

# Mark-Spectra: A convolutional neural network for quantitative spectral analysis overcoming spatial relationships

Yueling Wang, Minzan Li, Ronghua Ji, Minjuan Wang, Yao Zhang, Lihua Zheng <sup>\*</sup>

*Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China*



## ARTICLE INFO

**Keywords:**  
Chemometrics  
Convolutional neural network  
Spectroscopy  
Spatial relationships

## ABSTRACT

Spectral analysis is one of the most important and widely used methods for chemometrics in the field of agriculture, and convolutional neural network (CNN) models have achieved excellent performance on spectral analysis. The critical drawback of the CNN approach is that it preserves the spatial relationships among adjacent wavelengths, which contribute to collinearity and redundancies rather than relevant effective information. To confirm this observation, the distribution of characteristic wavelengths extracted by different methods (include F-test, importance weights, and CNN) are visualized in this paper. A convolutional neural network for quantitative spectral analysis, named Mark-Spectra, is presented to overcome spatial relationships and to improve the model performance. A layer (Mark layer) is introduced as part of Mark-Spectra, which is used to overcome spatial relationship of raw spectral data. Mark-Spectra model is compared with three CNN models using three open accessed visible and near infrared spectroscopic datasets (corn, wheat and soil). Mark-Spectra model outperforms the other three convolutional neural network models on two datasets (except dataset of wheat, due to lesser number of features), and it cost much less training time than the others. In addition, this paper compares Mark-Spectra with two classical neural network-based algorithms, principal component analysis – artificial neural network (PCA-ANN) and extreme learning machine (ELM). Mark-Spectra performed best in soil dataset, and ELM performed best in corn and wheat datasets, respectively. These results can illustrate that Mark-Spectra is still limited with the characteristic of raw spectral data (e.g., the number of samples and features), which is a fundamental fact of deep learning-based methods, but it performed better than the other CNN models and reduced the dependence of sample size due to overcoming spatial relationships.

## 1. Introduction

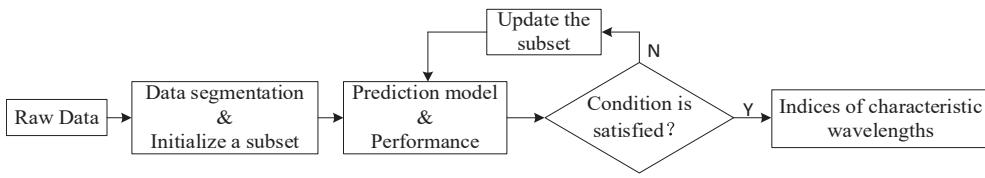
Non-destructive detection of elemental contents is an important application of spectral information detection technology in chemometrics fields. Spectroscopic techniques have the advantages of fast detection speed, low cost, and non-destructive (Li et al., 2020), and it has been widely used in agricultural products (Binetti et al., 2016; An et al., 2015; Zhang et al., 2016), pharmaceuticals (Herrick et al., 2017), petrochemical (Palou et al., 2017), and soil (An et al., 2012; Wang et al., 2021a,b; Zhang et al., 2019a,b,c). Numerous different technologies and algorithms for predicting elemental contents via spectral data have been studied and developed (Malek et al., 2018; Wang et al., 2020; Yang et al., 2019). After literature search and analysis, performance on predicting elemental contents was mainly depended on 2 aspects: data quality and data properties. Quality of the raw spectral data is widely noticed during spectroscopic analysis, and it is usually decreased by various signals

from environment and artifacts (Gerretzen et al., 2016). As for data properties of spectral information, especially in deep-learning algorithms, it is often disregarded. The typical spectral variables contain thousands of wavelengths. Moderate or high cross correlations exist among the independent spectral variables (i.e., multi-collinearity or collinearity). Due to collinearity, many wavelengths are redundant for predicting to-be-detected elements (i.e., target elements), which needs more computational resource and hampers development of portable devices (An et al., 2015; Wang et al., 2021b; Zhou et al., 2019). In other word, these large number of spectral variables often contribute to collinearity and redundancies rather than providing relevant effective information (Ng et al., 2019; Wang et al., 2021a). Therefore, to obtain a better performance prediction model, this study is conducted on these two aspects mentioned above.

