simultaneously predicting SPAD and water content in rice leaves using hyperspectral imaging with deep multi-task regression and transfer component analysis. *J. Sci. Food Agric.* **2025**, *105*, 554–568. [\[CrossRef\]](#)
18. Cherif, E.; Feilhauer, H.; Berger, K.; Dao, P.D.; Ewald, M.; Hank, T.B.; He, Y.H.; Kovach, K.R.; Lu, B.; Townsend, P.A.; et al. From spectra to plant functional traits: Transferable multi-trait models from heterogeneous and sparse data. *Remote Sens. Environ.* **2023**, *292*, 113580. [\[CrossRef\]](#)
19. Rehman, T.U.; Jin, J. Deep adversarial domain adaptation for hyperspectral calibration model transfer among plant phenotyping systems. *Biosyst. Eng.* **2022**, *224*, 246–258. [\[CrossRef\]](#)
20. Feng, X.P.; Zhan, Y.H.; Wang, Q.; Yang, X.F.; Yu, C.L.; Wang, H.Y.; Tang, Z.Y.; Jiang, D.A.; Peng, C.; He, Y. Hyperspectral imaging combined with machine learning as a tool to obtain high-throughput plant salt-stress phenotyping. *Plant J.* **2020**, *101*, 1448–1461. [\[CrossRef\]](#)
21. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144. [\[CrossRef\]](#)
22. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative Adversarial Networks An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [\[CrossRef\]](#)
23. Qi, H.; Huang, Z.; Jin, B.; Tang, Q.; Jia, L.; Zhao, G.; Cao, D.; Sun, Z.; Zhang, C. SAM-GAN: An improved DCGAN for rice seed viability determination using near-infrared hyperspectral imaging. *Comput. Electron. Agric.* **2024**, *216*, 108473. [\[CrossRef\]](#)
24. Xia, Y.; Cheng, X.; Hu, X. Soil organic matter content prediction in tobacco fields based on hyperspectral remote sensing and generative adversarial network data augmentation. *Comput. Electron. Agric.* **2025**, *233*, 110164. [\[CrossRef\]](#)
25. Tan, H.; Hu, Y.; Ma, B.; Yu, G.; Li, Y. An improved DCGAN model: Data augmentation of hyperspectral image for identification pesticide residues of Hami melon. *Food Control* **2024**, *157*, 110168. [\[CrossRef\]](#)
26. Liu, T.; Han, X.; Yin, Y.; Xi, G.; Zhang, Z.; Sun, J.; Chen, G.; Zhang, L.; Han, L. Multi-property prediction of waterlogged archaeological wood based on Wasserstein GAN-augmented tree models. *J. Cult. Herit.* **2025**, *76*, 86–98. [\[CrossRef\]](#)
27. Liu, Z.; Chen, Q.; Li, J.; Liu, H.; Shuai, B.; Ouyang, L.; Bian, Z.; Wang, J.; Yu, J.; Qian, J. Multi-indicator prediction of kiwifruit quality integrating hyperspectral analysis and WGAN-GP-MMoE framework. *Food Control* **2026**, *182*, 111854. [\[CrossRef\]](#)
28. Wu, T.; Zhang, H.; Xiao, Y.; Tang, Z.; Yang, F.; Qi, H.; Wang, D.; Zhang, C. Quantification of multiple elements in Anji white tea using hyperspectral imaging combined with machine learning regression. *J. Food Compos. Anal.* **2025**, *142*, 107520. [\[CrossRef\]](#)

29. Zhang, C.; Li, C.; He, M.; Cai, Z.; Feng, Z.; Qi, H.; Zhou, L. Leaf water content determination of oilseed rape using near-infrared hyperspectral imaging with deep learning regression methods. *Infrared Phys. Technol.* **2023**, *134*, 104921. [\[CrossRef\]](#)
30. Huang, P.; Yuan, J.F.; Yang, P.; Xiao, F.T.; Zhao, Y.P. Nondestructive Detection of Sunflower Seed Vigor and Moisture Content Based on Hyperspectral Imaging and Chemometrics. *Foods* **2024**, *13*, 1320. [\[CrossRef\]](#)
31. Bro, R.; Smilde, A.K. Principal component analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [\[CrossRef\]](#)
32. Sun, H.R.; Wu, G.X.; Hua, Z.L.; Yang, H.C.; Zhang, B.Z.; Zhao, J.M.; Deng, L.M. Prediction of soybean fatty-acid composition from hyperspectral imaging with spectral feature processing and structured tabular modeling. *Food Chem.* **2026**, *515*, 149249. [\[CrossRef\]](#)
33. Mas-Gómez, J.; Rubio, M.; Dicenta, F.; Martínez-García, P.J. Unlocking almond breeding for nutritional composition with hyperspectral imaging. *Plant Phenomics* **2026**, *8*, 100239. [\[CrossRef\]](#)
34. Li, H.; Liu, Q.; Han, Z.; Lu, T.; Li, Z. MoistureVision: Rapid non-destructive prediction of cotton seed moisture using hyperspectral imaging and machine learning. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2026**, *352*, 127511. [\[CrossRef\]](#)
35. Burnett, A.C.; Anderson, J.; Davidson, K.J.; Ely, K.S.; Lamour, J.; Li, Q.; Morrison, B.D.; Yang, D.; Rogers, A.; Serbin, S.P.; et al. A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *J. Exp. Bot.* **2021**, *72*, 6175–6189. [\[CrossRef\]](#) [\[PubMed\]](#)
36. De Silva, A.L.; Trueman, S.J.; Kämper, W.; Wallace, H.M.; Nichols, J.; Hosseini Bai, S. Hyperspectral imaging of adaxial and abaxial leaf surfaces as a predictor of macadamia crop nutrition. *Plants* **2023**, *12*, 558. [\[CrossRef\]](#)
37. Fraley, C.; Raftery, A.E. Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **2002**, *97*, 611–631. [\[CrossRef\]](#)
38. Wu, Q.Y.; Zhu, Z.Y.; Wu, J.M.; Xin, X. A Dataset Representativeness Metric and A Slicing Sampling Strategy for the Kennard-Stone Algorithm. *Chem. J. Chin. Univ.-Chin.* **2022**, *43*, 20220397. [\[CrossRef\]](#)
39. Brereton, R.G.; Lloyd, G.R. Partial least squares discriminant analysis: Taking the magic away. *J. Chemom.* **2014**, *28*, 213–225. [\[CrossRef\]](#)
40. Shen, L.Z.; Gao, M.F.; Yan, J.W.; Li, Z.L.; Leng, P.; Yang, Q.; Duan, S.B. Hyperspectral Estimation of Soil Organic Matter Content using Different Spectral Preprocessing Techniques and PLSR Method. *Remote Sens.* **2020**, *12*, 1206. [\[CrossRef\]](#)
41. Rimal, R.; Almoy, T.; Sæbo, S. Comparison of multi-response prediction methods. *Chemom. Intell. Lab. Syst.* **2019**, *190*, 10–21. [\[CrossRef\]](#)
42. Bec, K.B.; Grabska, J.; Huck, C.W. Near-Infrared Spectroscopy in Bio-Applications. *Molecules* **2020**, *25*, 2948. [\[CrossRef\]](#)
43. Hu, H.; Mei, Y.; Zhou, Y.; Zhao, Y.; Fu, L.; Xu, H.; Mao, X.; Huang, L. Optimizing starch content prediction in kudzu: Integrating hyperspectral imaging and deep learning with WGAN-GP. *Food Control* **2024**, *166*, 110762. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.