RECOGNITION OF DROUGHT STRESS IN MILLET ON HYPERSPECTRAL IMAGING
/
基于高光谱成像技术识别谷子干旱胁迫

Rongxia WANG[1)], Jiarui ZHANG[1)], Jianyu CHEN[1)], Yuyuan MIAO[1)], Jiwan HAN[1)], Lijun CHENG[1)]
[1)] Shanxi Agricultural University, College of Software / China
Tel: +86 16634257091; E-mail: wangrongxia@stu.sxau.edu.cn
Correspondent author: Jiwan HAN, Lijun CHENG
DOI: https://doi.org/10.35633/inmateh-74-62

Keywords: *hyperspectral imaging, drought stress, characteristic wavelengths, image features*

ABSTRACT

Millets are one of China's primary traditional food crops, and drought can adversely impact their yield and quality. To quickly detect the degree of drought stress in cereal grains, this study establishes a nondestructive classification model based on hyperspectral imaging technology. The raw spectral data underwent preprocessing using six pretreatment methods and various combinations of these methods. Subsequently, three distinct algorithms were employed for feature wavelength selection. To assess the severity of drought stress on millet, classification models were developed by integrating texture and color features, utilizing Support Vector Machine (SVM), Partial Least Squares Discriminant Analysis (PLS-DA), and Multilayer Perceptron (MLP) algorithms. The results indicate that the D1st-SVM model, based on CARS wavelength selection, exhibits the highest modeling performance when feature wavelengths are fused with significant texture and color variables, achieving an accuracy rate of 93%. These findings suggest that drought identification in millet can be performed quickly and nondestructively by integrating image features through hyperspectral imaging technology.

摘要

谷子是我国传统主要粮作之一，干旱对其的产量和品质都会产生不利影响。为了快速检测谷子受干旱胁迫程度，本研究基于高光谱成像技术建立一种无损的分类模型。本研究对原始光谱数据进行六种预处理方法，以及这些方法的不同组合，对光谱数据中的噪声进行处理。采用 3 种不同算法进行特征波长的选取。融合纹理特征和颜色特征，基于支持向量机（support vector machine，SVM）、最小二乘判别分析（partial least-squares discriminant analysis，PLS-DA）和多层感知机算法（Multilayer Perceptron，MLP）建立分类模型，来识别谷子受干旱胁迫程度。结果表明特征波段融合重要纹理特征、重要颜色特征变量时，基于 CARS 波长选择的 D1st-SVM 模型的建模性能最高，预测集的分类准确度为 93%。研究结果表明，利用高光谱成像技术融合图像特征可以快速、无损地识别谷子是否受到干旱胁迫。

INTRODUCTION

In northern China, precipitation levels are low, with uneven seasonal distribution and significant inter-annual variability, leading to the predominance of semi-arid and arid areas. Millet is one of the primary cereal crops in these dry and semi-arid regions, characterized by its drought tolerance, high water and fertilizer utilization rates, broad adaptability, and strong resistance to adverse conditions *(Wang et al., 2022)*. Despite these attributes, drought stress often threatens millet's growth and development, which significantly impacts yield *(Yang et al., 2022)*. Drought is one of the most common and severe abiotic stress factors, adversely affecting the growth and productivity of many plants in dry and semi-arid regions (Hussain et al., 2018). As one of the crops with the highest water utilization rates *(Zou et al., 2019)*, millet's photosynthesis mechanism can suffer irreversible damage due to drought stress *(Gerhards et al., 2019)*, inhibiting root system growth and reducing the efficiency of nutrient absorption and utilization, ultimately leading to slower growth and decreased yield. Therefore, accurately identifying the degree of drought stress in millet, supplying water effectively, and rationally managing water resources are crucial. In recent years, remote sensing technology *(Chen et al., 2022)* and near-infrared spectroscopy *(Wan et al., 2020)*, among other advanced technologies, have played significant roles in detecting drought stress by assessing efficient water utilization by crops and enabling rational irrigation. However, natural conditions and spectral resolution limitations make it challenging to capture the comprehensive changes in crops under drought stress. Consequently, high spectral imaging technology, which combines imaging and spectroscopy *(Mansoor et al., 2024)*, has become a current research hotspot.

In high-throughput plant phenotyping platforms, hyperspectral imaging technology is extensively employed for the non-destructive and close-range assessment of plant physiological traits *(Mohd Asaari et al., 2022)*. Hyperspectral imaging technology can detect subtle changes in plant responses to abiotic stressors, such as reduced crop growth and stomatal closure resulting from drought stress *(Mansoor et al., 2024; Barradas et al., 2021)*. This technology also minimizes plant damage and reduces chemical pollution *(Gerhards et al., 2019)*. It enables rapid prediction of metabolite profiles in crop leaves and accurate classification of drought-affected crops using spectral data *(C A B et al., 2021)*. Hyperspectral imaging technology has been extensively utilized to analyze crop drought stress. For example, using hyperspectral characteristics, *Zhou et al. (2021)* applied hyperspectral imaging as a high-throughput phenotyping method to detect drought stress in citrus trees early. *Ioannis et al. (2024)* utilized hyperspectral imaging technology to detect drought stress in broccoli within agricultural environments. However, these methods primarily rely on single spectral data analysis, limiting models' adaptability and decision-making capabilities in practical applications *(Xu et al., 2022)*. Therefore, integrating multiple types of information can enhance model performance and robustness. Fusing image features with hyperspectral data offers distinct classification, recognition, and model optimization advantages. *Dong et al. (2015)* improved the accuracy of wheat variety classification by combining hyperspectral imaging with image feature extraction techniques. Hyperspectral imaging can obtain color features that are correlated with specific characteristics *(Alessandro et al., 2024). Abdullah et al. (2024)* established a rice classification model through the integration of spectral and color features, thereby attaining accurate discrimination between viable and non-viable rice seeds. *Qiao et al. (2024)* optimized feature variables in the corn kernel moisture content prediction model by fusing color features and texture information using hyperspectral imaging technology, thus improving prediction accuracy. These studies indicate that hyperspectral imaging technology, combined with image feature fusion, holds great potential for identifying plant drought stress, providing insights into plant responses to water scarcity *(Mansoor et al., 2024).*

This study uses millet as the research object and uses a high-throughput plant phenotyping platform to collect spectral data. The spectral data is preprocessed, and feature wavelengths are selected. By combining high-spectral imaging technology with machine learning methods, the image, and spectral features are fused to classify and identify millet drought stress, and an optimal classification model is constructed to realize real-time identification of drought stress degree in millet fields, aiming to provide a fast and non-destructive detection method for millet drought stress identification.