In order to promote the quality of spectral data, preprocessing is usually introduced to remove the noise in raw data. The preprocessing

\* Corresponding author.

E-mail address: zhenglh@cau.edu.cn (L. Zheng).



**Fig. 1.** Schematics overview of selecting characteristic wavelengths according importance weights.

methods can be divided into general-purpose and dedicated types. Generic preprocessing methods for visible and near infrared spectral analysis include baseline correction, multiplicative scatter correction, scaling and so on (Jansen et al., 2013). External parameter orthogonalization (EPO) is one of the most typical dedicated approaches, which can be used for removing the effect of soil moisture from NIR diffuse reflectance (Minasny et al., 2011). Besides, wavelet analysis and continuum removal technology are also used for removing different noise sources from raw spectral data (Zhang et al., 2016). An appropriate pre-treatment method can improve the quality of spectral data, but misuse of preprocessing approaches can only increase processing workflow but decrease the modeling performance (Zhang et al., 2019a,b,c).

Deep learning approaches can discover intricate structures in high-dimensional data, reducing the need from prior knowledge and human effort in feature engineering (Yang et al., 2019). To avoid the drawbacks of preprocessing, more and more end-to-end deep learning approaches are used for quantitative spectral analysis (Bjerrum et al., 2017; Fukuhara et al., 2019; Zhang et al., 2020). Among them, the convolutional neural network (CNN) has shown strong fitting capability and been widely employed. Convolutional neural networks learn to extract high level representations hierarchically from low level features while maintaining the underlying spatial relationships (Cang and Wei, 2017), and it usually can be assessed by the sizes of convolution kernels. In fact, it is the reason that convolutional neural network processes a set of wavelengths rather than one waveband, and the correlations among adjacent wavelengths, which are one of the most important factors that affect the performance of prediction model, cannot be avoided. In other word, CNN model ignores the properties of spectral data (i.e., collinearity). In addition, complex structures and handling of high-dimensional data require more training to obtain satisfactory repeatability and accuracy, which makes the process more time consuming and computational overhead. Inspired by the application of CNN in computer vision, the small number of key-wavelengths (i.e., characteristic wavelengths or sensitive wavebands Abulaiti et al., 2020; An et al., 2015; Zhang et al., 2019a,b,c) can be marked and they can be used to predict target elements.

In this paper, a convolutional neural network with marked characteristic wavelengths of target elements is developed for quantitative spectral analysis, which is named Mark-Spectra. Mark-Spectra introduces an extra layer ‘Mark’ before convolutional process, which could screen characteristic wavelengths and obtain the indices of discrete wavelengths according to importance weights of different wavelengths. Firstly, the distributions of characteristic wavelengths are visualized by different methods (include F-test, importance weights and CNN), which displayed the effect of extracting characteristic wavelengths. Then, the impact of characteristic size on the accuracy and training time of Mark-Spectra approach is evaluated. Furthermore, Mark-Spectra is tested by three datasets and its model accuracy is compared with three CNN models and two conventional neural network-based algorithms (include PCA-ANN and ELM) on raw data. The results demonstrate that Mark-Spectra approach improves quantitative analysis of spectroscopic data on the datasets containing adequate number of features. On the one hand, Mark-Spectra inherits the advantages of convolutional neural networks. On the other hand, Mark-Spectra obtains better performance of prediction and reduces the dependence of sample size.

## 2. Materials and methods

### 2.1. Methods for selecting characteristic wavelengths of target elements

#### 2.1.1. Selecting characteristic wavelengths with F-test

Correlation analysis between the wavelengths and the target elements is one of the most common approaches for selecting characteristic wavelengths, and F-test is presented as one representative of methods based on correlation coefficients (Fernández-Ugalde et al., 2020; Jia et al., 2017). In F-test, linear regression diagnostics is first performed to check how well the raw spectral data meets the assumptions of the linear regression, and then the testing is carried out to determine whether the prediction effect of the target content is significant in the linear model. F-test is completed in 3 steps, which is used in this paper. First step, correlations between spectral data ( $X$ ) and target elements ( $Y$ ) are calculated using the Pearson correlation coefficient ( $r$ ), which is presented in Eq. (1).

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (1)$$

where  $\bar{X}$  is the arithmetic mean of  $X$ , and  $\bar{Y}$  is the arithmetic mean of  $Y$ .

