

Article

Not peer-reviewed version

Comparative Analysis of Machine Learning and Deep Learning Algorithms for Assessing Agricultural Product Quality Using NIRS

Jiwen Ren , Yuming Xiong , Xinyu Chen , [Yong Hao](#) *

Posted Date: 2 August 2024

doi: 10.20944/preprints202408.0024.v1

Keywords: Gramian Angular Difference Field; Convolutional Neural Networks; Near-infrared spectroscopy; Coordinate Attention; Robust model



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Comparative Analysis of Machine Learning and Deep Learning Algorithms for Assessing Agricultural Product Quality Using NIRS

Jiwen Ren ¹, Yuming Xiong ¹, Xinyu Chen ² and Yong Hao ^{1,*}

¹ School of Mechatronics and Vehicle Engineering, East China Jiaotong University, Nanchang, 330013, China

² Optoelectronics Department of Changzhou Institute of Technology, Changzhou, 213000, China

* Correspondence: haonm@163.com

Abstract: The success of near-infrared spectroscopy (NIRS) analysis hinges on the precision and robustness of the calibration model. Shallow learning (SL) algorithms like partial least squares discriminant analysis (PLS-DA) often fall short in capturing interrelationships between adjacent spectral variables, and the analysis results are easily affected by spectral noise, which dramatically limits the breadth and depth application of NIRS. Deep learning (DL) methods, with their capacity to discern intricate features from limited samples, are progressively integrated into NIRS. In this paper, two discriminant analysis problems including wheat kernels and Yali pears were used as examples, and several representative calibration models to research the robustness and effectiveness of the model. Additionally, this article proposed a near-infrared calibration model, which is based on gramian angular difference field and coordinate attention convolutional neural networks (G-CACNN). The research results show that, compared with SL, spectral preprocessing has a smaller impact on the analysis accuracy of consensus learning (CL) and DL, and the latter has the highest analysis accuracy in the modeling results of using the original spectrum. The accuracy of G-CACNN in two discrimination tasks is 98.48% and 99.39%, respectively. Finally, This research compared the performance of various models under noise to evaluate the robustness and noise resistance of the proposed method.

Keywords: gramian angular difference field; convolutional neural networks; near-infrared spectroscopy; coordinate attention; robust model

1. Introduction

Near-infrared spectroscopy (NIRS) analysis technology has been widely used in agriculture [1–3], food [4–6], medicine [7–9], and other fields due to its fast and non-destructive characteristics [10]. Generally, the response of NIRS is characterized by weak absorption and overlapping peaks. Therefore, it is necessary to use Principal Component Analysis (PCA), Partial Least Squares (PLS), and other chemometrics methods to reveal the complex relationship between multiple variables and targets [11]. Due to the susceptibility of NIRS to environment, instrument status, sample physical state, etc., it is necessary to introduce spectral preprocessing such as denoising, smoothing, and standardization to suppress or eliminate the interference before analysis [12]. To eliminate the influence of interfering variables on model accuracy and robustness, several methods, including monte carlo uninformative variable elimination (MC-UVE) [13], competitive adaptive reweighted sampling (CARS) [14], iterative variable subset optimization (IVSO) [15], etc. are introduced to select useful modeling variables [16]. Several recent studies on the use of NIRS in agricultural product detection have demonstrated the effectiveness of multivariate calibration models. They include some new model optimization algorithms such as genetic algorithm (GA), random frog leaping algorithm (RFLA), particle swarm optimization (PSO), and some classic shallow learning (SL) models such as PLS and support vector regression (SVR) [17–19]. In the above research, multiple preprocessing and

data-cleaning methods were integrated into the SL methods. These improvements have to some extent improved the performance of the model, but the shortcomings of the SL algorithm, such as lack of non-linear ability and inability to capture the interrelationships between variables, have not been well addressed. It is necessary to promote the updating of NIRS analysis algorithms.

In spectral analysis, the existence of outliers may lead to the instability of the analysis model and an increase in prediction error. To overcome the negative influence of outliers in analysis model, the consensus strategy is introduced in the model construction. Since the consensus learning (CL) uses the integration results of multiple-member models, the existence of outliers will only affect individual models and have relatively little influence on the final prediction. Chen, et al. [20] explored the feasibility of adopting NIRS and consensus learning to improve the diagnosis of colorectal cancer, multiple weak learners were integrated to construct diagnostic models, and a linear discriminant classifier (LDA) was used as the random subspace method (RSM) for weak learners. Zhao, et al. [21] analyzed the reflection spectra of ginseng, and the decision trees (DT) based random forest (RF) machine learning method was successfully used to establish the model for identifying the growth year of ginseng.

In the field of spectral combined deep learning (DL) algorithms, The most important benefit of DL is automatic feature engineering, which reduces the burden of variable selection and preprocessing. In addition, one-dimensional convolutional neural networks have poor generalization and low accuracy. The research on converting one-dimensional (1D) signals into two-dimension (2D) images by different methods including energy grayscale image [22], angle-amplitude image [23], and binary images [24], etc., and then classifying and recognizing based on DL technology is rising. Hao, et al. [25] established the DL model based on the NIRS stacked encoding of the Yali pear, which realized the non-destructive online detection of Yali pear pests. Pu, et al. [26] encoded the terahertz (THz) time-domain spectra of Pericarpium Citri Reticulatae (PCR) into Graham Angle Field (GAF) images and Markov Transition Field (MTF) images.

The DL methods have exhibited great potential in qualitative and quantitative analysis of NIRS due to its strong feature extraction ability and accurate analysis results. In this paper, to fully understand the application considerations and operational guidelines of DL methods in the construction of NIRS analysis models, two classification tasks of wheat kernels and Yali pears browning identification are discussed. Firstly, the spectra are preprocessed, and the qualitative analysis models of partial least squares discriminant analysis (PLS-DA), random forest (RF), and convolutional neural network (CNN) are established. The effects of different spectral preprocessing methods on the prediction accuracy of the above models are discussed. Secondly, the gramian angular difference field (GADF) is used to convert the spectra of the above two datasets into images, and four models including PLS-DA, RF, CNN, and coordinate attention convolutional neural networks (CACNN) based on the images are established [27]. The advantages of GADF in constructing spectral spatial domain information are determined. Finally, the noisy spectra are constructed by adding random noise, and various discriminant models representing three modeling strategies of SL, CL, and DL methods, respectively, were established to examine the robustness of each model under different noise levels. This research work is expected to fully clarify and utilize the advantages of DL methods in feature capture and simplify the construction process of DL models for NIRS analysis.

2. Theory and Algorithm

2.1. Gramian Angular Field (GAF)

GAF is a coding method for converting 1D signals to 2D images [28], which can retain the mutual difference or sum information of the original 1D signals. The specific calculation process is as follows.

First, the true value of the spectrum signal $X = \{x_1, x_2, \dots, x_n\}$ is scaled into the range [0, 1]. The calculation formula is shown in Eq. 1.

$$\tilde{x}_i = \frac{(x_i - \max(X)) + (x_i - \min(X))}{\max(X) - \min(X)}. \quad (1)$$

Among them, $\max(X)$ and $\min(X)$ represent the maximum and minimum values of the spectrum, respectively, and $\tilde{x}_i$ represents the scaled result of each element in the spectrum.