MATERIALS AND METHODS

Materials

The experiment was conducted in the phenotyping platform greenhouse at Shanxi Agricultural University, with an average temperature of 32.5 to 34.6°C and an average humidity of 51.3 to 53.8% RH. The materials were provided by Shanxi Agricultural University and included drought-resistant varieties of millet, specifically "Dao Ba Qi" (B15), "Xiao Miao Jin" (B19), and "Yin Tian Han" (B74), along with sensitive varieties "Liao Gu No. 1" (B294) and "Ji Gu 28" (B354). These varieties were planted in pots in the same batch and under identical conditions, using a uniformly mixed substrate of nutrient soil and vermiculite in a 3:1 ratio. Each variety was assigned three gradients and three replicate groups, resulting in 45 pots of millet plants cultivated.

Drought-stress experiments were initiated one week after the plants reached the four-leaf and one-heart stages. Soil moisture content was monitored using soil moisture sensors, and the drought stress state of the millet was further controlled using the weighing method. Three different drought treatment levels were established:

- no drought stress (soil moisture content at 65% ± 2%)
- moderate drought stress (soil moisture content at 40% ± 2%)
- severe drought stress (soil moisture content at 30% ± 2%)

The remaining fertilization schedules, methods, and management measures were maintained consistently. Prior to the manifestation of noticeable traits in the moderately drought-stressed plants, hyperspectral data were collected weekly, resulting in 315 data samples.

Hyperspectral acquisition system

Spectral data were acquired using a high-throughput crop phenotyping automatic extraction system, with the hyperspectral component of this system illustrated in Fig.1. This component comprises a hyperspectral camera, six halogen lamps (with a power of 500 W), a whiteboard, belts, a conveyor belt, a computer, and a three-phase asynchronous motor connected by circular wiring, among other components.

The spectrometer operates within a spectral range of 400 to 1,000 nm (ImSpector V10E, SPECIM, Finland), encompassing 224 distinct wavelength bands. It features a spectral resolution of 5.5 nm and a spatial pixel dimension of 1024 pixels. The system operates within an enclosed darkroom environment to minimize external light interference and enhance the quality of the collected data. The acquired data are stored in binary data stream format.

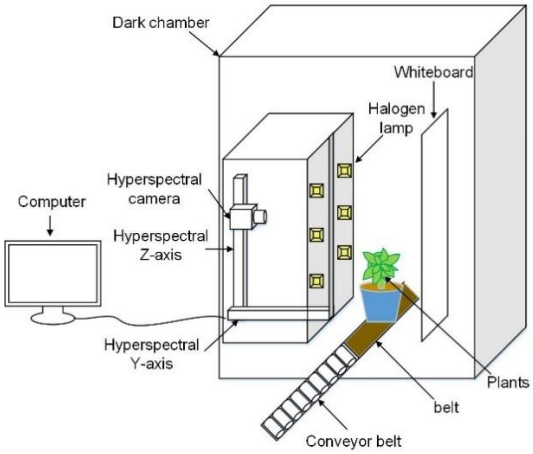


Fig. 1 - Hyperspectral part of the high-throughput automatic crop phenotype extraction system

Before collecting hyperspectral images, the system needs to be preheated for 30 minutes to eliminate the influence of uneven light and dark current on image quality *(Jia et al., 2020)*. Moreover, the relevant parameters of the system are set: the camera's exposure time is set to 11.6 ms, the platform's moving speed is set to 15 mm/s-1, and the object distance is set to 120 cm. During the data collection procedure, the whiteboard has to be calibrated to lower the complexity of calculations and the noise produced during the spectral data collection *(Pouria et al., 2021)*. In terms of a specific operation, first cover the lens cap, collect a dark current spectral data I_d, then open the lens cap and collect a whiteboard spectral data I_r and get the reflectance I of the corrected rice image according to formula 1:

$$I = \frac{I_r - I_d}{I_w - I_d} * 100 \quad (1)$$

where: I denotes the reflectance of the corrected rice image; I_r denotes the reflectance of the original image; I_d indicates the dark current-corrected image; and I_w signifies the whiteboard-corrected image.

The system was calibrated at one-hour intervals throughout the experiment.

Image Feature Extraction

The texture features of grains are extracted using the gray-level co-occurrence matrix (GLCM). In GLCM, the frequency of specific gray-level appearance between each pair of pixels is statistically calculated based on the given displacement direction and distance in an image, and this statistical information is integrated into a matrix *(Prasad et al., 2022)*. In this study, the distance is set to 1, and the energy, entropy, contrast (Contrast), correlation, and homogeneity are calculated at four different directions of 0°, 45°, 90°, and 135°.

Color features are extracted by creating a high-quality actual color (RGB) synthesis image using the red (625.67 nm), green (547.99 nm), and blue (470.31 nm) channels. The HSV color space provides an intuitive representation of color properties *(Yang et al.,2022)*. The RGB color information is mapped to the HSV color space, where the H, S, and V channels' first-order, second-order, and third-order moments are calculated. Color moments serve as a method for expressing color features; since color information is primarily concentrated in the lower-order moments *(Yang et al.,2022)*, the use of first-order, second-order, and third-order moments is sufficient to represent color distribution in digital images *(Jiang et al., 2022)*. The first-order moment indicates the average color, the second-order moment represents the dispersion of color, and the third-order moment reflects the skewness of color.

Hyperspectral image acquisition and preprocessing

From each sample dataset, a total of 224 images can be extracted, with segmentation performed to identify the region of interest (ROI) within the images, and the average spectral reflectance for this region calculated. Specifically, the grayscale image at 676.56 nm, which delineates the plant contours, is first extracted from the binary data.

This is due to the observation that in the spectral images of millet, the grayscale intensity initially deepens before lightening. At the 676.56 nm wavelength, the grayscale image reaches its darkest state, resulting in the most pronounced contrast between the plant and the background at this wavelength. Subsequently, a threshold segmentation method separates the millet plants from the background, producing a binary image. Image masking techniques are then applied to isolate the plant region in the image, designating it as the region of interest (ROI). Finally, the average reflectance of all pixels within the ROI is calculated to obtain the spectral reflectance, as illustrated in Fig. 2.

A range of preprocessing algorithms was utilized to mitigate the impacts of light scattering, baseline drift, and other confounding factors on the samples *(Zhang et al., 2021)*. These algorithms include first derivative (D1st), second derivative (D2nd), standard normalized variate (SNV), multiplicative scatter correction (MSC), Savitzky-Golay filtering (SG), and detrending. Furthermore, by the specific characteristics of each algorithm, SNV was integrated with detrending, while SG was paired with both D1st and D2nd to facilitate more effective preprocessing of the original spectrum.

