RESEARCH ON DETECTION METHOD OF SALINITY IN SALINE-ALKALI SOILS BASED ON FUSION DATA
/
基于融合数据的盐碱地盐分检测方法研究

Qinghai HE [1,2,3,4,5], Chengli GAO [1,3,4,5,7], Xueguan ZHAO [6], Hongen GUO [3,4,5], Wendong ZHANG [3,4,5], Peng QI [3,4,5*], Xiaoli LI [2,*], Yong HE [2], Wengang ZHENG [6] and Guoqiang LIU [7], Mohamed Mahmoud IBRAHIM[8], Maher Fathy Attia MORSY[9], Hani Abdelghani MANSOUR[9],

[1] School of Mechanical Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan / China;

[2] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou / China;

[3] Shandong Academy of Agricultural Machinery Science, National Key Laboratory of Nutrients, Jinan / China;

[4] Key Laboratory of Intelligent Agricultural Equipment in Hilly and Mountainous Areas of Shandong Province, Jinan / China;

[5] Huang Huai Hai Key Laboratory of Modern Agricultural Equipment, Ministry of Agriculture and Rural Affairs, Jinan / China

[6] Beijing Research Center of Intelligent Equipment for Agriculture, Beijing / China;

[7] Shandong Institute of Mechanical Design and Research, Jinan / China;

[8] Department of Agricultural Engineering, Faculty of Agriculture, Cairo University, Cairo / Egypt

[9] Agricultural and Biological Research Institute, National Research Centre, Cairo / Egypt

E-mail: qi-peng@139.com; xiaolili@zju.edu.cn

DOI: https://doi.org/10.35633/inmateh-78-81

Keywords: *Hyperspectral imaging; machine vision; data fusion; soil salinity.*

ABSTRACT

Efficient monitoring of soil salinity is of significant importance for the utilization and ecological restoration of saline–alkali soils. This study investigates typical saline–alkali soils by employing hyperspectral imaging and machine vision technologies to develop a quantitative prediction model for soil salinity. First, spectral and image data of soil samples were acquired. Subsequently, spectral preprocessing was performed using Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), and Moving Average smoothing (MA). Characteristic spectral bands were extracted using the Competitive Adaptive Reweighted Sampling (CARS), Variable Combination Population Analysis (VCPA), Iteratively Retaining Informative Variables (IRIV), and a combined VCPA–IRIV approach. Finally, the selected spectral features were fused with image features to establish a Support Vector Regression (SVR) model. The results demonstrated that, compared with single-source data, the SVR model based on feature-level data fusion significantly improved the prediction accuracy of soil salinity. Among the tested models, the SNV + VCPA–IRIV + SVR combination achieved the best performance (Rc^2 = 0.9889, RMSEC = 0.4790, Rp^2 = 0.9569, RMSEP = 1.0484, RPD = 3.4423). The results indicate that feature-level data fusion combining hyperspectral and machine vision data can effectively predict soil salinity in saline–alkali soils, while improving model stability compared to single-data-source approaches.

摘要

高效监测土壤盐分含量对盐渍土利用与生态修复具有重要的意义。本文研究以典型盐碱地土壤为研究对象，基于高光谱与机器视觉技术构建了土壤盐分含量的定量预测模型。首先收集土样的光谱与图像信息，采用SNV、MSC 和MA 进行光谱预处理，采用CARS、VCPA、IRIV 和VCPA-IRIV 提取光谱特征波长，并结合图像特征，融合两种数据建立 SVR 模型。结果表明，相较于单一数据建立的模型，基于特征级融合建立的 SVR 模型显著提升了盐碱地土壤盐分含量的预测精度。其中，SNV+VCPA-IRIV+SVR 是最佳的算法组合(R_c^2=0.9889，RMSEC=0.4790，R_p^2=0.9569，RMSEP=1.0484，RPD=3.4423)。研究表明，基于高光谱与机器视觉的特征级数据融合方法能够有效预测盐碱地土壤盐分含量，有效解决了单一数据建模稳定性差的问题。

INTRODUCTION

China currently has approximately 100 million hm² of saline soils, which vary widely in type and are extensively distributed, posing serious constraints on local ecological and economic sustainability *(Whitney et al., 2018)*. The composition of soil salinity is influenced by a complex interplay of factors, including climate, topography and geomorphology, hydrological conditions, parent material, and biological activity *(Jia et al., 2024; Yang et al., 2023)*.

In addition, anthropogenic activities such as land overexploitation, improper agricultural irrigation, and seawater intrusion further intensify primary and secondary salinization. Therefore, efficient monitoring of soil salinity is of great significance for the sustainable utilization of saline land resources and ecological restoration *(Sun et al. 2022b; Wang et al., 2012)*. Conventional soil salinity measurement methods—such as drying and electrical conductivity analyses—are largely laboratory-based and, although accurate, are labor-intensive, time-consuming, and reliant on extensive sample collection. These limitations hinder their applicability for large-scale and rapid monitoring required in modern precision agriculture *(Liu et al., 2016)*.

In recent years, machine vision and hyperspectral imaging—two representative non-destructive detection technologies—have been widely applied in agriculture *(Allbed et al., 2014),* food science *(Chen et al., 2024a)*, and pharmaceuticals *(Song et al., 2019)*. In particular, hyperspectral imaging integrates two-dimensional imaging with spectral analysis, enabling the acquisition of reflectance or absorption information across numerous continuous bands. Each pixel contains not only spatial information but also spectral signatures over multiple wavelengths, allowing comprehensive characterization of the sample's physical structure and chemical composition *(Achata et al., 2019)*. Based on spectral data, various methods have been developed to enhance the prediction accuracy of soil constituents. A novel spectral feature extraction strategy was introduced to construct spectral subsets for effective prediction of soil organic matter (SOM) content using partial least squares regression (PLSR) *(Sun et al., 2022a)*. Multiple regression models, including Multiple Linear Regression (MLR), support vector machine (SVM), and PLSR, were evaluated for soil salinity prediction, with SVM demonstrating superior performance *(Wang et al., 2019)*. In addition, characteristic band selection algorithms were applied to develop PLSR models for soil nutrient estimation, where the beetle optimization–based regression (BOR) optimized model achieved high predictive accuracy *(Ramasamy et al., 2025)*.

Although spectral technology shows great potential for soil component analysis, reliance on a single spectral source often poses inherent limitations. In recent years, the integration of spectral techniques with multivariate chemometric methods, coupled with multi-source data fusion strategies, has emerged as an effective approach for enhancing model accuracy and robustness. Data fusion techniques integrate information from different sensors, expanding sample information dimensions while improving model generalization and detection efficiency. For instance, A data fusion approach integrating spectral and image information was applied to evaluate salmon freshness, achieving a prediction accuracy of 92.3% *(Wu et al., 2019)*. Similarly, near-infrared spectroscopy combined with electronic-nose data was used to enhance tea quality assessment, with classification accuracies reaching 99.63% and 97.00%, respectively *(Xia et al., 2024)*. However, the application of fusion strategies in quantitative soil salinity detection remains at an exploratory stage, with limited studies and no consensus regarding optimal fusion approaches or expected outcomes.

Building on existing research, this study focuses on typical saline–alkali soils in China and investigates data fusion approaches that integrate spectral features with image features derived from hyperspectral imaging. Linear PLSR and nonlinear SVR models were established using full-band, characteristic band, and fused datasets, enabling a systematic comparison of the predictive performance of different modeling strategies. The aim is to develop a modeling framework suitable for rapid, non-destructive, and high-precision soil salinity monitoring, thereby providing theoretical and technical support for the classification and utilization of saline soils in precision agriculture.

MATERIALS AND METHODS

Study Area

As shown in Figure 1, Wudi County in Binzhou City and Guangrao County in Dongying City, Shandong Province, which are located in typical saline–alkali lands, were selected as the study areas for soil salinity assessment. Wudi County in Binzhou City (37°97′ N, 117°75′ E) has a temperate continental monsoon climate with distinct seasons and marked wet–dry cycles, featuring an annual mean temperature of approximately 13.61 °C and an average annual precipitation of 621.52 mm. Guangrao County in Dongying City (37°30′ N, 118°73′ E) experiences a warm temperate monsoon climate with distinct seasons and synchronized rainfall–temperature patterns, with an annual mean temperature of 12.3 °C and an average annual precipitation of 587.4 mm. Both locations are situated within the Yellow River Delta, one of the major regions of saline–alkali soil distribution in China, characterized by high representativeness, diverse salinization types, and substantial variability in salinity levels.