The second step, Pearson correlation coefficient ( $r$ ) is transformed to F-value ( $F$ ) by applying the Eq. (2).

$$F = \frac{r^2}{1 - r^2} * d_f \quad (2)$$

where  $d_f$  is the degree of freedom of dependent variables relative to the target elements, and its calculation fulfills the Eq. (3).

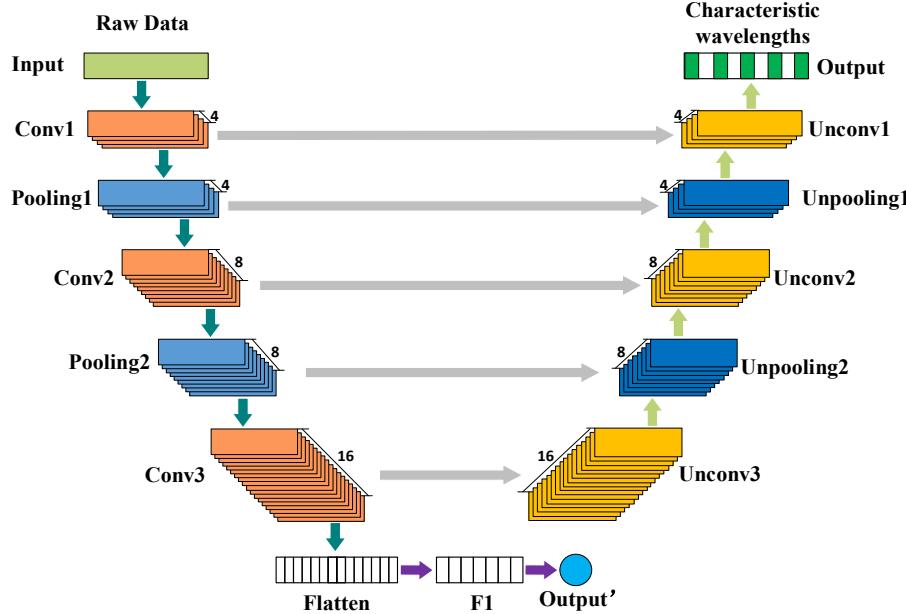
$$d_f = y_n - 1 \quad (3)$$

where  $y_n$  is sample size.

The third step, the wavelengths are sorted according to F-values (from large to small), and each wavelength is sequentially sampled and screened until the number of qualified wavelengths meets the setting values. At this point, screening characteristic wavelengths with F-test is completed.

#### 2.1.2. Selecting characteristic wavelengths with importance weights

Characteristic wavelengths are always allocated larger weights in prediction models, based on this concept, importance weights of wavelengths can be used to select characteristic wavelengths from raw spectral data. The algorithm (shown in the Fig. 1) is selected to screen the characteristic wavelengths of target elements, which has own built-in feature selection techniques, and performs variable selection as a part of the learning procedure. Meanwhile, the number of characteristic wavelengths can be limited by defined amount. In this algorithm, raw spectral data is randomly partitioned into a training/validation set and a test set first, and a subset is initialized as characteristic wavelengths. And then, the prediction model is built and performance is obtained based on the dataset. We use ordinary least square estimation (OLSE) as the prediction model in this algorithm. If the termination condition (defined amount) is satisfied, the indices of characteristic wavelengths are outputted. If defined number of characteristic wavelengths cannot meet the termination condition, alternative subset will be updated and screening is repeated in successive iterations until the stopping criterion



**Fig. 2.** Diagram of selecting the characteristic wavelengths based on CNN. Dark-yellow modules represent convolutional operations (labeled as Conv1, Conv2, and Conv3), and light-yellow modules represent deconvolutional operations (labeled as Unconv1, Unconv2, and Unconv3). Light-blue modules represent max-pooling operations (labeled as Pooling1 and Pooling2), dark-blue modules represent un-pooling operations (labeled as Unpooling1 and Unpooling2). The full connected layer (labeled as F1) uses linear model as modeling approach.