Second, the arccosine value θ_i of each element of the scaled spectrum $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n\}$ is calculated, and the radius R of the polar coordinates is divided into N segments to mark the position information of spectrum t at N sampling points. The calculation process is shown in Eq. 2.

$$\left[\begin{array}{l} \theta_i = \arccos(\tilde{x}_i), \quad 0 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r_i = \frac{i}{N}, \quad \text{with } t_i \leq N \end{array} \right]. \quad (2)$$

Finally, according to the definition of Eq. 3, the value of each pixel in the GAF image is obtained by using the trigonometric function to identify the difference information in different sampling intervals, where I is a unit row vector. A way to preserve structure dependency was provided, since the spatial position of the image represents the position of the sampling point. The piecewise proximate aggregation (PPA) of raw data during converting can be seen as a method of spectra smoothing.

$$\begin{aligned} GADF &= [\sin(\theta_i - \theta_j)] = \sqrt{I - \tilde{X}^2} \cdot \tilde{X} - \tilde{X}' \cdot \sqrt{I - \tilde{X}^2}. \\ GASF &= [\cos(\theta_i + \theta_j)] = \tilde{X}' \cdot \tilde{X} - \sqrt{I - \tilde{X}^2} \cdot \sqrt{I - \tilde{X}^2}. \end{aligned} \quad (3)$$

2.2. PLS-DA

PLS-DA is derived from PLS regression [29] and involves forming a regression model between the X and Y . Assuming there is a decomposition of X and Y , as shown in Eq. 4.

$$Y = TQ^T + F, \quad X = TP^T + E \quad (4)$$

Where $T (T \in \mathbb{R}^{n \times r})$ is the r principal components composed of linear combinations of observed values, $P (P \in \mathbb{R}^{p \times r})$ and $Q (Q \in \mathbb{R}^{l \times r})$ are the coefficient matrix of r principal components, $E (E \in \mathbb{R}^{n \times p})$ and $F (F \in \mathbb{R}^{n \times l})$ are the residual matrix. A set of vectors $W (W = w_1, w_2, \dots, w_r)$ is found under the condition of Eq. 4, and this process follows the optimization standard as Eq. 5.

$$\left\{ \begin{array}{l} \hat{w}_k = \arg \max_w w^T X^T Y Y^T X w \\ \text{s.t. } w^T w = 1, \quad w^T S_{XX} \hat{w}_j = 0, \quad j = 0, 1, \dots, r-1 \end{array} \right. \quad (5)$$

Where w is the linear combination coefficient of each row in X , and S_{XX} is the sample variance of X . The algorithm described above improves predictive performance in high-dimensional data by reducing the dimensionality of the independent variables to capture the maximum common variance with the response. The final model can be summarized as Eq. 6.

$$\hat{Y} = XWB^T + b_0 \quad (6)$$

Where $\hat{Y}$ is the predicted value, $X (m \times n)$ represents m samples, each with n features. B is the transpose of P , and b_0 is the deviation. PLS will pay attention to the influence of the extracted components on the dependent variable during principal component extraction. Therefore, PLS considerably retains the useful information of the data while considering the decisiveness of the independent variable to the dependent variable.

In PLS-DA, Y is a set of discrete variables representing categories. Usually represented by +1, -1, or +1, 0. However, the PLS is designed for continuous variables (regression tasks), and the predicted output is continuous variables. Therefore, Decision Rules (DR) are used to accurately convert into meaningful category labels. There are three popular DR methods: (a) naive (maximum value, Max); (b) cut-off point; and (c) boundary line [29]. The DR method of the PLS-DA algorithm involved in this article is naive (maximum value, Max).

2.3. Random Forest (RF)

RF method is one of the most widely used consensus learning models, which constructed an ensemble model by integrating the results of multiple sub-models developed with different training data subsets formed by using random sampling methods [30]. It consists of multiple decision trees (DTs) generated based on a combination of bootstrap aggregation and the random subspace method. Specifically, bootstrap aggregation is a method to construct multiple data subsets by extracting samples from the original data set. The random subspace method is to randomly select a feature subset as a candidate for splitting when constructing each split node of the decision tree. This further increases the diversity of models and helps to prevent over fitting. Each decision tree is generated based on CART (classification and regression trees). In this study Gini index minimization criterion is used for feature selection, and the binary tree is generated, recursively. The calculation formula for Gini index is shown in Eq. 7.

$$\text{Gini ind}(p) = \sum_{k=1}^K p_k (1 - p_k) \quad (7)$$

where p_k is the frequency of category k appearing in the dataset. The DT can select features that are more important for classification tasks by minimizing the Gini index [31].

In RF, each DT segments the data based on different features and generates a set of rules for predicting the target variables. During the construction of DTs, randomness is introduced through sampling with replacement from the original data and considering only a subset of features for node division. Each DT in RF makes independent predictions during testing, and classification problems are resolved by voting to select the category with the most votes as the final prediction result. Due to differences in sample subsets selected by each DT and integration of multiple DTs prediction results, the RF is relatively robust against missing data and outliers [32].

2.4. Coordinated Attention Convolutional Neural Networks (CACNN)

Convolutional neural network (CNN) is a variant of multi-layer perceptron (MLP), which uses convolutional operations to capture the features of data. It has been widely used in the classification of images, sound, and other datasets [33]. CNN is usually composed of three parts: convolution layer, pooling layer, and fully connected layer. The convolution layer extracts different features from the image through many different convolution kernels, and the features are kept in different channels. Different activation functions (AFs) are used to filter out some of these features. The pooling layer is adapted to retain an important part of the features extracted by the convolutional layer, reduce parameters. The fully connected layers are used to classify the extracted features.

Many excellent CNN networks have emerged in the evolution of network structures. Among these classic networks such as LeNet-5, VGG, GoogLeNet, and ResNet, etc., the VGG networks are widely used in image recognition due to their concise structure, small convolutional kernels, small pooling kernels, multiple channels, and feature maps with deeper and wider layers [34]. Each convolution layer uses a 3×3 convolution kernel. Based on this network, a coordinate attention (CA) module can help the model better locate and focus on important features. is added to CNN to form CACNN. The network structure is shown in Figure 1.