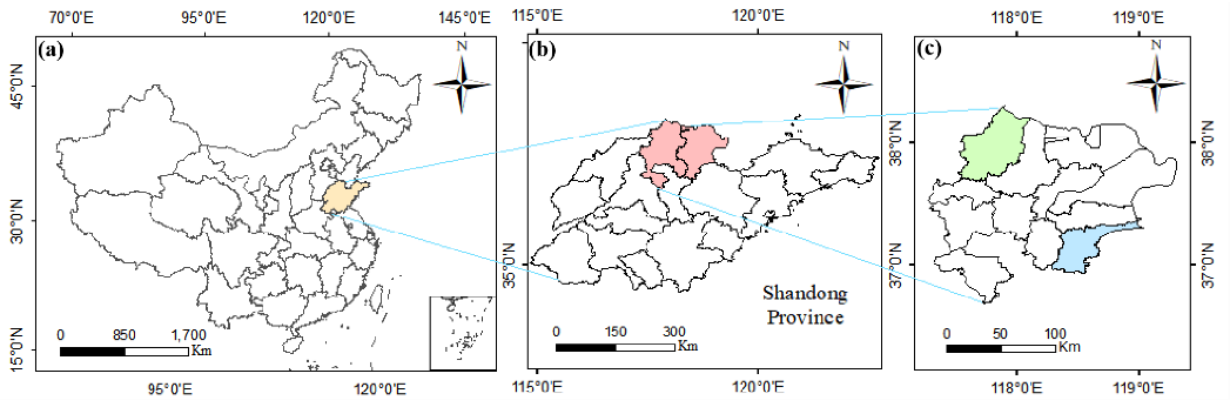


Fig. 1 - (a) China location map; (b) Location map of Shandong; (c) Location map of the study area in Binzhou and Dongying.

Sample Collection and Measurement

A five-point sampling method was employed in this study, and soil samples representing four salinization levels—Light Salinization (LS), Moderate Salinization (MS), Heavy Salinization (HS), and Salted Soil (SS). The sampling depth was set at 0–20 cm of the soil surface *(Cruz-Tirado et al., 2021)*. A total of 150 saline-alkali soil samples were collected, each weighing approximately 1 kg, covering multiple land-use types including saline–alkali cropland, saline–alkali forest–grassland, and unused saline–alkali land. The collected samples were sealed, labeled, and transported to the laboratory, where impurities such as gravel and organic debris were removed. The samples were then air-dried in a cool, dry, and ventilated environment, ground, and passed through a 100-mesh sieve. The sieved samples were divided into two portions: 60 g was used for spectral measurement, and the remaining portion was reserved for soil salinity determination.

Soil salinity was classified according to the Baosidan soil salinization classification standard, categorized as follows: LS (1.0–2.0 g/kg), MS (2.0–4.0 g/kg), HS (4.0–6.0 g/kg), and SS (>6.0 g/kg)(Bao et al., 2000). Soil salinity was determined using the gravimetric method. Results indicated a minimum salinity of 1.25 g/kg, a maximum of 19.45 g/kg, an average of 6.0 g/kg, and a standard deviation of 4.02 g/kg.

$$S_t = \frac{m_1}{m_2} \times 1000 \quad (1)$$

where m_1 is the mass of the dried residue (g), and m_2 is the mass of the oven-dried soil sample (g).

Hyperspectral Imaging Systems

The ISpecHyper-VS1000-Lab hyperspectral imaging system (Shenzhen, China) is presented in Figure 2. The spectrometer covers a spectral range of 334–986 nm with a spectral resolution of 2.5 nm, yielding a total of 300 spectral bands. Hyperspectral images of the saline–alkali soil samples were acquired using a line-scanning mode. The instrument was preheated for 30 minutes prior to data acquisition to ensure stabilization of the illumination source *(Alordzinu et al., 2021)*. During image acquisition, each soil sample was placed in a circular polyethylene Petri dish with a diameter of 10 cm and a depth of 1 cm. Excess soil was removed using a blade or plastic ruler to ensure a level and smooth sample surface. The prepared sample dish was then positioned at the center of the linear translation stage, the dark chamber door was closed, and hyperspectral image acquisition was performed via the control software. These operational steps were repeated for each sample until hyperspectral images of all 150 saline–alkali soil samples were acquired. To obtain clear and distortion-free hyperspectral images, the conveyor speed, camera exposure time, and the height between the camera lens and the sample were set to 6 mm/s, 4 ms, and 28 cm, respectively.

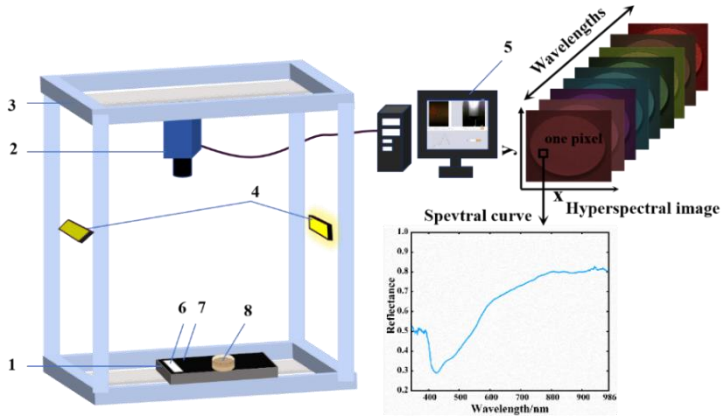


Fig. 2 - Hyperspectral imaging system.
1 – Conveyor belt; 2 – VS1000 hyperspectral camera; 3 – Black box; 4 – Light source; 5 – Computer; 6 – PTFE white board; 7 – Blackboard; 8 – Sample.

Sample Information Extraction

Spectral Feature Extraction

Using ENVI software, ten regions of interest (ROIs) were randomly selected from each hyperspectral soil image for spectral data extraction, with each ROI defined as a 1 × 1 cm square area. The average spectral value of the selected ROIs was used as the spectral data for the corresponding soil sample. Because uneven illumination intensity, camera dark current, and grating scattering can introduce deviations into the spectral data, black–white calibration was performed on the acquired raw hyperspectral images *(Sun et al., 2017)*. White calibration was conducted using a white polytetrafluoroethylene plate with a reflectance of 99.99% to obtain the standard white reference image. Dark calibration was carried out by covering the camera lens with an opaque cap to obtain the standard black reference image. Reflectance correction of the raw images was completed using the corresponding calibration formula *(Li et al., 2018)*.

$$I = \frac{I_{\text{Raw}} - I_{Dark}}{I_{White} - I_{Dark}} \tag{2}$$

where I is a hyperspectral corrected image; I_{Raw} is a hyperspectral original image; I_{Dark} is a black corrected image; and I_{White} is a white corrected image.

Image Feature Extraction

After standardized cropping of the hyperspectral images, an image processing program developed in the MATLAB GUI environment (software copyright number: 2014SR149549) was used to extract color and texture features from the soil samples. The color features included the mean values of the red (R), green (G), and blue (B) channels; the mean hue (H), saturation (S), and value (V) from the HSV color space; the excess green index (2G−R−B); the ratio of the red to green channel means (R/G); and the hue angle (hab*). The texture features included mean gray level (m), standard deviation (δ), smoothness (r), uniformity (U), and entropy (e) *(Liu et al., 2022)*. For each soil hyperspectral image, ten ROIs were randomly selected, from which image features were extracted; the mean value of these features was then computed and used as the image feature data for that sample.

Data Processing

Pre-Processing Method

During the acquisition of hyperspectral data from saline–alkali soils, the raw spectra inevitably contain unwanted information such as baseline drift, light scattering, and random noise due to the influence of the instrument system, soil samples, and environmental conditions. Therefore, spectral preprocessing is required to improve data quality and enhance modeling accuracy. The purpose of spectral preprocessing is to reduce the adverse effects caused by noise, dark current, and other external environmental factors during data acquisition, thereby decreasing spectral redundancy and improving the predictive accuracy and robustness of the model.