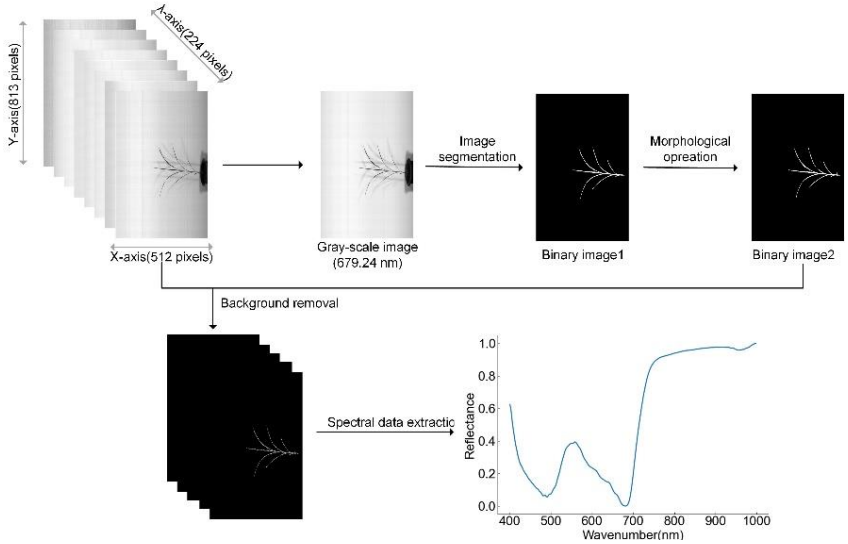


Fig. 2 - Spectral reflectance extraction process

Feature wavelength selection

Hyperspectral images encompass a substantial volume of spectral data, characterized by significant collinearity and redundant information; thus, selecting relevant feature wavelengths is imperative. The methodologies employed for the selection of feature wavelengths include Competitive Adaptive Reweighted Sampling (CARS), Successive Projections Algorithm (SPA), and Variable Iterative Space Shrinkage Approach (VISSA). CARS is a feature selection technique that integrates Monte Carlo sampling with partial least squares regression coefficients *(Li et al., 2009)*, facilitating the identification of the most pertinent spectral features while streamlining the data processing workflow. SPA sequentially selects projection directions that best preserve the key characteristics of the original data, mapping high-dimensional data into a lower-dimensional subspace to achieve dimensionality reduction *(Milanez et al., 2017)*. This process preserves critical and distinct data features while effectively capturing differences among various categories. VISSA assesses feature importance by evaluating the contributions of individual features to the spectral data, thereby selecting optimal wavelength combinations *(Zhang et al., 2020)*. This approach aids in identifying features that are rich in information and significance, while also accounting for inter-feature correlations.

Classification model

This study conducts a comparative analysis of three classification algorithms: Support Vector Machine (SVM), Partial Least Squares Discriminant Analysis (PLS-DA), and Multilayer Perceptron (MLP) to identify the optimal model for classifying millet based on varying levels of drought severity. SVM is a machine learning approach that can be employed in classification, regression, and outlier detection *(Kuswidiyanto et al., 2023)*. This research uses a polynomial kernel function to map the input data into a higher-dimensional feature space, where an optimal hyperplane is established to segregate data from distinct categories effectively. Additionally, five-fold cross-validation is implemented, with optimization performed on both the penalty factor and kernel parameters; specifically, the penalty factor is varied within a range of 0 to 100.

PLS-DA conducts rotation and projection of the data through Partial Least Squares, effectively handling the problem of multicollinearity among features and enhancing classification accuracy *(Allen et al., 2017)*. During the modeling process, the number of principal components is selected within the range of 5 to 30; too few may result in insufficient information that affects classification accuracy, while too many can lead to overfitting. Therefore, selecting an optimal number of principal components that retain critical information is essential for achieving effective classification *(Feng et al., 2024)*. MLP can achieve good prediction and classification performance on unseen data. It adapts to tasks of varying scales and complexities by adjusting both the number of layers and the number of neurons in each layer *(Saeideh et al., 2017)*. It enhances model performance and generalization capabilities by employing various activation functions, regularization parameters, and other techniques. MLP demonstrates notable proficiency in nonlinear modeling for classification tasks.

The quality of classification models is evaluated by accuracy. The calculation formula is shown in Equation (2), which represents the proportion of correctly classified samples out of the total sample number.

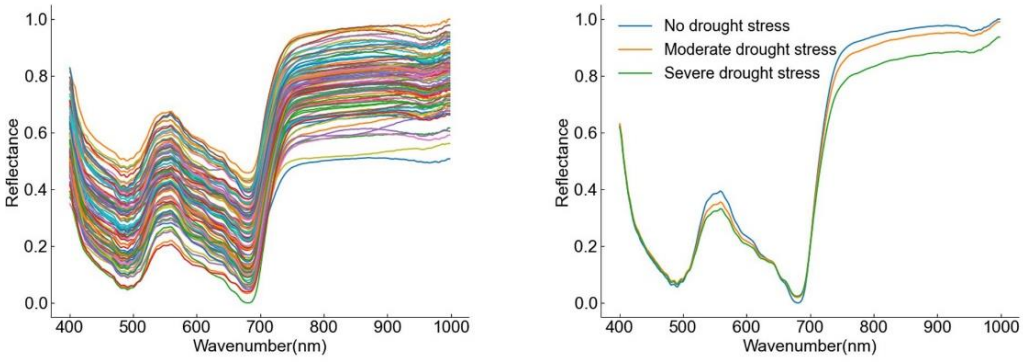
$$Accuracy = \frac{TP + TN}{TP + FN + TN + FN} * 100 \quad (2)$$

where: TP represents the number of samples predicted to be positive and they are also positive; TN represents the number of samples predicted to be negative and they are also negative; FP stands for the number of samples predicted to be positive, but they are negative; while FN represents the number of samples predicted to be negative, but they are positive.

RESULTS

Spectrum characteristics analyses

The spectral reflectance of millet plant samples shows similar trends in the wavelength range of 400 to 1,000 nm (Fig. 3-a). The reflectance trends of all samples within this range are generally similar. Specifically, within the 540 to 590 nm range, a prominent absorption peak forms due to chlorophyll's relatively weak light absorption in the green light region (520 to 600 nm). This spectral band corresponds to lower photosynthesis efficiency in plants, resulting in higher spectral reflectance in this range *(Zhao et al., 2016)*. Conversely, within the 660 to 710 nm range, an absorption trough appears due to chlorophyll's strong light absorption efficiency in the red edge region (630 to 690 nm). This area represents a spectral band with higher photosynthesis efficiency, enabling plants to exhibit vigorous photosynthetic activity *(Zhao et al., 2016)*. Starting from the 690 nm range, the reflectance sharply increases with wavelength until reaching a maximum near 760 nm, forming the plant-specific red edge position.