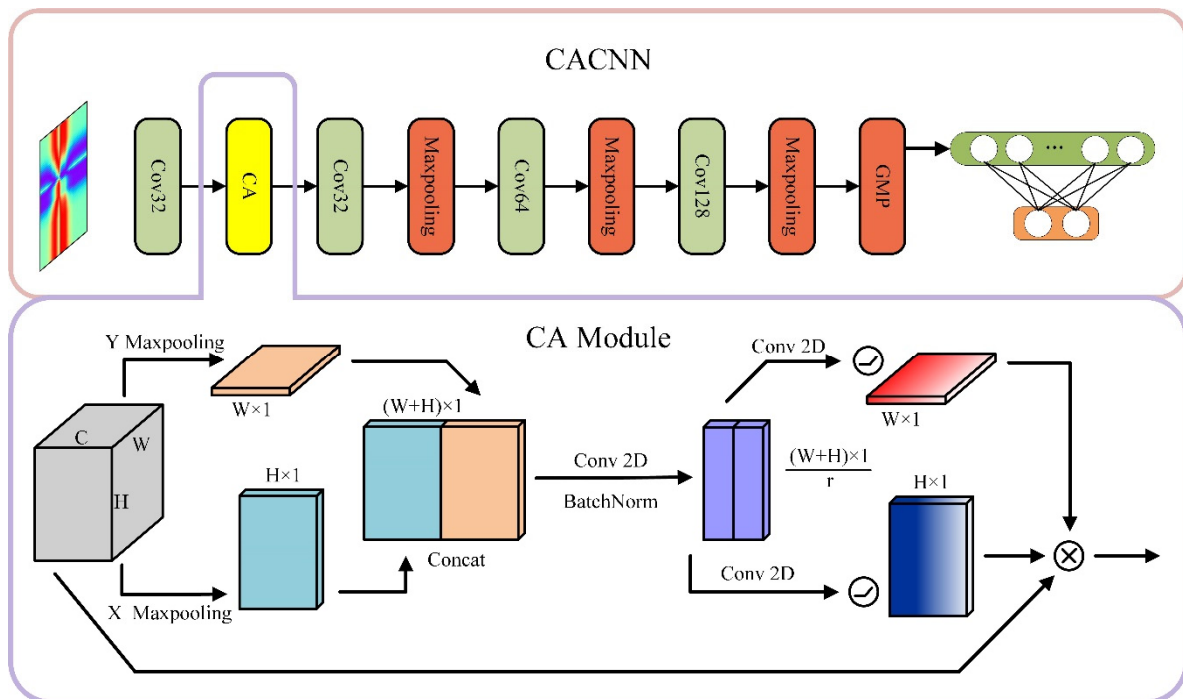


Figure 1. Network structure and CA internal structure.

The general attention module compresses the spatial information into the channel descriptor through global pooling, and it isn't easy to save the position information. CA comprises two parts: coordinate attention embedding (CAE) and coordinate attention generation (CAG). The spatial information of crucial features is preserved and enhanced in the new feature map CAE encodes each channel from two spatial ranges along horizontal and vertical coordinates using pooled cores, obtaining a global acceptance field and encoding precise position information. The above results are concatenated and then transformed using the convolutional transformation function. Another two convolutional transformations change the encoded image into a tensor with the same number of channels as input.

In this way, long-range dependencies can be captured along one spatial direction, while precise position information can be preserved along another spatial direction. The resulting feature maps are then separately encoded into a pair of "direction-aware and position-aware attention maps", which can be complementary applied to the input feature maps to enhance the representation of objects of interest.

3. Datasets and Experiments

3.1. Datasets

3.1.1. Wheat Kernel Dataset

The wheat kernel datasets contain NIRS of various wheat kernels. The wheat kernels were obtained from a seed company in Jiangsu Province, China, in 2019 [35]. They were kept under the same environmental conditions after harvest (dried, packaged in woven plastic bags, and delivered to the laboratory). Spectral images of the wheat kernels were collected by an NIR hyperspectral system with a spectral range from 874 to 1734 nm, and the NIRS were extracted from the hyperspectral images. The dataset can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.575810>.

The experiment in this article involve 200 samples were randomly selected from HM41 and JM919 respectively, totaling 400 samples to the analysis dataset. It is difficult to collect a large number of spectral samples in actual detection tasks, and random sampling of the original dataset simulates this situation. Therefore, this measure will be beneficial for the application of this research to other

spectral datasets. The SPXY method was used to divide the training set and test set at a ratio of 2:1. The training set contains 136 HM41 samples and 132 JM919 samples, a total of 268; the test set contains 64 HM41 samples and 68 JM919 samples, a total of 132 samples.

3.1.2. Yali Pear Dataset

Yali pear browning is a kind of pear heart disease caused by changes in temperature and long storage time during storage and transportation. Yali pears samples were collected from an orchard in Hebei Province, China. After picking, they were transported to the laboratory through cold chain transportation. Before the experiment, these samples were stored at a constant temperature for 24 hours, which can reduce the influence of temperature on spectral acquisition [36]. The pears selected in the experiment are about 65-75 mm in diameter, and the skin has no noticeable discoloration or damage. The Vis-NIR of the samples were collected by Ocean Optics INC, QE-65Pro (Dunedin, FL, USA) high-precision spectrometer. After the spectra collection, the samples were cut one by one along the center line, and three experienced agricultural experts judged whether browning occurred. A total of 495 pear samples were collected, and the spectral collection range was 370-1160 nm, with a total of 1024 data points, including 256 healthy pears and 239 browning pears. The SPXY method was used to divide the training set and test set at a ratio of 2:1. The training set contains 152 healthy pear samples and 178 browning pear samples, a total of 330; the test set contains 104 healthy pear samples and 61 browning pear samples, a total of 165 samples.

3.2. Model Evaluation

The performance of the model is evaluated by classification accuracy (Accuracy), positive sample classification accuracy (RP), and negative sample classification accuracy (RN). The closer the values of the above three evaluation indicators are to 100%, the stronger the classification ability of the model is. The calculation formulas of Accuracy, RP, and RN are as Eq. 8.

$$\begin{aligned} \text{Accuracy} &= (1 - \frac{P_e + N_e}{P + N}) \times 100\% \\ \text{RP} &= (1 - \frac{P_e}{P}) \times 100\% \\ \text{RN} &= (1 - \frac{N_e}{N}) \times 100\% \end{aligned} \quad (8)$$

Where P is the total number of positive samples; N is the total number of negative samples; P_e is the number of positive samples misclassified as negative samples; N_e is the number of negative samples misclassified as positive samples. In the wheat kernel dataset, the HM41 is defined as positive sample, and the JM919 is defined as negative sample. The positive sample in the Yali pear dataset is healthy pear, and the negative sample is browning pear.

3.3. Experiments

The experimental process is shown in Figure 2. First, the raw spectra were pretreated with different preprocessing methods, and the classification models including PLS-DA, RF, and CNN were developed based on SL, CL, and DL modeling strategies, respectively. The preprocessing methods involved in the experiment include no preprocessing (None), Savitzky-Golay filter (SG^{1st}), standard normal variate (SNV), multivariate scattering correction (MSC), and continuous wavelet transform (CWT). Secondly, GADF is used to convert the NIRS to images, and four models including PLS-DA, RF, CNN, and CACNN were built based on the images and are used again for modeling. The results of the above four models are analyzed and compared. Finally, the noisy datasets were introduced by artificially adding noise, and based on the optimal preprocessing method, the PLS-DA, RF, and CACNN models are established and compared, respectively. The robustness of the SL model, CL model, and DL model under different noise influences were discussed. At the same time,

the influence of GADF and CA attention on the models are analyzed by using the saliency visualization method.

The following are the algorithm optimization methods and parameter settings involved in the experiment. For the PLS-DA, the cross-validation method is used to search for the components within 1-30. For the RF, the grid search method was used to optimize the number of decision trees and the maximum depth of decision trees. The search interval of the former is 50-300, with a step size of 50; The search interval for the latter is 10-100, with a step size of 20. In the training of deep learning models, the initial learning rate is 0.0005 and decays by 50% every 200 iterations, for a total of 600 iterations. To prevent overfitting, a dropout optimization method was added after the fully connected layer, with a dropout rate set to 0.25. The structure and parameter settings of CNN are shown in Table 1. The structure of the CACNN network is roughly the same as that in Table 1, with the CA attention layer inserted between Conv1 and Conv2.

Table 1. The structure and parameters of CNN.