**Table 1**  
Hyperparameters setting in feature extractions and visualizations of four spectra datasets.

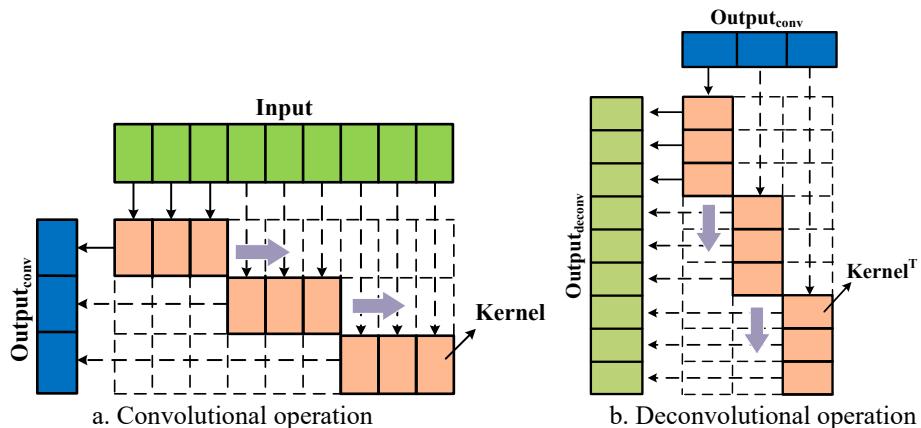
Hyperparameter	Corn	Wheat	Soil
Kernel size 1	7	3	7
Kernel size 2	3	2	5
Kernel size 3	5	3	3
Stride 1	3	3	5
Stride 2	3	2	3
Stride 3	3	2	2
Hidden number	16	32	64
Dropout rate	0.2	0.1	0.5
Batch size	32	128	256
Learning rate	0.01	0.01	0.01
Learning rate decay	0.001	0.001	0.001

is met.

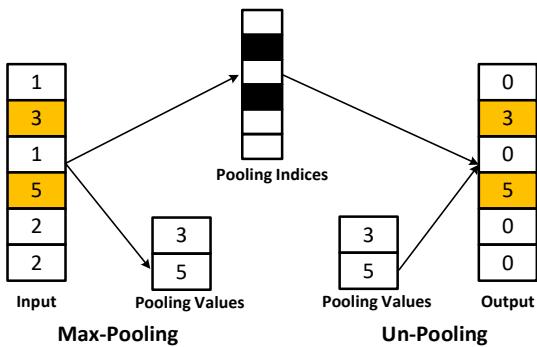
#### 2.1.3. Selecting characteristic wavelengths with convolutional neural networks

Convolutional Neural Networks (CNNs) are a popular deep learning

approach that has been well applied in image recognition (He et al., 2016), video analysis (Karpathy et al., 2014), and so on (Aikawa, 1992). In recent years, CNNs are used for spectral analysis in studies (Riese and Keller, 2019; Wang et al., 2020), and they can discover intricate structures in high-dimensional data with sufficient samples. In order to extract the characteristic wavelengths of target elements, these operations of feature extraction and feature visualization are needed. Among those, feature extraction is used for determining the characteristic wavelengths, and feature visualization is used for obtaining the indices of the characteristic wavelengths (Wang et al., 2021a). Therefore, a U-shaped network (shown in Fig. 2) is built to select the characteristic wavelengths (Weng and Zhu, 2021). The input of the model is raw spectroscopic data and the output is indices of characteristic wavelengths. The feature extraction stage contains 3 convolutional layers, 2 pooling layers and 1 fully connected layer, and feature visualization stage contains 3 deconvolutional layers, 2 un-pooling layers. Of those, deconvolution (also known as transposed convolutions) (Shi et al., 2016) and un-pooling blocks are the mirrored version of the corresponding convolution and pooling blocks, which are meant to increase the dimensionality of feature maps. Characteristic wavelengths are



**Fig. 3.** Diagram of convolutional and deconvolutional operations. The input data (labeled as Input) is of size 9, the filter (labeled as Kernel) is of size 3, and the output of convolutional operations (labeled as Output<sub>conv</sub>) is of size 3. Deconvolutional operation got its name because it is the backward propagation of a convolutional operation.



**Fig. 4.** Diagram of max-pooling and un-pooling operations. An example of pooling and un-pooling operation has pooling size of three. During the process of max-pooling, the maximum of each pooling size (i.e., 3 and 5) and indices (labeled as Pooling indices) are recorded. As for the process of un-pooling, the pooling values fill up according to the indices and the others are filled with zero.

selected according to predicting results of the target element (Output'). Hyperparameters are set in Table 1 to extract and visualize the characteristic wavelengths.