(a) Raw data (b) Average reflectance of different drought stresses

Fig. 3 - Characterization curves of spectrum of millet samples

To analyze the impact of drought severity on millet spectra, the average spectral curves of millet under no drought stress, moderate drought stress, and severe drought stress conditions are considered representatives of different drought severities. The overall trends of the spectral curves for millet under various drought severities are similar, as shown in Fig. 3 (b). However, significant differences in spectral curves are observed at the peak wavelength range of 540-590 nm, the trough wavelength range of 660-710 nm, and the high reflectance region of 730-1000 nm. Spectral reflectance decreases with increasing drought severity within the wavelength ranges of 540-590 nm and 730-1000 nm. This is attributed to the reduced water content in the millet, which affects the transmission and scattering of light, subsequently leading to a decline in spectral reflectance. Conversely, the reflectance increases with drought severity in the wavelength range of 660-

710nm, which falls within the red edge region where chlorophyll absorption in millet is strongest. This is due to decreased chlorophyll content under drought conditions, increasing reflectance. These spectral differences demonstrate the feasibility of utilizing hyperspectral imaging technology to identify the drought severity in millet.

Spectral data preprocessing results

After applying various preprocessing techniques and their combinations, significant variations were observed in the spectral curves, as illustrated in Fig. 4. Specifically, following the application of MSC and SNV preprocessing (Fig. 4-a, b), the spectral data exhibited increased concentration, with both methods yielding similar trends in the spectral curves. This similarity can be attributed to the shared objective of both preprocessing methods: standardizing the spectral morphology across all samples to a unified baseline and amplitude, thereby mitigating non-chemical variations within the spectral data. After applying SG filtering (Fig. 4-c), the curve appeared smoother compared to its original counterpart, with a notable reduction in fluctuations and noise within the dataset, facilitating subsequent analytical procedures. The spectral curve resulting from D1st processing (Fig. 4-d) revealed pronounced peaks in the 510-540 nm and 700-780 nm wavelength bands compared to the original spectrum, aiding in feature extraction and analysis within the spectral dataset. Following D2nd preprocessing (Fig. 4-e), significant fluctuations were observed in the spectral curve, highlighting the high sensitivity of this method to changes within the spectrum and its ability to accentuate subtle variations present. Upon Detrend from the original spectral data (Fig. 4-f), the initially observed concave feature within the 660-710 nm range became more pronounced, and a distinct downward trend was evident in the 750-1000 nm range. This preprocessing technique effectively eliminated trends or background components from the spectrum, thereby redirecting analytical focus towards the intrinsic fluctuation characteristics of each spectrum and enhancing both analytical accuracy and reliability.

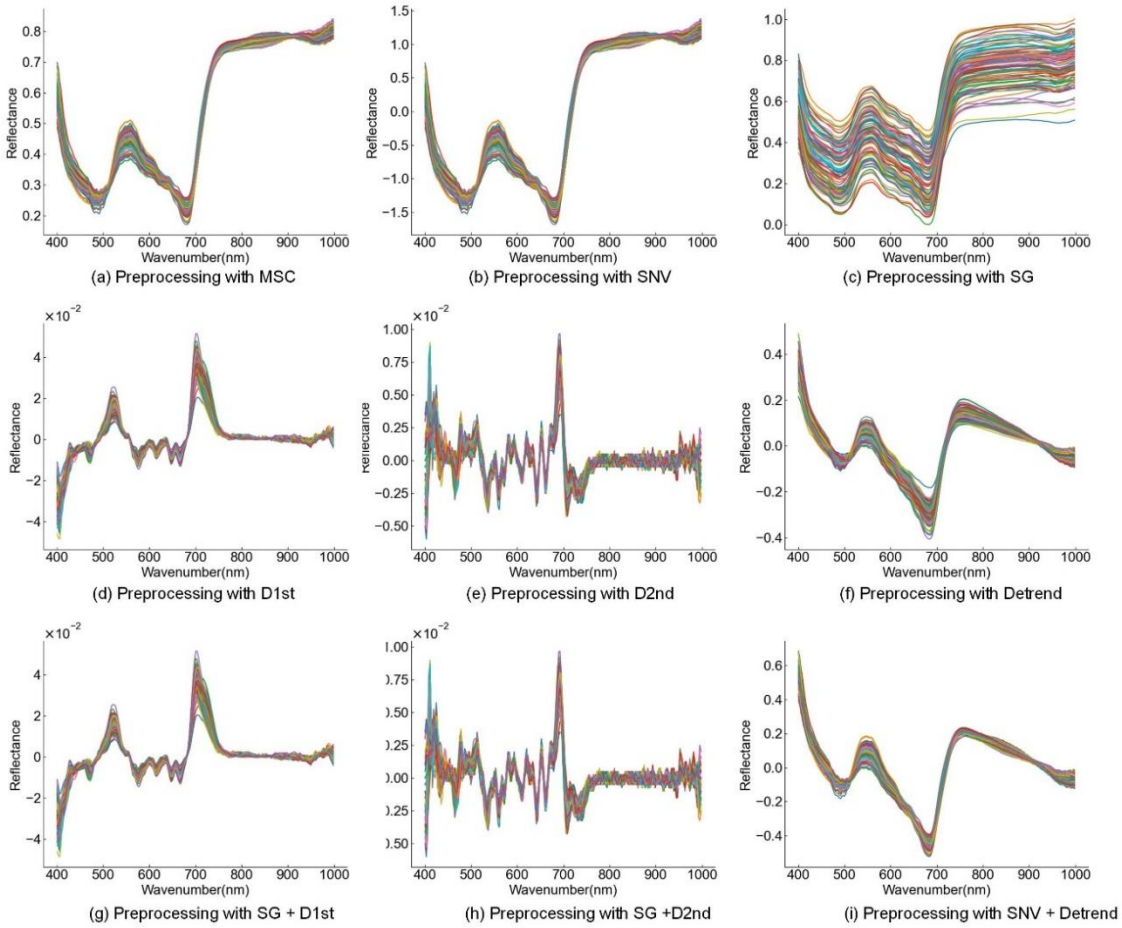


Fig. 4 - Spectral curves of different preprocessing methods

Two distinct combinations are employed to process the spectral data by integrating the characteristics of various preprocessing methods. The spectral data subjected to SG filtering is utilized for D1st calculation (Fig. 4-g), resulting in both smoothing and enhancement of local features within the spectral dataset. When SG filtering is combined with D2nd processing (Fig. 4-h), it not only smooths the data but also accentuates

details within the spectral curve, thereby improving the accuracy of subsequent classification efforts. Additionally, when SNV is integrated with detrending (Fig. 4-i), the results indicate that the spectral curve retains consistency with trends observed post-detrending while exhibiting a more concentrated distribution. This combination effectively corrects systematic errors in the spectral data while mitigating effects introduced by detrending.