In this study, three commonly used preprocessing methods were applied to the acquired soil hyperspectral reflectance data: Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), and Moving Average smoothing (MA). SNV effectively reduces spectral interference caused by variations in external light-scattering intensity and enhances the correlation between the spectra and the target attributes *(Li et al., 2021)*. MSC is capable of correcting spectral variations among soil samples that arise from differences in scattering intensity *(Yang et al., 2020)*. MA reduces noise and fluctuations in the spectral data, producing smoother and more stable spectra, highlighting the major spectral features, and strengthening the association between the spectra and the corresponding data attributes *(Al-Mbaideen, 2019)*.

Characteristic Band Screening Method

The hyperspectral dataset used in this study covers a band range of 344–986 nm. Within this range, many bands are not directly related to soil salinity prediction, including wavelength intervals without pronounced absorption features, highly redundant adjacent bands, and edge bands with low signal-to-noise ratios. Such irrelevant information may reduce modeling efficiency and degrade predictive accuracy. The selection of characteristic bands enables the removal of irrelevant information while retaining those bands most strongly associated with soil salinity. This procedure not only simplifies the modeling process but also improves computational efficiency and enhances model accuracy *(He et al., 2024)*. Four characteristic band selection algorithms were employed in this study: Competitive Adaptive Reweighted Sampling (CARS), Variable Combination Population Analysis (VCPA), Iteratively Retaining Informative Variables (IRIV), and a hybrid VCPA–IRIV method.

CARS selects subsets of variables through repeated sampling and identifies the optimal subset with the lowest root mean square error based on cross-validation *(Wu et al., 2019)*. VCPA is an evolution-inspired feature selection algorithm that simulates natural selection; through iterative optimization, it retains superior variable combinations and gradually identifies the bands contributing most to model performance *(Jiang et al., 2021)*. IRIV iteratively evaluates the contribution of each variable to the model, removing redundant variables while retaining the most informative bands, thereby improving predictive capability *(Yun et al., 2014)*. The VCPA–IRIV hybrid method integrates the global search ability of VCPA with the local refinement ability of IRIV, first performing coarse screening of important bands via VCPA and then further eliminating redundant variables using IRIV. This hybrid approach exploits the strengths of both algorithms while overcoming their individual limitations *(Yang et al., 2021)*.

Data Fusion Method

Data fusion technology integrates information from multiple sources to generate outputs that are more accurate, comprehensive, and informative than those derived from a single data source. This approach not only reduces the likelihood of errors within individual datasets but also enhances data reliability and improves modeling performance *(Chen et al., 2022; Chen et al., 2024b)*. Depending on data characteristics and analytical requirements, fusion strategies can be categorized into data-level fusion, feature-level fusion, and decision-level fusion *(Yu et al., 2021)*. The overall workflow of data fusion is illustrated in Figure 3.

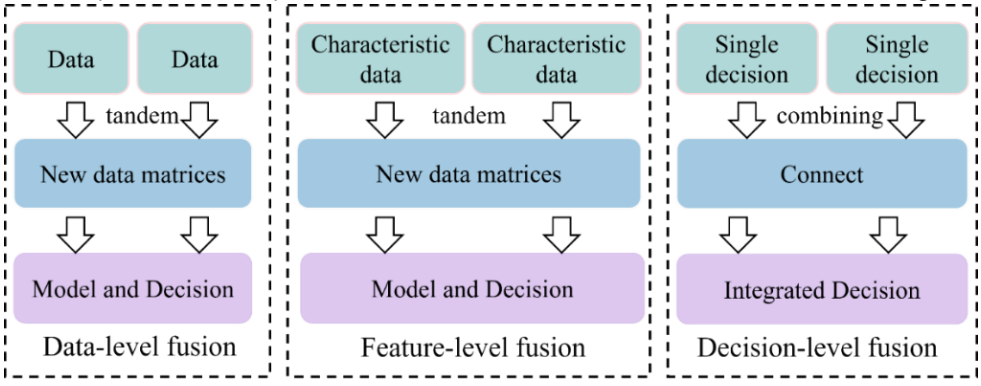


Fig. 3 - Data fusion strategy workflow

Data-level fusion, also referred to as low-level fusion, involves concatenating raw or preprocessed data from different sources to form a new composite dataset for subsequent modeling. Feature-level fusion, also known as mid-level fusion, concatenates selected feature variables from each data source after feature extraction, thereby reducing redundancy and overlapping information.

Decision-level fusion, or high-level fusion, combines independently generated results from multiple data sources according to predefined rules to obtain an integrated decision output. However, decision-level fusion may lead to the loss of important information if the correlations among results from different sources are not properly considered *(An et al., 2022)*. Therefore, data-level fusion and feature-level fusion were selected in this study to integrate spectral data with image feature data. The fused datasets were subsequently normalized and used to construct regression models for soil salinity prediction, through which the optimal fusion strategy was identified.

Model Construction and Performance Evaluation

Multiple algorithms were applied to process the spectral data, and quantitative prediction models for soil salinity in saline–alkali soils were developed using full band, characteristic band, and fused datasets. PLSR and SVR were employed to construct models and conduct comparative analyses of their predictive performance. PLSR effectively addresses challenges arising from multicollinearity by deriving latent variables that minimize the loss of variance contained in the original variable set, making it one of the most widely used linear modeling approaches *(Subi et al., 2024)*. SVR constructs models by identifying an optimal hyperplane in a high-dimensional feature space while allowing a defined margin of error, offering particular advantages for high-dimensional datasets and nonlinear relationships *(He et al., 2023)*. In this study, a radial basis function (RBF) kernel was applied in the SVR method, which was implemented using the Li-SVMLAB toolbox in the MATLAB environment. To ensure model stability, the Kennard–Stone (K–S) algorithm was used to partition all samples into a training set of 112 samples and a prediction set of 38 samples at a 3:1 ratio.

Model accuracy was evaluated comprehensively using the coefficient of determination (R^2), root mean squared error (RMSE), and the ratio of performance to deviation (RPD)*(Alsaleh et al., 2025; Wang et al., 2020)*. A higher R^2, lower RMSE, and higher RPD indicate better predictive performance, whereas the opposite reflects poorer prediction accuracy.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_i)^2}{n}} \quad (4)$$

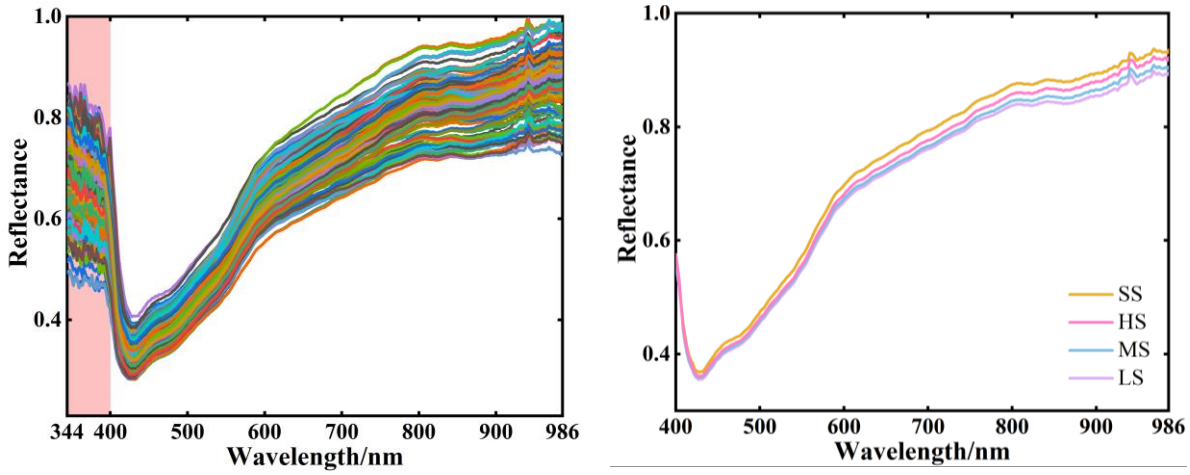
$$RPD = \frac{SD}{RMSE} \quad (5)$$

where n is the number of samples (i=1,2,3,... ,n); $y_i$and $\hat{y}_i$ are the standardized and predicted values of the sample; $\overline{y_i}$ is the mean of the sample and SD is the standard deviation of the sample.