Layers	Size	Number	Activation	Output
Input	64*64*1	-	-	-
Conv1	3*3	32	ReLU	64*64*32
Conv2	3*3	32	ReLU	64*64*32
Max-Pooling1	2*2	-	-	32*32*32
Conv3	3*3	64	ReLU	32*32*64
Max-Pooling2	2*2	-	-	16*16*64
Conv4	3*3	128	ReLU	16*16*128
Max-Pooling3	2*2	-	-	8*8*128
GlobalMaxPooling2D	-	-	-	128
Dense1	128	-	ReLU	128
Dense2	2	-	Softmax	2

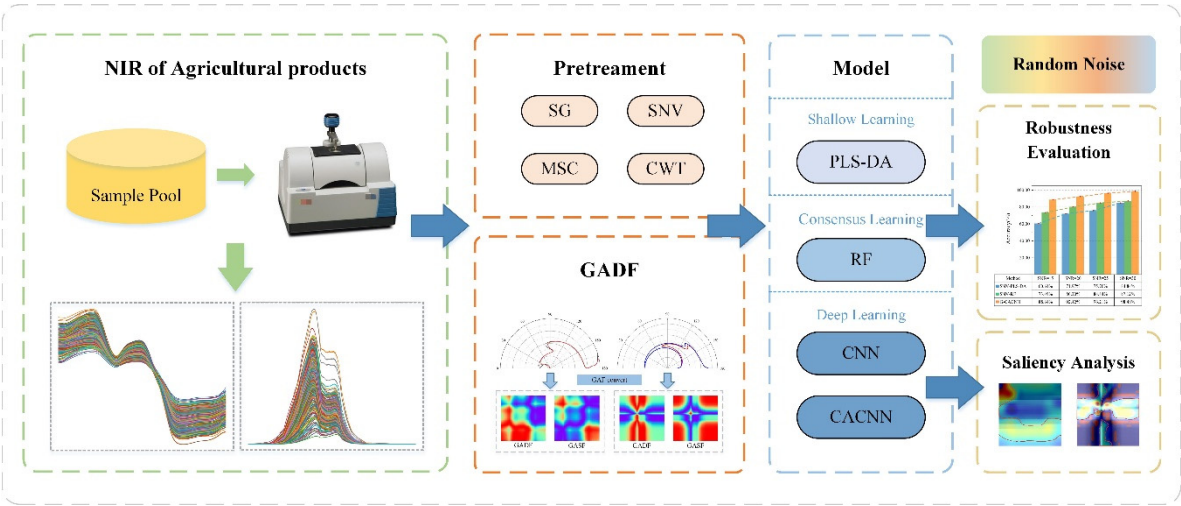


Figure 2. Experimental flow chart.

All data processing and modeling experiments were performed on the personal computer (CPU: intel i5-12400f, GPU: GeForce RTX 2060 Super). The software environment is Python3.7.0, and the DL model is built using the TensorFlow architecture.

4. Results and Analysis

4.1. Spectral Analysis and GAF Converting

The average spectra of the two categories of wheat kernels are shown in Figure 3a. The difference between the average spectra of the two categories of samples in the 870-1100 nm and 1200-1400 nm band is more significant. Absorbance peaks at different bands in the following wavelengths were most noteworthy: 965-985 nm N-H stretching second overtone band associated with proteins, 1130-1200 nm second overtone of C-H stretching, which was related to carbohydrates, 1330-1385 nm combination of C-H stretching and C-H deformation [37].

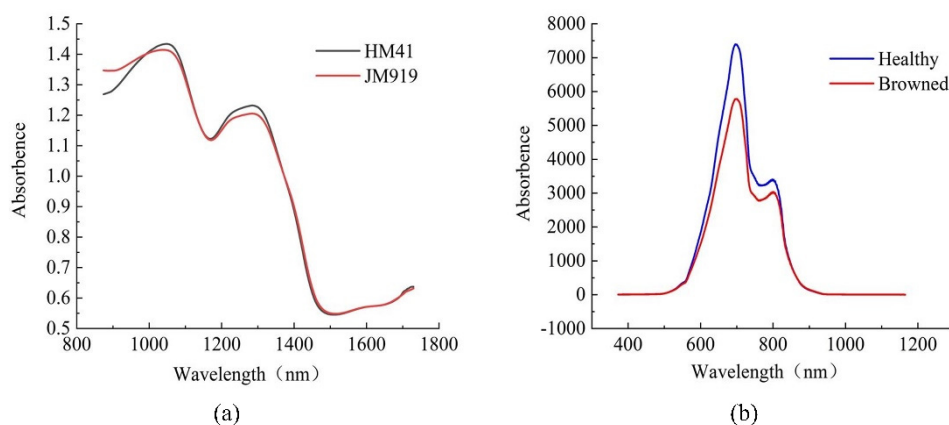


Figure 3. Average spectra. (a) The average spectra of wheat kernel. (b) The average spectra of Yali pears.

The average spectra of the two categories of Yali pears are shown in Figure 3b. The average spectrum of healthy pears is higher than that of brownd pears on the whole, and the band of 600-800 nm is the most obvious, which may be due to the strong absorption of transmitted light by browning tissue inside the fruit. As the spectra show, there are two absorption peaks at approximately 700 and 800 nm. The absorption peak at around 700 nm may result from the stretching and contraction of the fourth overtone of the C-H functional group, while that at around 800 nm may be related to the stretching and contraction of the third overtone of the N-H functional group [38].

The GAF encoding process of the average spectra of the two datasets are shown in Figure 4. and the color change from blue to red corresponds to the increment of the value in the pixel. After the polar coordinate transformation, the NIR numerical fluctuations are transformed into angular changes. The encoded GASF image is symmetrical along the main diagonal, and the main diagonal represents the original 1D NIRS sequence. In the GADF image, the main diagonal is always 0, and the pixel values of two symmetrical points along the main diagonal are opposite to each other, which is caused by the exchange of the difference order. The spectral feature information is enriched with GADF by exploring the difference in the value of different positions in a single spectral sequence.

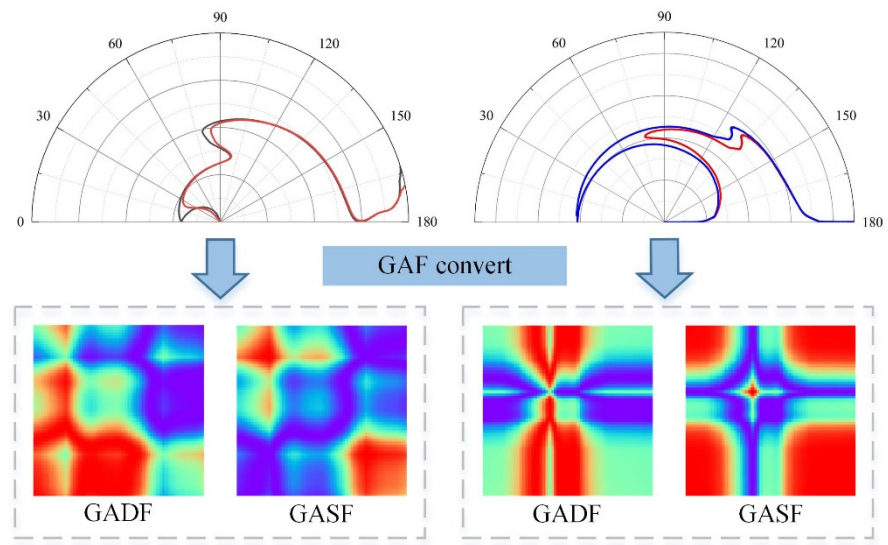


Figure 4. Schematic diagram of GAF converting process.