Modeling analysis based on full wavelengths

Utilizing data that has undergone preprocessing with varying techniques as input can have a notable impact on the outcomes of modeling exercises. Table 1 compares the performance of different preprocessing methods within diverse modeling contexts. When employing raw, unprocessed data (RAW) for modeling, the MLP model exhibited optimal performance, achieving an accuracy rate of 81.3% in the predictive set classification task. A comparative analysis of model classification performance across various preprocessing methods reveals that spectral data processed using D1st, SNV, Detrend, and the combined SNV-Detrend techniques demonstrate superior modeling outcomes. In the context of the SVM models, applying D1st processing led to the highest performance, with a predictive set accuracy of 86.0%.

Furthermore, implementing the D1st processing technique in both the PLS-DA and the MLP models resulted in accuracy rates surpassing those of other preprocessing methods, precisely 79.0% and 84.8%, respectively. This underscores the advantage of the D1st processing method across diverse modeling paradigms. By emphasizing pivotal features within the original data, the D1st processing technique substantially enhances classification accuracy. Consequently, the decision was made to advance with spectral data processing and analysis based on the 1Dst processing methodology.

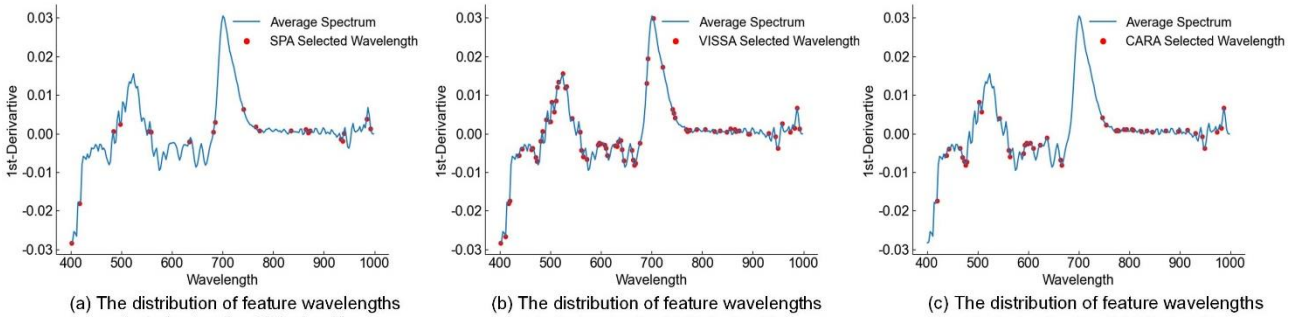
Table 1

Classification results of millet drought degree based on Full spectrum

	Training set accuracy(%)			Test set accuracy(%)		
	SVM	PS-DA	MLP	SVM	PLS-DA	MLP
Raw	91.9	80.4	87.9	81.3	77.9	82.5
D1st	89.9	91.4	91.4	86.0	79.0	84.8
D2nd	81.4	79.3	81.4	80.2	74.4	79
SNV	84.9	79.8	90.4	84.8	79.0	83.7
MSC	89.4	84.2	84.9	82.5	76.7	77.9
Detrend	86.4	85.9	89.4	82.5	79.0	82.5
SG	87.4	80.4	83.4	76.7	70.9	73.2
SG-D1st	89.9	93.9	86.4	80.2	75.5	75.5
SG-D2nd	75.3	79.8	74.8	70.9	66.2	70.9
SNV-Detrend	89.4	83.4	87.9	84.8	77.9	83.7

Feature wavelength selection results

This study employed three distinct algorithms, CARS, SPA, and VISSA, to extract characteristic wavelengths from the preprocessed spectral data. The precise distribution of the selected wavelengths is graphically depicted in Fig. 5.



(a) The distribution of feature wavelengths selected using the SPA algorithm

(b) The distribution of feature wavelengths selected using the VISSA algorithm

(c) The distribution of feature wavelengths selected using the CARS algorithm

Fig. 5 - Distribution of characteristic wavelengths selected by different algorithms

The SPA algorithm selects 21 bands, the VISSA algorithm selects 74 bands, and the CARS algorithm selects 70 bands. The SPA algorithm selects fewer characteristic wavelengths from the spectral data, and the distribution is more dispersed. In contrast, the VISSA and CARS algorithms concentrate their selected characteristic wavelengths in 590~680 nm and 760~1000 nm, respectively, while exhibiting a more scattered distribution in other bands.

The CARS algorithm utilizes Monte Carlo sampling iterations and an exponential decay function to adjust the selection probability of each wavelength band adaptively, ultimately selecting the optimal combination of bands that contribute most significantly to modeling. Simultaneously, the number of Monte Carlo sampling iterations is determined through cross-validation assessment to achieve optimal algorithm performance and efficiency. In the CARS analysis, the Monte Carlo runs are set to 50, with a ratio of 3:1 for the training set to the test set in each run, and a 5-fold cross-validation is employed for Root Mean Square Error of Cross-Validation (RMSECV). The process of selecting characteristic variables by CARS is illustrated in Figure 6. Specifically, Fig. 6 (a) shows the exponential decrease in characteristic wavelengths as the number of Monte Carlo sampling iterations increases. Fig. 6 (b) depicts the changes in RMSECV during the Monte Carlo sampling iterations, with the RMSECV reaching a minimum value of 0.386 at the 18th iteration. Fig. 6(c) displays the trend of changes in the regression coefficient paths during each Monte Carlo sampling process. The vertical blue line in the figure indicates that the RMSECV value is minimum at the 18th iteration, with 70 characteristic wavelengths extracted.