RESULTS AND ANALYSIS

Spectral Characteristics of Soil Samples

Figure 4(a) displays the raw spectrum of soil samples, with each curve representing the spectral characteristics of a specific soil sample. As seen in the figure, the reflectance curves in the 344–400 nm range exhibit significant fluctuations and considerable overlap, indicating substantial noise interference, which may arise from instrument instability or environmental light effects. To enhance the reliability of subsequent analyses, this study excludes this range and retains only the data within the 400–986 nm range for modeling analysis. After denoising, the spectral curves exhibit an overall trend of "decrease followed by increase." In the 400–429 nm region, the spectral reflectance continuously decreases with increasing band, while in the 429–986 nm range, the spectral reflectance increases with band, initially at a higher rate, then the rate of increase diminishes, and finally stabilizes. Additionally, a significant absorption trough is observed near 429 nm, which may be attributed to the strong absorption of light at this band by certain soil components, such as iron oxides or organic matter *(Xie et al., 2013)*. A distinct absorption peak is detected at approximately 940 nm, primarily because this spectral region is highly sensitive to the presence of moisture *(Tao et al., 2009)*.



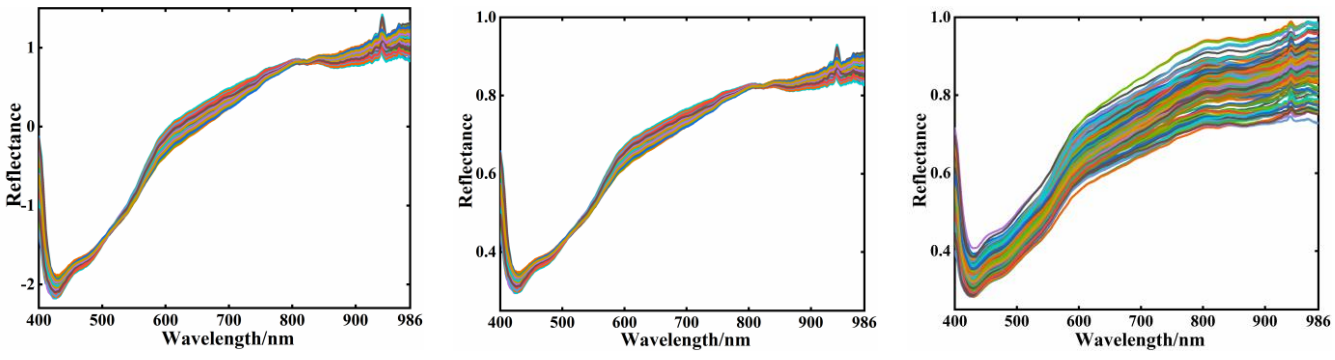
(a) (b)

Fig. 4 - (a) Original spectrum; (b) Spectral curves at different degrees of salinization

To more clearly examine the differences in spectral characteristics among soils with varying degrees of salinization, the spectral curves corresponding to each salinization level were averaged, as shown in Figure 4(b). Despite variations in salinization intensity, the averaged spectral curves of the different soil categories exhibit consistent overall shapes and similar variation trends, with the curves generally distributed in a nearly parallel manner. This indicates that the patterns of spectral reflectance are largely comparable across salinization levels, with differences primarily manifested in the magnitude of reflectance values rather than in the spectral trend. Further analysis reveals that spectral reflectance increases progressively with rising soil salinity, which is consistent with the findings reported by An et al *(Bouaziz et al., 2011; An et al., 2015)*. This phenomenon may be attributed to the enhanced light-scattering effect on the soil surface induced by higher salt content, ultimately resulting in increased overall reflectance.

Soil Salinity Prediction Model Based on Full Band

To reduce noise originating from environmental disturbances and instrument errors in the raw soil spectral data and to enhance the predictive accuracy of the models, three preprocessing techniques—Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), and Moving Average (MA)—were applied. The objective was to evaluate how different preprocessing approaches influence the performance of soil salinity prediction models. Figure 5 illustrates the spectra obtained after applying the respective preprocessing methods. Subsequently, PLSR and SVR models were developed using the raw and preprocessed spectral datasets to predict soil salinity content. The modeling results are summarized in Table 1.



(a) (b) (c)

Fig. 5 - Spectra after preprocessing: (a) Spectrum after SNV; (b) Spectrum after MSC; (c) Spectrum after MA.

Table 1

PLSR and SVR prediction models for soil salinity based on full band data

Preprocessing Methods	Models	Calibration Set		Prediction Set		RPD
		Rc^2	RMSEC	Rp^2	RMSEP	
None	PLSR	0.6788	2.3293	0.6457	2.6638	0.7568
	SVR	0.8452	1.8079	0.6916	2.9325	1.3194
SNV	PLSR	0.7801	2.0411	0.6330	2.8850	0.9839
	SVR	**0.9777**	**0.7393**	**0.9307**	**1.5115**	**2.3333**
MSC	PLSR	0.7775	2.0120	0.7723	2.2498	1.1080
	SVR	0.9692	0.7853	0.8929	1.5738	2.2478
MA	PLSR	0.7488	2.3272	0.3882	3.3069	0.9203
	SVR	0.8467	1.8982	0.7247	2.6192	1.3694

By comparing all predictive models listed in Table 1, it is observed that each preprocessed model exhibits superior performance to the model constructed from the raw spectra. The models developed using SNV and MSC demonstrate overall better predictive performance than those processed with MA. This is because, although MA can smooth high-frequency noise to some extent, it is insufficient for correcting the systematic errors introduced by scattering. In contrast, SNV and MSC more effectively compensate for scattering effects within the spectra, thereby enhancing the accuracy of the resulting models *(Zhang et al., 2023)*. The SVR model based on SNV preprocessing achieved the best performance, with an increase of 0.2391 in Rp^2, a reduction of 1.421 in RMSEP, and an improvement of 1.0139 in RPD. These results indicate that preprocessing effectively mitigates scattering effects and instrumental noise in the spectral data, improving data quality for modeling and significantly enhancing the accuracy and robustness of the prediction models.

Soil Salinity Prediction Model Based on Characteristic Bands

To reduce model complexity and enhance predictive accuracy, the full band spectra were first preprocessed, and four band selection techniques—CARS, VCPA, IRIV, and VCPA-IRIV—were subsequently employed to extract informative features. Figure 6 illustrates the distribution of characteristic bands extracted by each characteristic selection method. As shown in the figure, the VCPA-IRIV method identifies the largest number of characteristic bands across all preprocessing conditions, indicating its strong overall capability in capturing informative variables. In contrast, the IRIV method extracts the fewest bands, potentially posing a risk of insufficient information extraction. Moreover, the selected bands are predominantly concentrated around 460 nm, 620–670 nm, and 910 nm, implying that these spectral regions may exhibit strong sensitivity to variations in soil salinity.

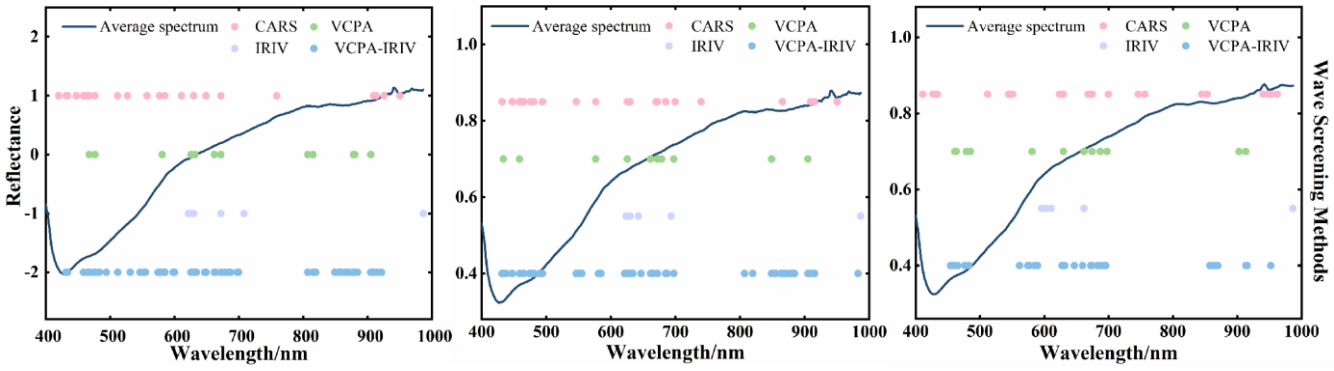


Fig. 6 - Distribution of characteristic bands: (a) Distribution of feature bands after SNV; (b) Distribution of feature bands after MSC; (c) Distribution of feature bands after MA.