4.2. Spectra Discriminative Model Analysis

4.2.1. Wheat Kernel Dataset

The results of the PLS-DA, RF, and CNN discrimination models for wheat kernels are shown in Table 2, and the influence of different spectral preprocessing methods on the modeling results of the above various models is analyzed. The standard deviation (SD) between each set of evaluation indicators is used to measure the impact of preprocessing methods on modeling results. From the table, it can be seen that different preprocessing methods have the greatest influence on the PLS-DA model, and the SD of the accuracy is 1.27%. The average accuracy of the three modeling methods is 96.67%, 96.82%, and 97.27%, respectively. The SD of the training accuracy of several preprocessing methods for the RF model is 0.64%. The CNN model is less affected by different preprocessing methods, and the SD of the accuracy is only 0.42%. For the PLS-DA model, the optimal model result of spectra pretreated by SNV or MSC is 97.37%. The analysis results of RF and CNN models varied slightly with the different spectral preprocessing methods and had good stability. Moreover, the CNN model of the original spectra can obtain similar analysis results to the optimal SNV-PLS-DA model. The spectral preprocessing methods had little influence on the analysis results of the DL model, and the model shows better robustness. The results of the deep model on the original spectra were still good.

Table 2. Wheat kernel dataset modeling results.

Classifier	Pretreatment	Accuracy (%)	RP (%)	RN (%)
PLS-DA	None	94.70	96.88	92.65
	SG-1 st	96.21	96.88	95.59
	SNV	97.73	98.44	97.06
	MSC	97.73	98.44	97.06
	CWT	96.97	96.88	97.06
	-	96.67±1.27 ^a	97.50±0.85	95.88±1.92
RF	None	96.21	96.88	95.59
	SG-1 st	96.97	98.44	95.59

CNN	SNV	97.73	98.44	97.06
	MSC	96.97	96.88	97.06
	CWT	96.21	96.88	95.59
	-	96.82±0.64	97.50±0.85	96.18±0.81
	None	96.97	98.44	95.59
	SG-1 st	96.97	98.44	95.59
	SNV	97.73	98.44	97.06
	MSC	97.73	98.44	97.06
	CWT	96.97	96.88	97.06
	-	97.27±0.42	98.13±0.70	96.47±0.81

a: Mean±SD of evaluation indicators

4.2.2. Yali Pear Dataset

The results of the PLS-DA, RF, and CNN discrimination models for Yali pears are shown in Table 3, Similarly, spectral preprocessing methods have different effects on model analysis results. Among them, the preprocessing methods had the greatest influence on the analysis results of the PLS-DA model, and the SD of the accuracy of the modeling results with different preprocessing methods is 4.74%. The RF model and the PLS-DA model are less affected, and the SD of the accuracy of varying preprocessing methods are 0.69% and 0.52%, respectively. The average accuracy of the three modeling methods is 91.58%, 87.51%, and 95.65%, respectively. For the task of identifying Yali pears browning, the accuracy of CNN modeling is significantly better than the other two models. Moreover, the result of CNN without any spectral preprocessing methods is much higher than that of the other two models, which is close to the optimal result of 95.78% of the PLS-DA. The results show that the DL model can complete the qualitative analysis of samples based on NIRS without complex spectral preprocess.

Table 3. Yali pears dataset modeling results.

Classifier	Pretreatment	Accuracy (%)	RP (%)	RN (%)
PLS-DA	None	84.24	80.26	87.64
	SG-1 st	89.70	90.79	88.76
	SNV	95.78	94.73	96.63
	MSC	95.15	94.08	98.07
	CWT	93.03	93.42	92.70
	-	91.58±4.74 ^a	90.66±6.00	92.76±4.62
RF	None	87.27	86.84	87.64
	SG-1 st	86.67	87.50	85.69
	SNV	88.48	88.16	88.76
	MSC	87.27	86.84	87.64
	CWT	87.88	88.81	87.08
	-	87.51±0.69	87.63±0.86	87.36±1.12
CNN	None	95.15	95.39	94.94

SG-1 st	95.78	94.74	96.63
SNV	96.39	96.71	96.07
MSC	95.15	96.05	96.63
CWT	95.78	94.74	96.63
-	95.65±0.52	95.53±0.86	96.18±0.73

a: Mean±SD of evaluation indicators

4.3. Advantages of Image in Modeling

For images, CNN and CACNN are used for modeling and prediction, respectively. And flatten the images into a one-dimensional vector. For vectors, RF and PLS-DA are used for modeling and prediction, respectively. Table 4 shows the results of all models developed with GADF images. Compared with the results with spectra, the accuracy of both PLS-DA and CNN were improved, which shows that GADF can effectively enhance spectral feature information. For several modeling methods, the ability of CNN to process images is superior to that of traditional PLS-DA and RF models, and the CA module can further improve the modeling results of CNN. The method proposed in this paper has achieved the best results in a series of experiments; the accuracy of Wheat kernel classification is 98.48%; the accuracy of Yali pear classification is 99.39%.

Table 4. Modeling Results Based on GADF Images.

Model	Wheat kernel Dataset			Yali pear Dataset		
	Accuracy (%)	RH (%)	RB (%)	Accuracy (%)	RH (%)	RB (%)
G-PLS-DA	95.45	96.88	94.12	90.91	92.31	88.24
G-RF	94.96	95.31	95.45	96.88	94.12	88.52
G-CNN	96.97	97.06	96.35	97.98	98.46	97.06
G-CACNN	98.48	98.44	98.53	99.39	100	98.36

4.4. Optimal Model Analysis

The convolution kernels are used to scan images and extract different features, which have a receptive field. The regional scanning makes the different variables (pixels) no longer independent, which means the different arrangements of variables are bound to affect modeling results. When the spectra are convoluted and scanned, only small-span spectral information can be displayed in the local area. When the convolution operation is performed on the image encoded by the GADF, the small span can reflect the long-range spectral information. CNN is not sensitive to the global position of features, and it is generally agreed that it is biased toward whether there are decisive features in a small range. From this point of view, the dataset encoded by GADF is more suitable for CNN feature capture preferences.

Gradient-weighted Class Activation Mapping (Grad-CAM) is used to visualize the attention distribution of the network: The CAM is obtained according to the different contributions of each channel of the feature map to the decision and combined with the prediction result and the gradient between the feature maps, it shows the important area for the network to make a judgment. The first line of Figure 5 shows the grayscale image input into the network, the grayscale image facilitates coverage of the attention heat map. The second line shows the Grad-CAM of the four methods of CNN, CACNN, G-CNN, and G-CACNN before adding the CA module, and the second line shows the Grad-CAM of the four methods after adding the CA module. The yellow area is paid attention to by the network, and the blue area network lacks attention. The ratio of attention (ROA) refers to the ratio of the area in the entire image where the attention weight exceeds a certain threshold. In this paper, the ROA is used to evaluate the degree of network attention concentration.