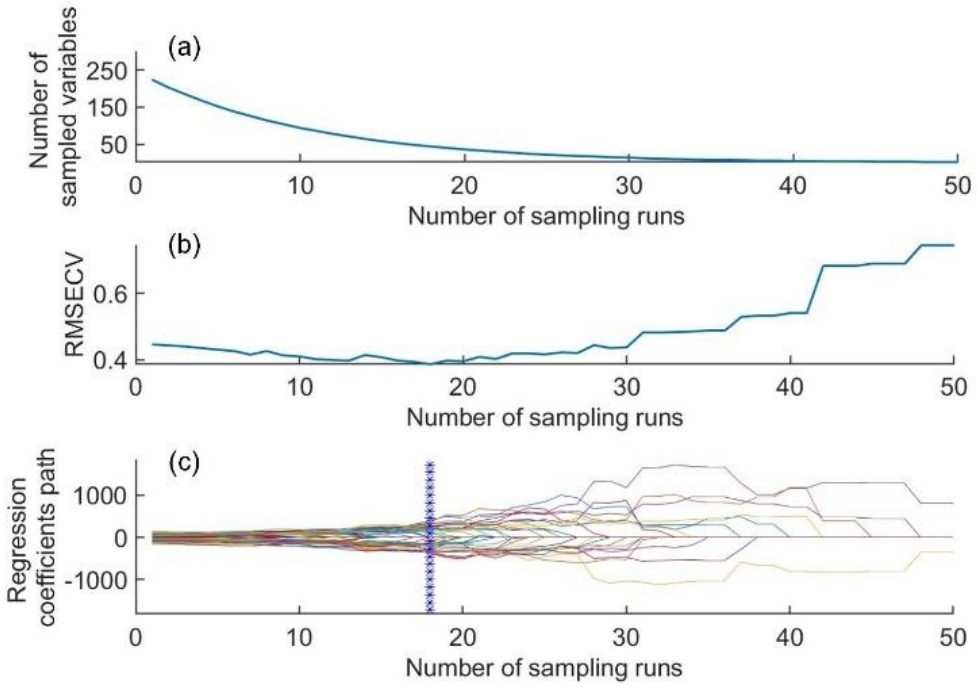


Fig. 6 - Changes of related parameters during the sampling operations of CARS

Modeling analysis based on feature wavelength

The results of modeling the important bands for different feature selection algorithms are shown in Table 2. When analyzing the data based on the entire wavelength spectrum, the SVM classification model achieved notably better results, exhibiting training set and prediction set classification accuracies of 89.9% and 86%, respectively. In comparison, the MLP classification model performed slightly worse, achieving accuracies of 88.9% on the training set and 84.8% on the prediction set. On the other hand, the PLS-DA classification model showed a high accuracy of 91.4% on the training set but struggled on the prediction set, achieving only 79% accuracy. Three feature wavelength selection algorithms were employed to extract critical wavelengths from the spectrum for modeling to enhance classification accuracy further and simplify the model. The classification model performance improved significantly after applying the CARS algorithm for feature selection, surpassing the performance of the other two algorithms. The VISSA algorithm yielded classification results similar to the CARS algorithm, but its prediction set accuracy was 2% lower. In contrast, the performance of the models declined after applying the SPA for feature selection, indicating that the SPA might not have captured the most relevant information for classification during feature extraction.

Table 2

Classification results of millet drought degree based on important spectral bands

	Training set accuracy (%)			Test set accuracy (%)		
	SVM	PS-DA	MLP	SVM	PLS-DA	MLP
Full spectrum	89.9	91.4	88.9	86	79	84.8
SPA	80.4	74.8	79.4	77.9	75.5	76.8
VISSA	86.4	83.4	88.9	87.2	77.9	83.7
CARS	91.9	88.9	87.9	87.2	79	83.7

Overall, the classification performance of the CARS-SVM and VISSA-SVM models improved compared to the original spectral modeling. Notably, the CARS-SVM model demonstrated an accuracy of 91.9% on the training set and 87.2% on the prediction set, exhibiting superior performance among all models. However, the classification accuracy of the other models, despite using selected feature wavelengths, did not match the level achieved with the original spectral data. This suggests that, while the selected feature wavelengths are representative, they do not fully encapsulate the spectral information required for precise classification of drought severity in millet.

Results of texture and color feature selection

Following the application of principal component analysis to the hyperspectral image, three principal component images were extracted, with contribution rates of 99.7%, 0.13%, and 0.11% (Fig. 7). Consequently, the optimal principal component image exhibiting a contribution rate of 99.7% was utilized for subsequent texture feature extraction. A total of 20 texture features were identified using the GLCM method. To mitigate information redundancy, correlation analyses were conducted between the extracted feature values at various angles and different drought severities in millet, ultimately resulting in the selection of seven significant texture feature variables.

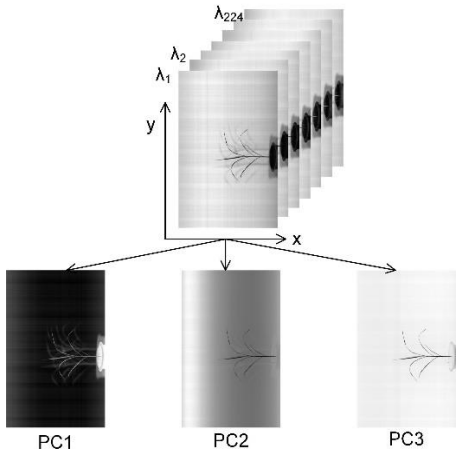


Fig. 7 - The first three principal component images of the millet samples

Color features were derived by calculating the first-order, second-order, and third-order moments of the H, S, and V channels from the RGB image, yielding a total of 9 feature values. A correlation analysis was performed to optimize further the feature combination between the color features and millet drought severity, ultimately resulting in the selection of 6 significant color feature variables. The extracted image texture information, color information, and spectral data were subsequently fused for modeling analysis.

Modeling analysis based on the fusion of spectral information and image information

Incorporating texture and color features into the spectral data significantly enhanced the classification accuracy of drought severity in millet (Table 3). Following the integration of spectral data with texture and color features, the performance of the three classification models remains broadly consistent with that achieved using only spectral feature wavelength extraction. The spectral data were fused with texture and color features, respectively, and the performance of the three classification models was consistent with that of the spectral data feature wavelength alone. However, when the spectral data was independently fused with important texture features (Texture*) and important color features (Color*), the performance of all models was slightly improved.

Specifically, adding important texture features improved the test set accuracy of the SVM model by 4.6%, the PLS-DA model by 3.5%, and the MLP model by 4.6%. The fusion of important color features improved the test set accuracy of the PLS-DA and MLP models by 2.3% and 4.6%, respectively, but had no significant impact on the performance of the SVM model.

Furthermore, integrating spectral data with important texture and color features significantly enhances the performance of classification models. Among these, the SVM model demonstrates superior classification accuracy, achieving an accuracy rate of 98.9% for the training set and 93% for the test set; closely following is the MLP model, which attains accuracy rates of 92.4% for the training set and 89.5% for the test set; in contrast, the PLS-DA model exhibits relatively lower classification accuracy, with rates of 90.9% for the training set and 83.7% for the test set. These findings indicate that integrating selected important texture and color features can effectively improve drought classification accuracy in millet.

Table 3

Classification results of millet drought degree by integrating spectrum, texture features and color features

	Training set accuracy (%)			Test set accuracy (%)		
	SVM	PS-DA	MLP	SVM	PLS-DA	MLP
Raw	91.9	80.4	87.9	81.3	77.9	82.5
D1st	89.9	91.4	91.4	86.0	79.0	84.8
CARS-D1st	91.9	88.4	87.9	87.2	79.0	83.7
Spectrum + Texture	92.9	85.4	94.4	87.2	80.2	87.2
Spectrum + Texture*	99.4	84.9	90.4	91.5	82.5	88.3
Spectrum + Color	92.9	85.4	93.4	86.0	80.2	87.2
Spectrum + Color*	91.4	87.9	88.4	87.2	81.3	88.3
Spectrum + Texture* + Color*	98.9	90.9	92.4	93.0	83.7	89.5

In conclusion, hyperspectral imaging technology effectively facilitates the acquisition of millet data across varying drought levels, thereby enabling precise classification of drought severity. Integrating spectral data with texture and color features has significantly enhanced classification accuracy, offering a rapid, non-destructive, and efficient approach for monitoring drought conditions in millet.