The performance of the soil salinity prediction models constructed using the selected characteristic bands is presented in Table 2. Among the linear models, the MSC+VCPA-IRIV+PLSR combination yielded the best performance, achieving Rc^2 and Rp^2 values of 0.9882 and 0.9558, respectively, with RMSEC and RMSEP values of 0.4818 and 1.1459, and an RPD of 3.3056.

In contrast, within the nonlinear models, the SNV+VCPA-IRIA+SVR combination exhibited the highest predictive ability, with Rc^2 and Rp^2 values of 0.9863 and 0.9541, RMSEC and RMSEP values of 0.5341 and 1.0763, and an RPD of 3.3698. Among the characteristic band selection methods, the CARS, VCPA, and VCPA-IRIV algorithms all contribute positively to improving model performance. The models constructed using VCPA-IRIV achieve the best performance, likely because this method retains a greater number of feature variables closely associated with soil salinity during extraction, thereby enhancing model stability and generalization *(Dong et al., 2021; Yun et al., 2019)*. However, the models constructed using IRIV demonstrated the poorest predictive performance. This can be attributed to the fact that IRIV selected only six bands, leading to the exclusion of several informative variables related to soil salinity and, consequently, a decline in modeling performance. Therefore, the IRIV method was excluded from subsequent analyses.

Table 2

PLSR and SVR prediction models for soil salinity based on full band data

Preprocessing Methods	Data Dimensionality Reduction	Models	Calibration Set		Prediction Set		RPD
			R_C^2	RMSEC	R_P^2	RMSEP	
SNV	CARS	PLSR	0.9650	0.9728	0.8302	1.5700	1.7371
		SVR	0.979	0.6613	0.9429	1.1819	3.0196
	VCPA	PLSR	0.9515	1.0164	0.8358	2.2787	1.6131
		SVR	0.9771	0.7108	0.9444	1.2979	2.7934
	IRIV	PLSR	0.8124	2.0516	0.7375	2.5914	1.3641
		SVR	0.9083	1.4522	0.7697	1.9469	1.5202
	VCPA-IRIV	PLSR	0.9685	0.7880	0.9206	1.3614	2.5686
		SVR	0.9863	0.5341	0.9541	1.0763	3.3698
MSC	CARS	PLSR	0.9866	0.5849	0.8364	1.4902	1.8292
		SVR	0.9551	1.0096	0.9381	1.2515	2.7032
	VCPA	PLSR	0.9125	1.4184	0.7522	1.9975	1.4788
		SVR	0.9868	0.5237	0.9375	1.2583	2.8091
	IRIV	PLSR	0.9103	1.4874	0.7555	1.6887	1.4602
		SVR	0.9504	1.1163	0.7796	1.8060	1.5423
	VCPA-IRIV	PLSR	0.9882	0.4818	0.9558	1.1459	3.3056
		SVR	0.9858	0.5869	0.9700	1.0791	3.3369
MA	CARS	PLSR	0.8114	1.9731	0.7176	3.1068	1.2183
		SVR	0.9694	0.9429	0.9014	1.1123	2.3077
	VCPA	PLSR	0.8288	1.7187	0.7785	2.4541	1.0413
		SVR	0.9731	0.7734	0.9016	1.6346	2.2132
	IRIV	PLSR	0.7870	1.9946	0.7009	2.5102	1.0234
		SVR	0.9257	1.2652	0.7731	2.3643	1.4997
	VCPA-IRIV	PLSR	0.9089	1.4000	0.8447	2.2305	1.7217
		SVR	0.9753	0.7433	0.9238	1.4410	2.4530

Soil Salinity Prediction Model Based on Image Feature

Based on the image feature data extracted from 150 soil image samples, two regression models—PLSR and SVR—were constructed to predict soil salinity content. Table 3 presents the predictive performance of the soil salinity models developed using image features. The results indicate substantial performance discrepancies between the PLSR and SVR models, suggesting that image features exhibit limited stability in soil salinity prediction and that model performance is strongly influenced by both algorithm choice and sample characteristics.

Table 3

PLSR and SVR prediction models for soil salinity based on image feature

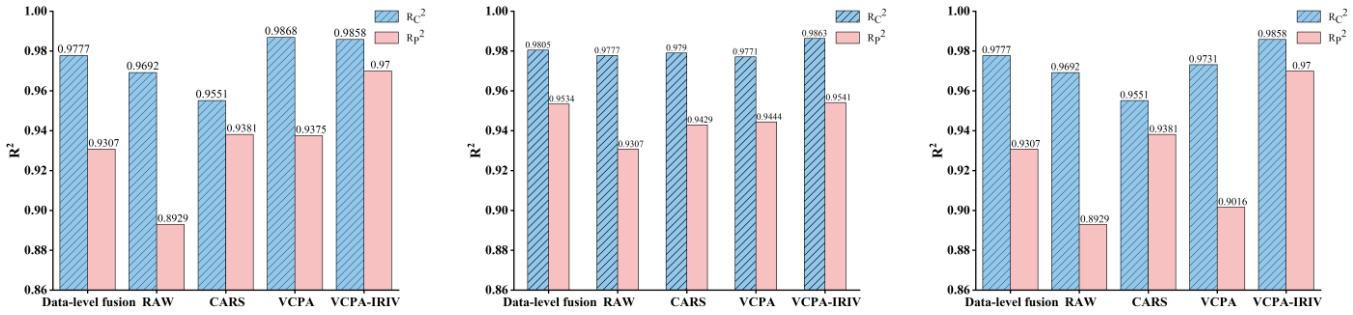
Preprocessing Methods	Calibration Set		Prediction Set		RPD
	R_C^2	RMSEC	R_P^2	RMSEP	
PLSR	0.8493	1.7281	0.7047	2.7846	1.3452
SVR	0.9806	0.6666	0.9553	1.2173	2.9783

Based on the above modeling results, the SVR model demonstrates higher prediction accuracy than the PLSR model across both the full band and the feature band. This superiority is reflected in the higher R^2 values and lower RMSE values obtained in both the training and prediction sets. However, despite the strong descriptive capability of spectral data, uncertainties remain when modeling under complex soil conditions. Moreover, the variability observed in the image feature models further highlights the limitations of relying on a single information source for soil salinity monitoring. These findings suggest that a single data source is insufficient to comprehensively characterize soil salinity, whereas multi-source data fusion can leverage the complementary strengths of spectral and image information, effectively overcoming the limitations in stability and generalization associated with single-source models *(Chen et al., 2021)*. Accordingly, subsequent work will employ the SVR model as the foundation to explore fusion strategies fusion spectral and image features, with particular emphasis on comparing data-level and feature-level fusion approaches in terms of their effectiveness in enhancing model stability and accuracy.

Soil Salinity Prediction Model Based on Data Fusion

Data Level Fusion Data Modelling

In data-level fusion, the three preprocessed spectral datasets are concatenated with the image feature data to generate a new fused dataset. After normalization of the fused dataset, an SVR model for soil salinity prediction is constructed. Figure 7 presents a comparison of the predictive performance between the data-level fusion model and the models constructed using single spectral datasets.



(a) (b) (c)

Fig. 7 - Comparison of prediction performance before and after data level fusion: (a) Comparison of R^2 before and after data level fusion under SNV; (b) Comparison of R^2 before and after data level fusion under MSC; (c) Comparison of R^2 before and after data level fusion under MA.

As shown in the figure, the data level fusion model yields a substantial improvement in predictive performance compared with the spectral models constructed using individual preprocessing methods. However, relative to the spectral models developed using characteristic bands, the data-level fusion models exhibit comparatively inferior predictive performance. This is primarily attributed to the fact that data-level fusion, while integrating comprehensive information, simultaneously introduces substantial redundant variables, resulting in markedly increased data dimensionality and ultimately hindering model training efficiency and generalization capability *(Gholizadeh et al., 2021; Meng et al., 2022)*. This finding suggests that the applicability of data-level fusion in soil salinity modeling is limited. Therefore, subsequent analyses will further investigate the impact of feature-level fusion on improving the accuracy of soil salinity prediction models.