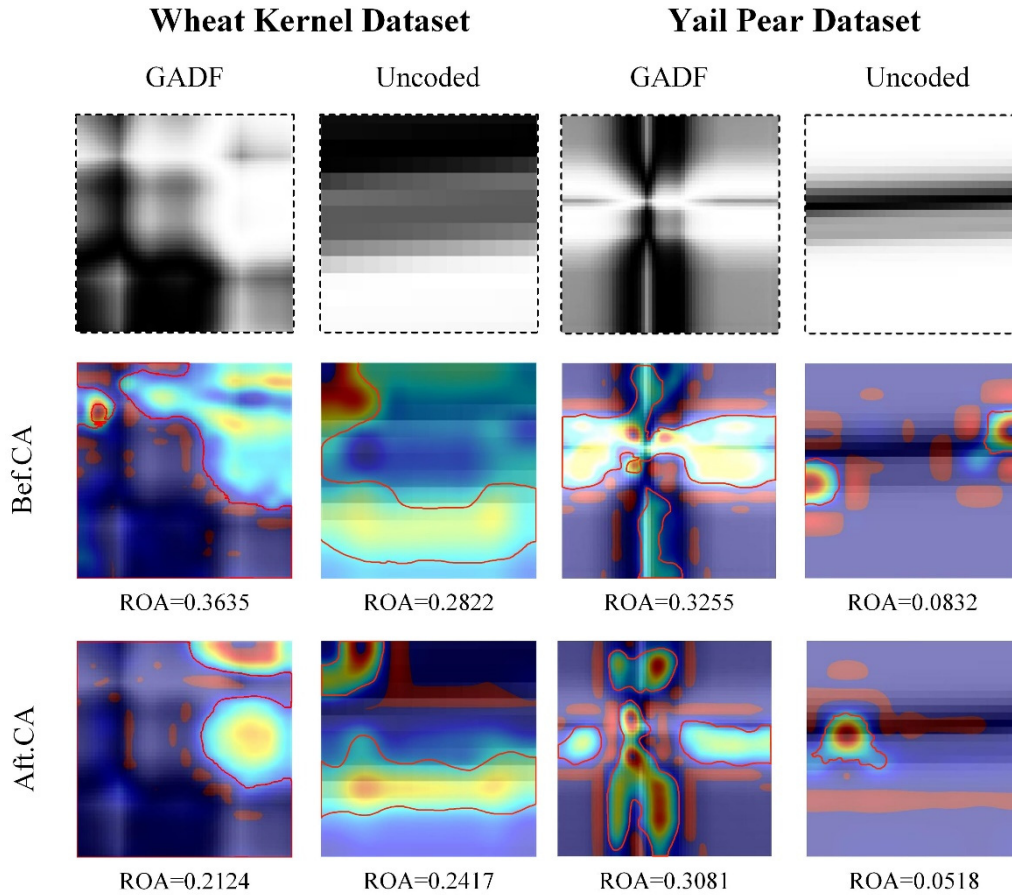


Figure 5. Grad-CAM of deep learning model.

CNN perceives particular features of the uncoded spectrum through the complex mechanism and makes relatively correct judgments but cannot reasonably explain them. However, the attention area of the network on the GADF spectral image is interpretable. Taking the spectra of different samples in the Yali pears dataset as an example, there are differences in the shape of NIR absorption peaks between different types of samples. These differences are encoded by GADF and fall into specific areas of the image. In particular, the network attention area coincides with the different information areas. The CNN can effectively extract different characteristics of absorption peaks from different samples and make judgments, which has been proven. Compared with the DL model of unencoded spectral images, the model trained with GADF spectral images is more interpretable and has higher accuracy (section 4.3).

For the wheat kernel dataset, the ROA of the G-CNN network decreased from 0.3635 to 0.2124 after adding the CA module. For the Yali pear dataset, the ROA of the G-GACNN network decreased from 0.3255 to 0.3081 after adding the CA module. The G-CACNN pays more attention to information regions.

4.5. Robustness Analysis of the Models

To evaluate the robustness of the models, the artificial noise was added to the original spectra, and the signal-to-noise ratio (SNR) was used to evaluate the noise level. The SNR calculation formula is shown in Eq. 9.

$$SNR = 10 \lg \frac{P_s}{P_n} \quad (9)$$

Where P_s is the power of the useful signal, P_n is the power of the noise signal, and $\lg$ represents the logarithm based on 10. This work simulates spectra data acquisition with additive noise under harsh acquisition conditions. Figure 6 shows the average spectra before and after adding noise.

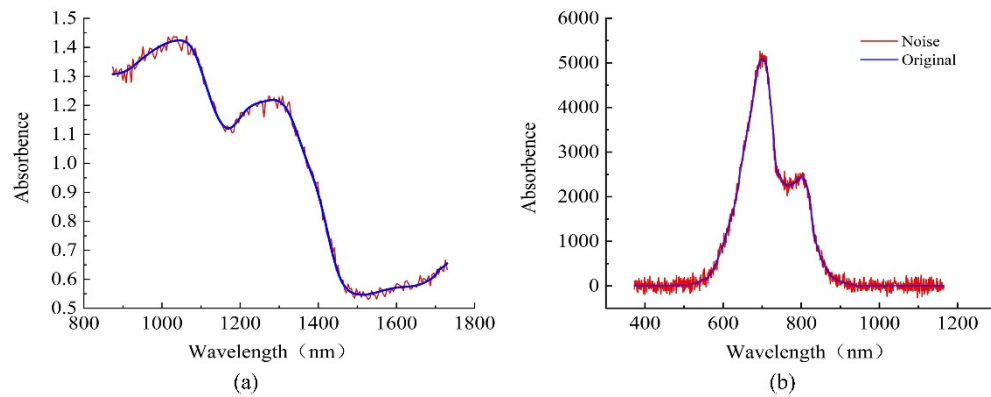


Figure 6. Schematic diagram of additive noise of spectra. (a) Wheat kernel dataset. (b) Yali pear dataset.

The confidence index (CI) is the basis on which networks are ultimately classified. The higher the CI, the greater the likelihood that the unknown sample belongs to the class. Figure 7 illustrates the Grad-CAM and confidence index of No. 7 healthy pear noise (SNR=20) spectral encoding map modeled with G-CNN and G-CACNN. Before using the CA module, the CI is 0.8324. The attention of the network is focused on a few points after adding the CA module and the CI increased to 0.9024. The introduction of the CA module makes CNN achieve better performance. At the same time, the noise immunity of the model is also enhanced.

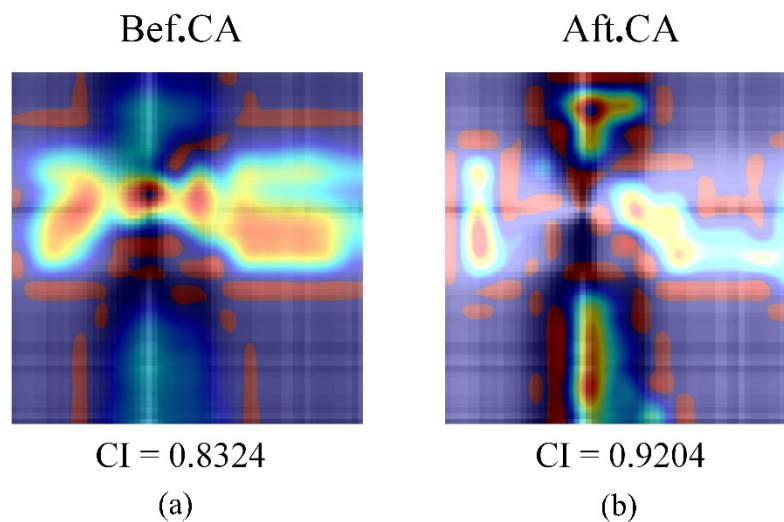


Figure 7. Grad-CAM results with noise spectrum. (a) G-CNN. (b) G-CACNN.