CONCLUSIONS

This paper employed various preprocessing algorithms and their combinations alongside different feature band selection techniques to process hyperspectral data. Concurrently, image feature information was integrated to construct a classification model to categorize millet based on varying degrees of drought stress accurately. The research findings indicate that the model integrating pre-processing, feature wavelength extraction, and machine learning algorithms is feasible for predicting the maturity of rapeseed. Among them, the 1st-CARS-SVM model, which incorporates significant texture and color features, has the optimal classification performance, with a classification accuracy of 93%, attaining an enhancement in model classification accuracy. The study validates the potential of hyperspectral imaging technology in the detection of plant drought stress.

This model offers a rapid and non-destructive approach for identifying the degree of drought stress in millet, which holds the potential to provide significant scientific support for enhancing millet yield and irrigation management. However, the model still requires improvement, such as the influence of geographical environment, different years, and growing conditions on the growth process of millet, which may result in variations in the spectral characteristics of millet. In the subsequent working process, these issues will be further taken into account to enhance the model's universality.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all authors for their support and contributions to this manuscript. This research is supported by the Shanxi Province Basic Research Program for Young Scientists (NO. 202303021222039).

REFERENCES

[1] Alessandro, M., Joby, M. P. C., Sathishkumar, S., & Robert M. (2024). Are unmanned aerial vehicle-based hyperspectral imaging and machine learning advancing crop science?[J], *Trends in plant science*, 29(2): 196-209. https://doi.org/10.1016/j.tplants.2023.09.001

[2] Allen, A., Williams, M. R., & Sigman, M. E. (2019). Application of likelihood ratios and optimal decision thresholds in fire debris analysis based on a partial least squares discriminant analysis (PLS-DA) model[J]. *Forensic Chem,* 16: 100188. https://doi.org/10.1016/j.forc.2019.100188

[3] Barradas, A., Correia, P. M. P., Silva, S., Mariano, P., Pires, M. C., Matos, A. R., da Silva, A. B., & da Silva, J. M. (2021). Comparing Machine Learning Methods for Classifying Plant Drought Stress from Leaf Reflectance Spectra in Arabidopsis thaliana[J]. *Applied Sciences*, 11(14): 6392-6392. *https://doi.org/10.3390/app11146392*

[4] Burnett, A. C., Serbin, S. P., Davidson, K. J., Ely, K. S., & Rogers, A. (2021). Detection of the Metabolic Response to Drought Stress Using Hyperspectral Reflectance[J]. *Journal of experimental botany*, 72(18): 6474-6489. https://doi.org/10.1093/jxb/erab255

[5] Chen, Y., Wang, P., Zhang, Y., & Yang, J. (2022). Comparison of Drought Recognition of Spring Maize in Northeast China Based on 3 Remote Sensing Indices[J](基于 3 种遥感指数的东北春玉米干旱识别对比). *Journal of Applied Meteorological Science*, 33(04): 466-476. https://doi.org/10.11898/1001-7313.20220407

[6] Dong, G., Guo, J., Wang, C., Chen, Z., Zheng, L., & Zhu, D. (2015). The Classification of Wheat Varieties Based on Near Infrared Hyperspectral Imaging and Information Fusion[J](基于近红外高光谱成像及信息融合的小麦品种分类研究). *Spectroscopy and Spectral Analysis*, 35(12): 3369-3374. https://doi.org/10.3964/j.issn.1000-0593(2015)12-3369-06

[7] Feng, H., Chen, Y., Song, J., Lu, B., Shu, C., Qiao, J., Liao, Y., & Yang, W. (2024). Maturity Classification of Rapeseed Using Hyperspectral Image Combined with Machine Learning[J]. *Plant phenomics (Washington, D.C.)*, 6:0139. https://doi.org/10.34133/plantphenomics.0139

[8] Gao, C., Ji, X., He, Q., Gong, Z., Sun, H., Wen, T., & Guo, W. (2023). Monitoring of Wheat Fusarium Head Blight on Spectral and Textural Analysis of UAV Multispectral Imagery[J]. *Agriculture*, 13(2):293-293. https://doi.org/10.3390/agriculture13020293

[9] Gao, S., & Xu, J.H(2022). Hyperspectral image information fusion-based detection of soluble solids content in red globe grapes[J]. *Computers and Electronics in Agriculture*, 196. https://doi.org/10.1016/j.compag.2022.106822

[10] Gerhards, M., Schlerf, M., Mallick, K., & Udelhoven, T. (2019). Challenges and Future Perspectives of Multi-/Hyperspectral Thermal Infrared Remote Sensing for Crop Water-Stress Detection: A Review [J]. *Remote Sensing*, 11(10): 1240. https://doi.org/10.3390/rs11101240

[11] Hussain, H. A., Hussain, S., Khaliq, A., Ashraf, U., Anjum, S. A., Men, S., & Wang, L. (2018). Chilling and Drought Stresses in Crop Plants: Implications, Cross Talk, and Potential Management Opportunities [J]. *Frontiers in Plant Science*, 9393. https://doi.org/10.3389/fpls.2018.00393

[12] Jia, B., Wang, W., Ni, X., Lawrence, K. C., Zhuang, H., Yoon, S.-C., & Gao, Z. (2020). Essential processing methods of hyperspectral images of agricultural and food products[J]. Chemometrics and Intelligent Laboratory Systems, 198. https://doi.org/10.1016/j.chemolab.2020.103936

[13] Jiang, X., Tian, J., Huang, H., Hu, X., Han, L., Huang, D., & Luo, H. (2022). Nondestructive visualization and quantification of total acid and reducing sugar contents in fermented grains by combining spectral and color data through hyperspectral imaging [J]. *Food Chemistry*, 386: 132779-132779. https://doi.org/10.1016/j.foodchem.2022.132779

[14] Kuswidiyanto, L. W., Kim, D. E., Fu T., Kim, K. S., &Han X. (2023). Detection of Black Spot Disease on Kimchi Cabbage Using Hyperspectral Imaging and Machine Learning Techniques [J]. *Agriculture*, 13 (12). https://doi.org/10.3390/agriculture13122215

[15] Li, H., Liang, Y., Xu, Q., & Cao, D. (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. *Analytica Chimica Acta*, 648(1): 77-84. https://doi.org/10.1016/j.aca.2009.06.046

[16] Malounas, I., Paliouras, G., Nikolopoulos, D., Liakopoulos, G., Bresta, P., Londra, P., Katsileros, A., & Fountas S. (2024). Early Detection of Broccoli Drought Acclimation/stress in Agricultural Environments Utilizing Proximal Hyperspectral Imaging and AutoML [J], *Smart agricultural technology*, 8: 100463 https://doi.org/10.1016/j.atech.2024.100463