Feature Level Fusion Data Modelling

In the feature-level fusion process, the spectral features obtained after preprocessing and characteristic band selection are concatenated with the image features to construct a fused feature dataset. Subsequently, an SVR model is developed after applying Z-score normalization to the fused features dataset. Figure 8 presents a comparison of the predictive performance between the feature-level fusion model and the models constructed using single spectral datasets.

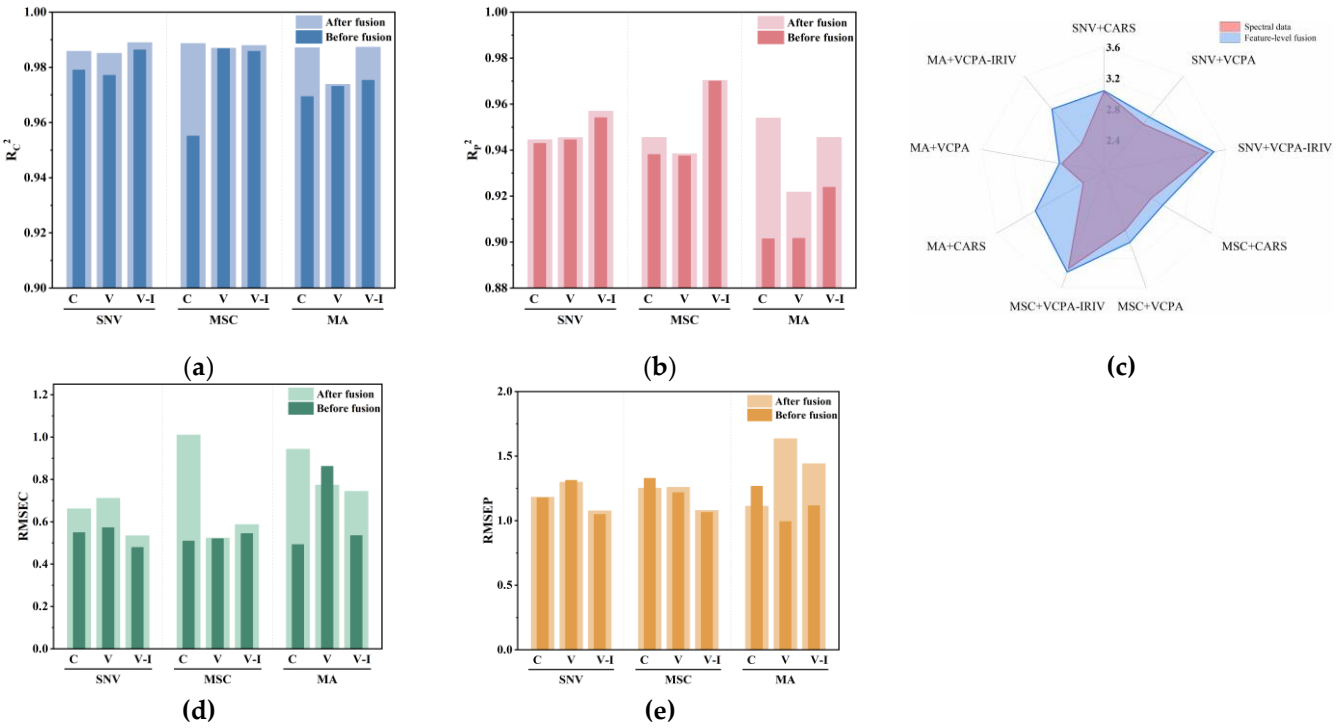


Fig. 8 - Comparison of model performance before and after feature level fusion: (a) Comparison of R_C^2 before and after feature level fusion; (b) Comparison of R_P^2 before and after feature level fusion; (c) Comparison of RPD before and after feature level fusion; (d) Comparison of RMSEC before and after feature level fusion; (e) Comparison of RMSEP before and after feature level fusion.

Compared with the models developed using single spectral datasets, the feature-level fusion model demonstrates superior overall performance in both prediction accuracy and computational efficiency. This improvement can be attributed to the feature selection performed prior to the fusion process, which effectively removes redundant information while preserving key features that are discriminative for soil salinity variation, thereby substantially enhancing the prediction accuracy and stability of the model *(Yin et al., 2023)*. As shown in Figure 8, among all fusion combinations, the MA+CARS+SVR approach exhibits the most substantial performance improvement. Specifically, Rc^2 increases from 0.9694 to 0.9871, Rp^2 increases from 0.9014 to 0.9539, RMSEC and RMSEP reach 0.4925 and 1.2662, respectively, and RPD rises from 2.3077 to 3.0241. Figures 9(a) and 9(b) further illustrate the relationships between the training and prediction sets before and after feature-level fusion with this combination.

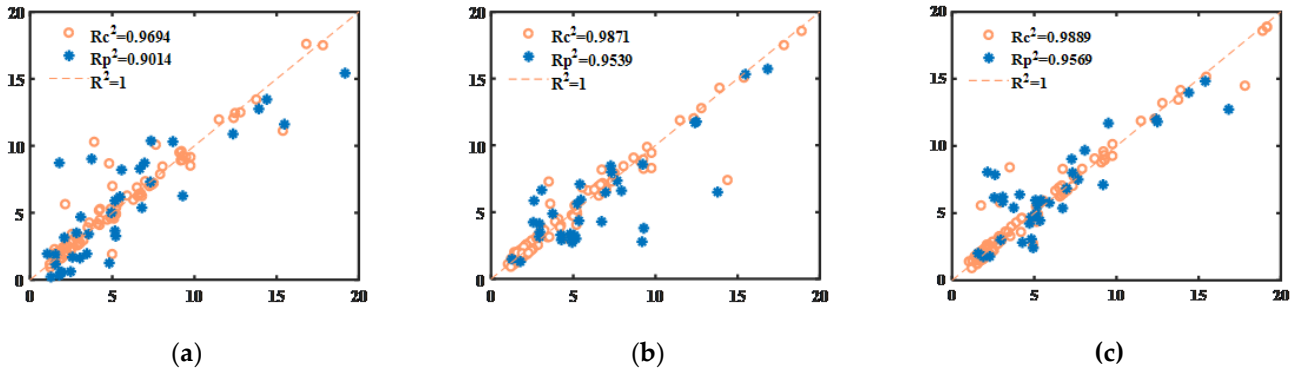


Fig. 9 - (a) Relationship between the MA+VCPA-IRIV training and prediction sets for single-spectral data; (b) Relationship between the MA+VCPA-IRIV training and prediction sets based on feature level data fusion; (c) Relationship between the training and prediction sets based on the optimal model.

A further comparison of the fusion models indicates that the SNV+VCPA-IRIV+SVR combination yields the best overall predictive performance. Under this model, Rc^2 and Rp^2 reach 0.9889 and 0.9569, respectively, demonstrating excellent fitting capability in both the training and testing phases. The RMSEC and RMSEP values of 0.4790 and 1.0408, together with an RPD value of 3.4423, further indicate strong robustness and a high degree of generalization.

Figure 9(c) illustrates the relationship between the training and prediction sets for the model. The data points are distributed uniformly around the regression line without evident deviations or systematic errors, further confirming the high predictive accuracy and strong fitting capability of this fusion model in soil salinity estimation. The data points are distributed uniformly around the regression line without evident deviations or systematic errors, further confirming the high predictive accuracy and strong fitting capability of this fusion model in soil salinity estimation.

CONCLUSIONS

This study demonstrates the feasibility of quantitatively predicting soil salinity in saline-alkali soils using hyperspectral data in combination with machine vision techniques. SVR models were constructed using both data-level fusion and feature-level fusion strategies. The results indicate that feature-level fusion models outperform data-level fusion models in both prediction accuracy and computational efficiency. Among all model combinations, the SVR model constructed by fusing the spectral features selected through the SNV+VCPA-IRIVA algorithm with the image features achieves the best performance, yielding an Rp^2 of 0.9569 and an RPD of 3.4423. These findings suggest that integrating hyperspectral data with machine vision enables reliable quantitative detection of soil salinity in saline-alkali soils, effectively addressing the limitations in modeling stability and generalization associated with single-source data.

ACKNOWLEDGEMENT

This research was funded by National Key R&D Program of China, grant number 2023YFD2001403, and the Major Science and the Major Science and Technology Innovation Project of Shandong Province, grant number 2023CXGC010701, 2022CVGC010610, 2022CXGC020708.