The models were constructed using PLS-DA, and RF after SNV of the noise spectra in the presence of additive noise. Figure 8 shows the comparison of the modeling influence of the preprocessed SL model, the CL model, and the DL model proposed under different levels of noise.

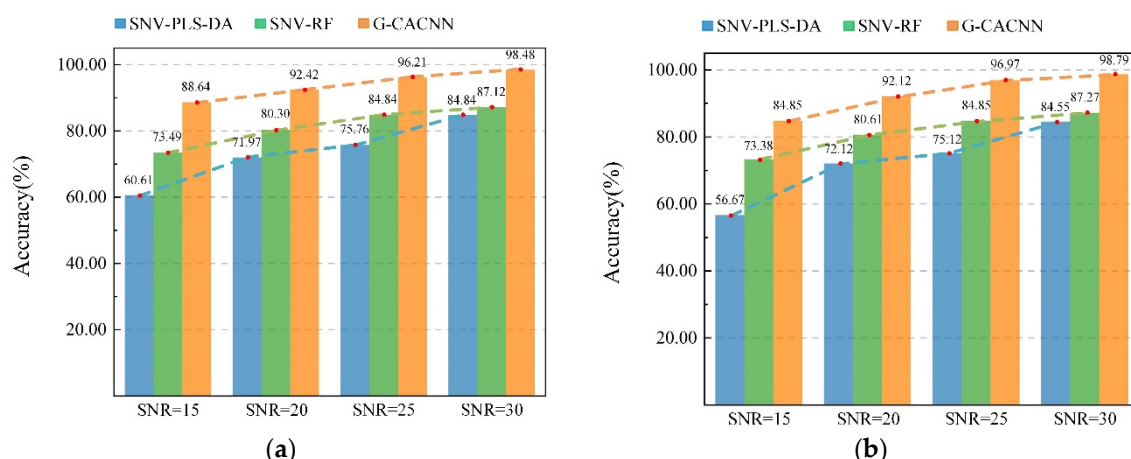


Figure 8. Results of different methods under different levels of noise. (a) Robustness test of the models based on the wheat kernel dataset. (b) Robustness test of the models based on the Pears dataset.

As the Figure8 shows, with the improvement of the SNR, the prediction accuracies of the three types of models have different improvements. The PLS-DA is very sensitive to noise in data, so the preprocessing method cannot eliminate the influence of spectral noise on its analysis results. Under the same level of noise conditions, PLS-DA always has the lowest prediction accuracy among the three models. In contrast, the EL model SNV-RF performs better, which demonstrates its robustness to noisy data. The result of GADF-CACNN is better than SNV-RF at all noise levels. For the wheat kernel dataset, they are 88.64%, 92.42%, 96.21%, and 98.48%. For the Pears dataset, they are 84.85%, 92.12%, 96.97%, and 98.79%. That means the method has good noise immunity and the best robustness when dealing with spectral data with additive noise. At the same time, from the perspective of the changing trend of accuracy, the effect of the change of noise level on the DL model and ensemble model is less than that on the SML model, and the influence of GADF-CACNN is the least, that is also an important feature to demonstrate the robustness operation of the model.

5. Conclusion

In this paper, a robustness NIRS modeling method based on GADF and CACNN methods is proposed. The GADF method converts 1D spectral signals into 2D images, enriching and highlighting the spectral characteristics. The CACNN network is designed to make full use of the excellent feature extraction ability of the DL model, and the CA module is added to improve the prediction accuracy of the deep structure model. Without using the preprocessing, the GADF-CACNN model has an accuracy of 99.39% in the Pear browning classification and an accuracy of 98.48% in the wheat kernel classification. In addition, the robustness analysis of the traditional shallow structure model and the GADF-CACNN model is carried out, and the experimental results show that the result of the DL method is better than that of the SML model and the EL model. The method proposed in this paper has the characteristics of simple modeling steps and good model interpretability with good accuracy and robustness for NIRS analysis

Author Contributions: JR: Supervision, formal analysis. YX: Experiments, software, methodology, writing original and editing. XC: Data curation, Visualization. YH: Conceptualization, funding acquisition, writing review and editing. All authors contributed to this article and approved the submitted version.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 31960497, Jiangxi Provincial Natural Science Foundation of China, grant number 20212BAB204009 & 20202ACB211002.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to the data may need to be used for subsequent research or further analysis.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mishra, P.; Woltering, E. Semi-supervised robust models for predicting dry matter in mango fruit with near-infrared spectroscopy. *Postharvest Biology and Technology* **2023**, *200*, doi:10.1016/j.postharvbio.2023.112335.
2. Esquerre, C.A.; Achata, E.M.; Garcia-Vaquero, M.; Zhang, Z.; Tiwari, B.K.; O'Donnell, C.P. Use of an NIR MEMS spectrophotometer and visible/NIR hyperspectral imaging systems to predict quality parameters of treated ground peppercorns. *Lwt-Food Science and Technology* **2020**, *131*, doi:10.1016/j.lwt.2020.109761.
3. Chadalavada, K.; Anbazhagan, K.; Ndour, A.; Choudhary, S.; Palmer, W.; Flynn, J.R.; Mallayee, S.; Pothu, S.; Prasad, K.V.S.V.; Varijakshapanikar, P.; et al. NIR Instruments and Prediction Methods for Rapid Access to Grain Protein Content in Multiple Cereals. *Sensors* **2022**, *22*, doi:10.3390/s22103710.
4. Cozzolino, D. The Ability of Near Infrared (NIR) Spectroscopy to Predict Functional Properties in Foods: Challenges and Opportunities. *Molecules* **2021**, *26*, doi:10.3390/molecules26226981.
5. Rahmawati, L.; Widodo, S.; Kurniadi, D.P.; Daud, P.; Triyono, A.; Sriharti, Susanti, N.D.; Mayasti, N.K.I.; Indriati, A.; Yulianti, L.E.; et al. Determination of colorant type in yellow tofu using Vis-NIR and SW-NIR spectroscopy. *Food Science and Technology* **2023**, *43*, e112422-e112422, doi:10.1590/fst.112422.
6. Zareef, M.; Chen, Q.; Hassan, M.M.; Arslan, M.; Hashim, M.M.; Ahmad, W.; Kutsanedzie, F.Y.H.; Agyekum, A.A. An Overview on the Applications of Typical Non-linear Algorithms Coupled With NIR Spectroscopy in Food Analysis. *Food Engineering Reviews* **2020**, *12*, 173-190, doi:10.1007/s12393-020-09210-7.
7. Assi, S.; Arafat, B.; Lawson-Wood, K.; Robertson, I. Authentication of Antibiotics Using Portable Near-Infrared Spectroscopy and Multivariate Data Analysis. *Applied Spectroscopy* **2021**, *75*, 434-444, doi:10.1177/0003702820958081.
8. Bec, K.B.; Grabska, J.; Huck, C.W. NIR spectroscopy of natural medicines supported by novel instrumentation and methods for data analysis and interpretation. *Journal of Pharmaceutical and Biomedical Analysis* **2021**, *193*, doi:10.1016/j.jpba.2020.113686.
9. Yin, L.; Zhou, J.; Chen, D.; Han, T.; Zheng, B.; Younis, A.; Shao, Q. A review of the application of near-infrared spectroscopy to rare traditional Chinese medicine. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2019**, *221*, doi:10.1016/j.saa.2019.117208.
10. Pasquini, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta* **2018**, *1026*, 8-36, doi:<https://doi.org/10.1016/j.aca.2018.04.004>.
11. Zhang, J.; Li, B.; Hu, Y.; Zhou, L.; Wang, G.; Guo, G.; Zhang, Q.; Lei, S.; Zhang, A. A parameter-free framework for calibration enhancement of near-infrared spectroscopy based on correlation constraint. *Analytica Chimica Acta* **2021**, *1142*, 169-178, doi:<https://doi.org/10.1016/j.aca.2020.11.006>.
12. Jiao, Y.; Li, Z.; Chen, X.; Fei, S. Preprocessing methods for near-infrared spectrum calibration. *Journal of Chemometrics* **2020**, *34*, doi:10.1002/cem.3306.
13. Cai, W.; Li, Y.; Shao, X. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems* **2008**, *90*, 188-194, doi:<https://doi.org/10.1016/j.chemolab.2007.10.001>.
14. Li, H.; Liang, Y.; Xu, Q.; Cao, D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta* **2009**, *648*, 77-84, doi:<https://doi.org/10.1016/j.aca.2009.06.046>.
15. Wang, W.; Yun, Y.; Deng, B.; Fan, W.; Liang, Y. Iteratively variable subset optimization for multivariate calibration. *RSC Advances* **2015**, *5*, 95771-95780, doi:10.1039/C5RA08455E.
16. Yun, Y.-H.; Li, H.-D.; Deng, B.-C.; Cao, D.-S. An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends in Analytical Chemistry* **2019**, *113*, 102-115, doi:<https://doi.org/10.1016/j.trac.2019.01.018>.
17. Shi, S.J.; Zhao, D.; Pan, K.Q.; Ma, Y.Y.; Zhang, G.Y.; Li, L.A.; Cao, C.G.; Jiang, Y. Combination of near-infrared spectroscopy and key wavelength-based screening algorithm for rapid determination of rice protein content. *Journal of Food Composition and Analysis* **2023**, *118*, doi:10.1016/j.jfca.2023.105216.
18. Fu, D.; Li, Q.; Chen, Y.; Ma, M.; Tang, W. Assessment of integrated freshness index of different varieties of eggs using the visible and near-infrared spectroscopy. *International Journal of Food Properties* **2023**, *26*, 155-166, doi:10.1080/10942912.2022.2158866.
19. Hu, L.; Yin, C.; Ma, S.; Liu, Z. Vis-NIR spectroscopy Combined with Wavelengths Selection by PSO Optimization Algorithm for Simultaneous Determination of Four Quality Parameters and Classification of Soy Sauce. *Food Analytical Methods* **2019**, *12*, 633-643, doi:10.1007/s12161-018-01407-1.