[17] Mansoor, S., & Chung, Y. S. (2024). Functional phenotyping: Understanding the dynamic response of plants to drought stress[J]. *Current Plant Biology*, 38: 100331. https://doi.org/10.1016/j.cpb.2024.100331

[18] Milanez, K., Nóbrega, T. C. A., Nascimento, D. S., Galvao, R. K. H., & Pontes, M. J. C. (2017). Selection of robust variables for transfer of classification models employing the successive projections algorithm. *Anal. Chim. Acta* 984: 76–85. https://doi.org/10.1016/j.aca.2017.07.037

[19] Mohd Asaari, M. S., Mertens S., Verbraeken, L., Dhondt, S., Inze, D., Koirala, B., & Scheunders, P. (2022). Non-destructive analysis of plant physiological traits using hyperspectral imaging: A case study on drought stress [J]. *Computers and Electronics in Agriculture*, 195. https://doi.org/10.1016/j.compag.2022.106806

[20] Pouria, S., Nicolas, V., & Malcolm, J. H. (2021). A Neural Network Method for Classification of Sunlit and Shaded Components of Wheat Canopies in the Field Using High-Resolution Hyperspectral Imagery [J], *Remote sensing*, 13(5): 898-898. https://doi.org/10.3390/rs13050898

[21] Prasad, G., Vijay, G. S., & Kamath, C. (2022). Comparative study on classification of machined surfaces using ML techniques applied to GLCM based image features [J]. *Materials Today: Proceedings*, 62(P3): 1440-1445. https://doi.org/10.1016/j.matpr.2022.01.285

[22] Qiao, M., Xia, G., Yang, X., Cui, T., Fan, C., Li, Y., Han, S., & Jun, Q. (2024). Prediction of moisture content for single maize kernel based on viscoelastic properties: constitutive model and force-time graph [J]. *Journal of the science of food and agriculture*, 140, 104(11): 6594-6604. https://doi.org/10.1002/jsfa.13483

[23] Saeideh, F., Hossein, A. M., Abbas, R., S. Amirhassan M, & Hassan S. (2017). Identification and Classification of Three Iranian Rice Varieties in Mixed Bulks Using Image Processing and MLP Neural Network [J]. *International Journal of Food Engineering*. 13(5): 20160121-20160121. https://doi.org/10.1515/ijfe-2016-0121

[24] Siam, A. A., Salehin, M. M., Alam, M. S., Ahamed, S., Islam, M. H., & Rahman, R. (2024). Paddy Seed Viability Prediction Based on Feature Fusion of Color and Hyperspectral Image with Multivariate Analysis [J], *Heliyon*, 10(17). https://doi.org/10.1016/j.heliyon.2024.e36999

[25] Sun, J., Zhou, X., Wu, X., Zhang, X., & Li, Q. (2016). Identification of moisture content in tobacco plant leaves using outlier sample eliminating algorithms and hyperspectral data[J]. *Biochemical and Biophysical Research Communications*, 471(1): 226-232. https://doi.org/10.1016/j.bbrc.2016.01.125

[26] Wan, Y., Ouyang, J., Yuan, H., Wu, Q., Xiang, D., & Zhao, G. (2020). Effect of Drought Stress on Physiological Characteristics and Infrared Spectrum Characterization of Tartary Buckwheat [J] (干旱胁迫对苦荞生理特征和红外光谱表征特性的影响). *Journal of Chengdu University (Natural Science Edition)*, 39(03): 230-233+240. https://doi.org/10.3969/j.issn.1004-5422.2020.03.002

[27] Wang, Z., Liu, X., Yu, A., Cheng, K., Li, H., Tian, G., Wang, Y., Chen, X., Zhang, P., & Liu, H. (2022). Changes of Physiological Response to Drought Stress and Selection of Drought Resistance Indexes in Different [J] (不同谷子品种萌发期对干旱胁迫生理响应的变化及抗旱指标筛选). *Journal of Agricultural Science and Technology*, 22(12): 39-49. https://doi.org/10.13304/j.nykjdb.2019.0749

[28] Xu, P., Zhang, Y., Tan, Q., Xu, K., Sun, W., Xing, J., & Yang, R. (2022). Vigor identification of maize seeds by using hyperspectral imaging combined with multivariate data analysis[J]. *Infrared physics and technology*, 126. https://doi.org/10.1016/j.infrared.2022.104361

[29] Yang, L., & Qin, H. (2022). Study on Peanut Appearance Quality Detection Based on Color and Texture Features [J] (基于颜色和纹理特征的花生仁外观品质检测研究). *Chinese Agricultural Science Bulletin*, 38(27): 151-156. https://doi.org/10.11924/j.issn.1000-6850.casb2021-0918

[30] Yang, Y., Li, K., Wei, S., Guga, S., Zhang, J., & Wang, C. (2022). Spatial-temporal distribution characteristics and hazard assessment of millet drought disaster in Northern China under climate change [J]. *Agricultural Water Management*, 272. https://doi.org/10.1016/j.agwat.2022.107849

[31] Zhang, L., Sun, J., Zhou, X., Nirere, A., Wu, X., & Dai, R. (2020). Classification detection of saccharin jujube based on hyperspectral imaging technology [J]. *Journal of Food Processing and Preservation*, 44(8). https://doi.org/10.1111/jfpp.14591

[32] Zhang, X., Sun, J., Li, P., Zeng, F., & Wang, H. (2021). Hyperspectral detection of salted sea cucumber adulteration using different spectral preprocessing techniques and SVM method [J]. *LW T*, 152. https://doi.org/10.1016/j.lwt.2021.112295

[33] Zhao, Z.J., Shan, G.L., Duan, X.H., Jiang, H., Ren, J., Chen, G., & Chu, X.H. (2016). Study on spectral reflectance and physiological characteristics of three cool-season turfgrass under drought stress [J]. *Grassland and Turf*, 36(06): 23-29. https://doi.org/10.3969/j.issn.1009-5500.2016.06.004

[34] Zhou, J. J., Zhang, Y. H., Han, Z. M., Liu, X. Y., Jian, Y. F., Hu, C. G., & Dian, Y. Y. (2021). Evaluating the Performance of Hyperspectral Leaf Reflectance to Detect Water Stress and Estimation of Photosynthetic Capacities[J]. *Remote Sensing*, 13(11): 2160-2160. https://doi.org/10.3390/rs13112160

[35] Zou, C., Li, L., Miki, D., Li, D., Tang, Q., Xiao, L., Rajput, S., Deng, P., Peng, L., Jia, W., Huang, R., Zhang, M., Sun, Y., Hu, J., Fu, X., Schnable, P. S., Chang, Y., Li, F., Zhang, H., Feng, B., Zhu, X., Liu, R., & Schnable, J. C. (2019). The genome of broomcorn millet[J]. *Nature communications*, 10(1): 436. https://doi.org/10.1038/s41467-019-08409-5