REFERENCES

[1] Achata, E. M., Inguglia, E. S., Esquerre, C. A., Tiwari, B. K., & O'Donnell, C. P. (2019). Evaluation of Vis-NIR hyperspectral imaging as a process analytical tool to classify brined pork samples and predict brining salt concentration. *Journal of Food Engineering*, vol.246, 134–140. https://doi.org/10.1016/j.jfoodeng.2018.10.022

[2] Allbed, A., Kumar, L., & Aldakheel, Y. Y. (2014). Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region. *Geoderma*, vol.230, 1–8. https://doi.org/10.1016/j.geoderma.2014.03.025

[3] Al-Mbaideen, A. A. (2019). Application of Moving Average Filter for the Quantitative Analysis of the NIR Spectra. *Journal of Analytical Chemistry*, vol.74, 686–692. https://doi.org/10.1134/S1061934819070013

[4] Alordzinu, K. E., Li, J., Lan, Y., Appiah, S. A., AL Aasmi, A., Wang, H., Liao, J., Sam-Amoah, L. K., & Qiao, S. (2021). Ground-Based Hyperspectral Remote Sensing for Estimating Water Stress in Tomato Growth in Sandy Loam and Silty Loam Soils. *Sensors*, vol.21, 5705. https://doi.org/10.3390/s21175705

[5] Alsaleh, A. R. S., Alcibahy, M., Gafoor, F. A., Hashemi, H. A., Athamneh, B., Al Hammadi, A. A., Seneviratne, L., & Al Shehhi, M. R. (2025). Estimation of soil organic carbon in arid agricultural fields based on hyperspectral satellite images. *Geoderma*, vol.453, 117151. https://doi.org/10.1016/j.geoderma.2024.117151

[6] An, D.; Zhao, G.; Chang, C.; Wang, Z.; Zhang, T.; Li, P.; Jia, J. (2015). Estimation of Coastal Saline Soil Salinity in the Yellow River Delta Based on Field Hyperspectral. *Chinese Journal of Soil Science*, vol.*46*, 843-850.

[7] An, T., Huang, W., Tian, X., Fan, S., Duan, D., Dong, C., Zhao, C., & Li, G. (2022). Hyperspectral imaging technology coupled with human sensory information to evaluate the fermentation degree of black tea. *Sensors and Actuators B: Chemical*, vol.366, 131994. https://doi.org/10.1016/j.snb.2022.131994

[8] Bao, S. (2000). Soil Agrochemical Analysis, 3rd Ed.; China Agriculture Press: Beijing, China, 2000; pp. 178–200.

[9] Bouaziz, M., Matschullat, J., & Gloaguen, R. (2011). Improved remote sensing detection of soil salinity from a semi-arid climate in Northeast Brazil. *Comptes Rendus Geoscience*, vol.343, 795–803. https://doi.org/10.1016/j.crte.2011.09.003

[10] Chen, J., Yang, C., Yuan, C., Li, Y., An, T., & Dong, C. (2022). Moisture content monitoring in withering leaves during black tea processing based on electronic eye and near infrared spectroscopy. *Scientific Reports*, vol.12, 20721. https://doi.org/10.1038/s41598-022-25112-6

[11] Chen, Y., Chen, Z., Yan, Q., Liu, Y., & Wang, Q. (2024a). Non-destructive detection of egg white and yolk morphology transformation and salt content of salted duck eggs in salting by hyperspectral imaging. *International Journal of Biological Macromolecules*, vol.262, 130002. https://doi.org/10.1016/j.ijbiomac.2024.130002

[12] Chen, Y., Guo, M., Chen, K., Jiang, X., Ding, Z., Zhang, H., Lu, M., Qi, D., & Dong, C. (2024b). Predictive models for sensory score and physicochemical composition of Yuezhou Longjing tea using near-infrared spectroscopy and data fusion. *Talanta*, vol.273, 125892. https://doi.org/10.1016/j.talanta.2024.125892

[13] Chen, Y., Xu, Y., Di, Y., Cui, X., Zhang, J., Zhou, X., Xiao, C., & Li, S. (2021). COD Concentration Prediction Model Based on Multi-Spectral Data Fusion and GANs Algorithm (多光谱数据融合和 GANs 算法的 COD 浓度预测). *Spectroscopy and Spectral Analysis*, vol.41, 188. https://doi.org/10.3964/j.issn.1000-0593(2021)01-0188-06

[14] Cruz-Tirado, J. P., Oliveira, M., de Jesus Filho, M., Godoy, H. T., Amigo, J. M., & Barbin, D. F. (2021). Shelf life estimation and kinetic degradation modeling of chia seeds (*Salvia hispanica*) using principal component analysis based on NIR-hyperspectral imaging. *Food Control*, vol.123, 107777. https://doi.org/10.1016/j.foodcont.2020.107777

[15] Dong, C., Ye, Y., Yang, C., An, T., Jiang, Y., Ye, Y., Li, Y., & Yang, Y. (2021). Rapid detection of catechins during black tea fermentation based on electrical properties and chemometrics. *Food Bioscience*, vol.40, 100855. https://doi.org/10.1016/j.fbio.2020.100855

[16] Gholizadeh, A., Coblinski, J. A., Saberioon, M., Ben-Dor, E., Drábek, O., Demattê, J. A. M., Borůvka, L., Němeček, K., Chabrillat, S., & Dajčl, J. (2021). Vis–NIR and XRF Data Fusion and Feature Selection to Estimate Potentially Toxic Elements in Soil. *Sensors*, vol.21, 2386. https://doi.org/10.3390/s21072386

[17] He, Q., Guo, Y., Li, X., He, Y., Lin, Z., & Zeng, H. (2024). Spectral Fingerprinting of Tencha Processing: Optimising the Detection of Total Free Amino Acid Content in Processing Lines by Hyperspectral Analysis. *Foods*, vol.13, 3862. https://doi.org/10.3390/foods13233862

[18] He, Q., Zhang, H., Li, T., Zhang, X., Li, X., & Dong, C. (2023). NIR Spectral Inversion of Soil Physicochemical Properties in Tea Plantations under Different Particle Size States. *Sensors*, vol.23, 9107. https://doi.org/10.3390/s23229107

[19] Jia, P., He, W., Hu, Y., Liang, Y., Liang, Y., Xue, L., Zamanian, K., & Zhao, X. (2024). Inversion of coastal cultivated soil salt content based on multi-source spectra and environmental variables (基于多源光谱和环境变量的滨海耕地土壤盐分含量反演). *Soil and Tillage Research*, vol.241, 106124. https://doi.org/10.1016/j.still.2024.106124

[20] Jiang, H., He, Y., Xu, W., & Chen, Q. (2021). Quantitative Detection of Acid Value During Edible Oil Storage by Raman Spectroscopy: Comparison of the Optimization Effects of BOSS and VCPA Algorithms on the Characteristic Raman Spectra of Edible Oils. *Food Analytical Methods*, vol.14, 1826–1835. https://doi.org/10.1007/s12161-020-01939-5

[21] Li, J., Chen, L., & Huang, W. (2018). Detection of early bruises on peaches (*Amygdalus persica* L.) using hyperspectral imaging coupled with improved watershed segmentation algorithm. *Postharvest Biology and Technology*, vol.135, 104–113. https://doi.org/10.1016/j.postharvbio.2017.09.007

[22] Li, X., Li, Z., Yang, X., & He, Y. (2021). Boosting the generalization ability of Vis-NIR-spectroscopy-based regression models through dimension reduction and transfer learning. *Computers and Electronics in Agriculture*, vol.186, 106157. https://doi.org/10.1016/j.compag.2021.106157

[23] Liu, Y.Q.; Chen, H.Y.; Wang, R.Y.; Chang, C.Y.; Chen, Z. (2016). Quantitative Analysis of Soil Salt and Its Main Ions Based on Visible/Near Infrared Spectroscopy in Estuary Area of Yellow River (基于可见/近红外光谱的黄河口区土壤盐分及其主要离子的定量分析). *Scientia Agricultura Sinica*, vol.49, 1925–1935.