20. Chen, H.; Lin, Z.; Tan, C. Random subspace-based ensemble modeling for near-infrared spectral diagnosis of colorectal cancer. *Analytical Biochemistry* **2019**, *567*, 38-44, doi:10.1016/j.ab.2018.12.009.
21. Zhao, L.M.; Liu, S.M.; Chen, X.F.; Wu, Z.W.; Yang, R.; Shi, T.T.; Zhang, Y.L.; Zhou, K.W.; Li, J.G. Hyperspectral Identification of Ginseng Growth Years and Spectral Importance Analysis Based on Random Forest. *Applied Sciences-Basel* **2022**, *12*, doi:10.3390/app12125852.
22. Azad, M.; Khaled, F.; Pavel, M. A NOVEL APPROACH TO CLASSIFY AND CONVERT 1D SIGNAL TO 2D GRAYSCALE IMAGE IMPLEMENTING SUPPORT VECTOR MACHINE AND EMPIRICAL MODE DECOMPOSITION ALGORITHM. *International Journal of Advanced Research* **2019**, *7*, 328-335, doi:10.21474/IJAR01/8331.
23. Yilmaz, B.H.; Yilmaz, C.M.; Kose, C. Diversity in a signal-to-image transformation approach for EEG-based motor imagery task classification. *Medical & Biological Engineering & Computing* **2020**, *58*, 443-459, doi:10.1007/s11517-019-02075-x.
24. Naz, M.; Shah, J.H.; Khan, M.A.; Sharif, M.; Raza, M.; Damaševičius, R. From ECG signals to images: a transformation based approach for deep learning. *PeerJ. Computer science* **2021**, *7*, e386, doi:10.7717/peerj-cs.386.
25. Hao, Y.; Zhang, C.; Li, X.; Lei, Z. Establishment of online deep learning model for insect-affected pests in "Yali" pears based on visible-near-infrared spectroscopy. *Frontiers in Nutrition* **2022**, *9*, doi:10.3389/fnut.2022.1026730.
26. Pu, H.; Yu, J.; Sun, D.-W.; Wei, Q.; Li, Q. Distinguishing pericarpium citri reticulatae of different origins using terahertz time-domain spectroscopy combined with convolutional neural networks. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2023**, *299*, 122771.
27. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. Available online: (accessed on
28. Wang, Z.; Oates, T. Imaging Time-Series to Improve Classification and Imputation. *AAAI Press* **2015**.
29. Lee, L.C.; Liong, C.-Y.; Jemain, A.A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst* **2018**, *143*, 3526-3539, doi:10.1039/C8AN00599K.
30. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5-32, doi:10.1023/A:1010933404324.
31. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *23*.
32. Lee, S.; Choi, H.; Cha, K.; Chung, H. Random forest as a potential multivariate method for near-infrared (NIR) spectroscopic analysis of complex mixture samples: Gasoline and naphtha. *Microchemical Journal: Devoted to the Application of Microtechniques in all Branches of Science* **2013**.
33. Yao, G.; Lei, T.; Zhong, J. A review of Convolutional-Neural-Network-based action recognition. *Pattern Recognition Letters* **2019**, *118*, 14-22, doi:10.1016/j.patrec.2018.05.018.
34. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Available online: (accessed on
35. Zhou, L.; Zhang, C.; Taha, M.F.; Wei, X.; He, Y.; Qiu, Z.; Liu, Y. Wheat Kernel Variety Identification Based on a Large Near-Infrared Spectral Dataset and a Novel Deep Learning-Based Feature Selection Method. *Frontiers in Plant Science* **2020**, *11*, doi:10.3389/fpls.2020.575810.
36. Liu, Y.; Chen, M.; Hao, Y. Spectral diagnostic technology and its application in agricultural product quality testing. *Journal of East China Jiaotong University* **2018**, *35*, 1-7, doi:10.16749/j.cnki.jecjtu.2018.04.001.
37. Wadood, S.A.; Guo, B.; Zhang, X.; Wei, Y. Geographical origin discrimination of wheat kernel and white flour using near-infrared reflectance spectroscopy fingerprinting coupled with chemometrics. *International Journal of Food Science & Technology* **2019**.
38. Xiaobo, Z.; Jiewen, Z.; Povey, M.J.W.; Holmes, M.; Hanpin, M. Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta* **2010**, *667*, 14-32, doi:<https://doi.org/10.1016/j.aca.2010.03.048>.

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