[24] Liu, Z., Zhang, R., Yang, C., Hu, B., Luo, X., Li, Y., & Dong, C. (2022). Research on moisture content detection method during green tea processing based on machine vision and near-infrared spectroscopy technology. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol.271, 120921. https://doi.org/10.1016/j.saa.2022.120921

[25] Meng X., Yu L., Zhou Y., and Li S. (2022). Predicting Organic Carbon Using Data Fusion of Visible Near-Infrared and Middle Infrared Spectra by Proximal Soil Sensing (基于可见近红外和中红外近地面光谱数据融合的土壤有机碳含量反演). *Chinese Journal of Soil Science*, vol.53, 301–307. https://doi.org/10.19336/j.cnki.trtb.2021072202

[26] Ramasamy, J., Raju, A., Ranganathan, K. K., Dhanaraju, M., Saliha, B., Ramalingam, K., & Samiappan, S. (2025). Battle Royale Optimization for Optimal Band Selection in Predicting Soil Nutrients Using Visible and Near-Infrared Reflectance Spectroscopy and PLSR Algorithm. *Journal of Imaging*, vol.11. https://doi.org/10.3390/jimaging11030083

[27] Song, J., Hu, M., Wang, J., Zhou, M., Sun, L., Qiu, S., Li, Q., Sun, Z., & Wang, Y. (2019). ALK positive lung cancer identification and targeted drugs evaluation using microscopic hyperspectral imaging technique. *Infrared Physics & Technology*, vol.96, 267–275. https://doi.org/10.1016/j.infrared.2018.12.001

[28] Subi, X., Eziz, M., Zhong, Q., & Li, X. (2024). Estimating the chromium concentration of farmland soils in an arid zone from hyperspectral reflectance by using partial least squares regression methods. *Ecological Indicators*, vol.161, 111987. https://doi.org/10.1016/j.ecolind.2024.111987

[29] Sun, W., Liu, S., Zhang, X., & Li, Y. (2022a). Estimation of soil organic matter content using selected spectral subset of hyperspectral data. *Geoderma*, vol.409, 115653. https://doi.org/10.1016/j.geoderma.2021.115653

[30] Sun, M.; Li, Q.; Jiang, X.; Ye, T.; Li, X.; Niu, B. (2022b). Estimation of Soil Salt Content and Organic Matter on Arable Land in the Yellow River Delta by Combining UAV Hyperspectral and Landsat-8 Multispectral Imagery. Sensors, vol.22, 3990. https://doi.org/10.3390/ s22113990

[31] Sun, Y., Wang, Y., Xiao, H., Gu, X., Pan, L., & Tu, K. (2017). Hyperspectral imaging detection of decayed honey peaches based on their chlorophyll content. *Food Chemistry*, vol.235, 194–202. https://doi.org/10.1016/j.foodchem.2017.05.064

[32] Tao, S., Yi-dan, B. a. O., & Yong, H. E. (2009). Research on the Method for Rapid Detection of Soil Moisture Content Using Spectral Data (利用光谱数据快速检测土壤含水量的方法研究). *Spectroscopy and Spectral Analysis*, vol.29, 675. https://doi.org/10.3964/j.issn.1000-0593(2009)03-0675-03

[33] Wang, Z., Chen, J., Fan, Y., Cheng, Y., Wu, X., Zhang, J., Wang, B., Wang, X., Yong, T., Liu, W., Liu, J., Du, J., Yang, W., & Yang, F. (2020). Evaluating photosynthetic pigment contents of maize using UVE-PLS based on continuous wavelet transform. *Computers and Electronics in Agriculture*, vol.169, 105160. https://doi.org/10.1016/j.compag.2019.105160

[34] Wang, Z., Wang, L., Xu, R., Huang, H., & Wu, F. (2012). GIS and RS based Assessment of Cultivated Land Quality of Shandong Province. *Procedia Environmental Sciences*, vol.12, 823–830. https://doi.org/10.1016/j.proenv.2012.01.354

[35] Wang, D.; Chen, H.; Wang, G.; Cong, J.; Wang, X.; Wei, X. (2019). Salinity Inversion of Severe Saline Soil in the Yellow River Estuary Based on UAV Multi-Spectra (无人机多光谱反演黄河口重度盐渍土盐分的研究). *Scientia Agricultura Sinica*, vol.52, 1698–1709.

[36] Whitney, K., Scudiero, E., El-Askary, H. M., Skaggs, T. H., Allali, M., & Corwin, D. L. (2018). Validating the use of MODIS time series for salinity assessment over agricultural soils in California, USA. *Ecological Indicators*, vol.93, 889–898. https://doi.org/10.1016/j.ecolind.2018.05.069

[37] Wu, T., Yang, L., Zhou, J., Lai, D. C., & Zhong, N. (2019). An improved nondestructive measurement method for salmon freshness based on spectral and image information fusion. *Computers and Electronics in Agriculture*, vol.158, 11–19. https://doi.org/10.1016/j.compag.2019.01.039

[38] Wu, L., Jiang, Q., Zhang, Y., Wang, S. *(2023).* Nontradional Detection of Soil Moisture Content Based On Hyperspectral Imaging Technique. *Bangladesh Journal of Botany*, vol.*51*, 1039–1049.

[39] Xia, H., Chen, W., Hu, D., Miao, A., Qiao, X., Qiu, G., Liang, J., Guo, W., & Ma, C. (2024). Rapid discrimination of quality grade of black tea based on near-infrared spectroscopy (NIRS), electronic nose (E-nose) and data fusion. *Food Chemistry*, vol.440, 138242. https://doi.org/10.1016/j.foodchem.2023.138242

[40] Xie X., S. B. and H. H. (2013). Relationship between visible-near infrared reflectance spectroscopy and heavy metal of soil concentration (土壤可见光-近红外反射光谱与重金属含量之间的相关性). *Acta Pedologica Sinica*, vol.44, 982–993. https://doi.org/10.11766/trxb200607060604

[41] Yang, C., Zhao, Y., An, T., Liu, Z., Jiang, Y., Li, Y., & Dong, C. (2021). Quantitative prediction and visualization of key physical and chemical components in black tea fermentation using hyperspectral imaging. *LWT*, vol.141, 110975. https://doi.org/10.1016/j.lwt.2021.110975

[42] Yang, H., Wang, Z., Cao, J., Wu, Q., & Zhang, B. (2023). Estimating soil salinity using Gaofen-2 imagery: A novel application of combined spectral and textural features. *Environmental Research*, vol.217, 114870. https://doi.org/10.1016/j.envres.2022.114870

[43] Yang, J., Liu, Q., Zhao, N., Chen, J., Peng, J., Pan, L., & Tu, K. (2020). Hyperspectral Imaging for Non-destructive Determination and Visualization of Moisture and Carotenoid Contents in Carrot Slices during Drying (基于高光谱成像的干燥胡萝卜片水分及类胡萝卜素含量无损检测和可视化分析). *Food Science*, vol.41, 285–291. https://doi.org/10.7506/spkx1002-6630-20190225-169

[44] Yin, Y., Li, J., Ling, C., Zhang, S., Liu, C., Sun, X., & Wu, J. (2023). Fusing spectral and image information for characterization of black tea grade based on hyperspectral technology. *LWT*, vol.185, 115150. https://doi.org/10.1016/j.lwt.2023.115150

[45] Yu, H.-D., Qing, L.-W., Yan, D.-T., Xia, G., Zhang, C., Yun, Y.-H., & Zhang, W. (2021). Hyperspectral imaging in combination with data fusion for rapid evaluation of tilapia fillet freshness. *Food Chemistry*, vol.348, 129129. https://doi.org/10.1016/j.foodchem.2021.129129

[46] Yun, Y.-H., Li, H.-D., Deng, B.-C., & Cao, D.-S. (2019). An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends in Analytical Chemistry*, vol.113, 102–115. https://doi.org/10.1016/j.trac.2019.01.018

[47] Yun, Y.-H., Wang, W.-T., Tan, M.-L., Liang, Y.-Z., Li, H.-D., Cao, D.-S., Lu, H.-M., & Xu, Q.-S. (2014). A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. *Analytica Chimica Acta*, vol.807, 36–43. https://doi.org/10.1016/j.aca.2013.11.032

[48] Zhang, H., He, Q., Yang, C., Lu, M., Liu, Z., Zhang, X., Li, X., & Dong, C. (2023). Research on the Detection Method of Organic Matter in Tea Garden Soil Based on Image Information and Hyperspectral Data Fusion. *Sensors*, vol.23, 9684. https://doi.org/10.3390/s23249684