Fig. 3 shows the schematic diagram of convolution and deconvolution for ease of understanding. In this example, the raw spectral data is the input of convolutional operations. The size of the convolutional kernel is 3 and the stride length is 3. Through the calculation of convolutional kernel, the outputs of convolutional operation (Output<sub>conv</sub>) are obtained, and one neuron in the Output<sub>conv</sub> covers a receptive field of 3 original spectral wavelengths, respectively. As for deconvolutional operations, the input is the feature maps obtained by convolution (Output<sub>conv</sub>). Deconvolutional kernel is the transpose convolutional kernel, that's why that it is also called transposed convolution (Shi et al., 2016). The output of deconvolutional calculation (Output<sub>deconv</sub>) has the same size of the raw spectral data, but the values of Output<sub>deconv</sub> have changed.

Un-pooling can be thought of an inverse operation to pooling and it expands the information that is compressed by pooling operations (shown in Fig. 4). Values and indices of the feature maps will be recorded in process of max-pooling, the max values will be filled up in the indices and the others are set to zeros during un-pooling processing.

## 2.2. Mark-Spectra model architecture

To improve predictive efficacy of CNNs-based methods, increasing the numbers of convolution and pooling layer is the most common method, which are able to capture more features (shown in Fig. 3.a). More convolution and pooling layers require more samples to train for good results, but adequate samples are somewhat harder to collect due to the tedious and complicated procedure in real applications. Besides,

more convolution and pooling layers mean that they need more computational resource and time. Different from the methods mentioned above, we adopted a novel approach to achieve the same goal, which is by introducing the Mark-Spectra model.

Mark-Spectra model developed in this study is based on a convolutional neural network (CNN) and selecting characteristic wavelengths (shown in Fig. 5). The input of Mark-Spectra model is raw spectral data, and the output is the target element content (e.g., the protein content of corn, soil total nitrogen content and so on). Comparison with general CNNs, Mark layer (labeled as Mark) is introduced as part of Mark-Spectra model for reducing the impact of collinearity and redundancies between adjacent wavelengths. To overcome the interference of noise (e.g., introduced by instrumental and experimental artifacts) and extract features, three convolutional layers (labeled as Conv1, Conv2, and Conv3), two max-pooling layers (labeled as Pooling1 and Pooling2), a flatten layer (labeled as Flatten) and one fully connected layer (labeled as F1) are built. The rectified linear unit (ReLU, Eq. (4)) is selected as the activation function for the convolutional layers and fully connected layer. A batch normalization (BN) is used after convolution layer and Flatten layer. Adam optimizer is used to search the local minimum of the objective function. Mean squared error (MSE) is adopted as the loss function, which is presented in Equation (5).

$$f(x) = \begin{cases} x, & (x \geq 0) \\ 0, & (x < 0) \end{cases} \quad (4)$$

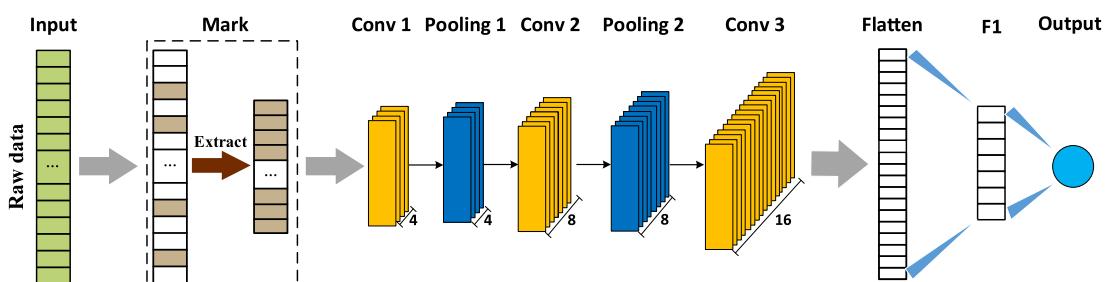
$$\text{Loss} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (5)$$

where  $y_i$  and  $\hat{y}_i$  are measured values and predicted values respectively;  $n$  is the number of samples in the training set;  $i$  is the  $i$ -th sample.

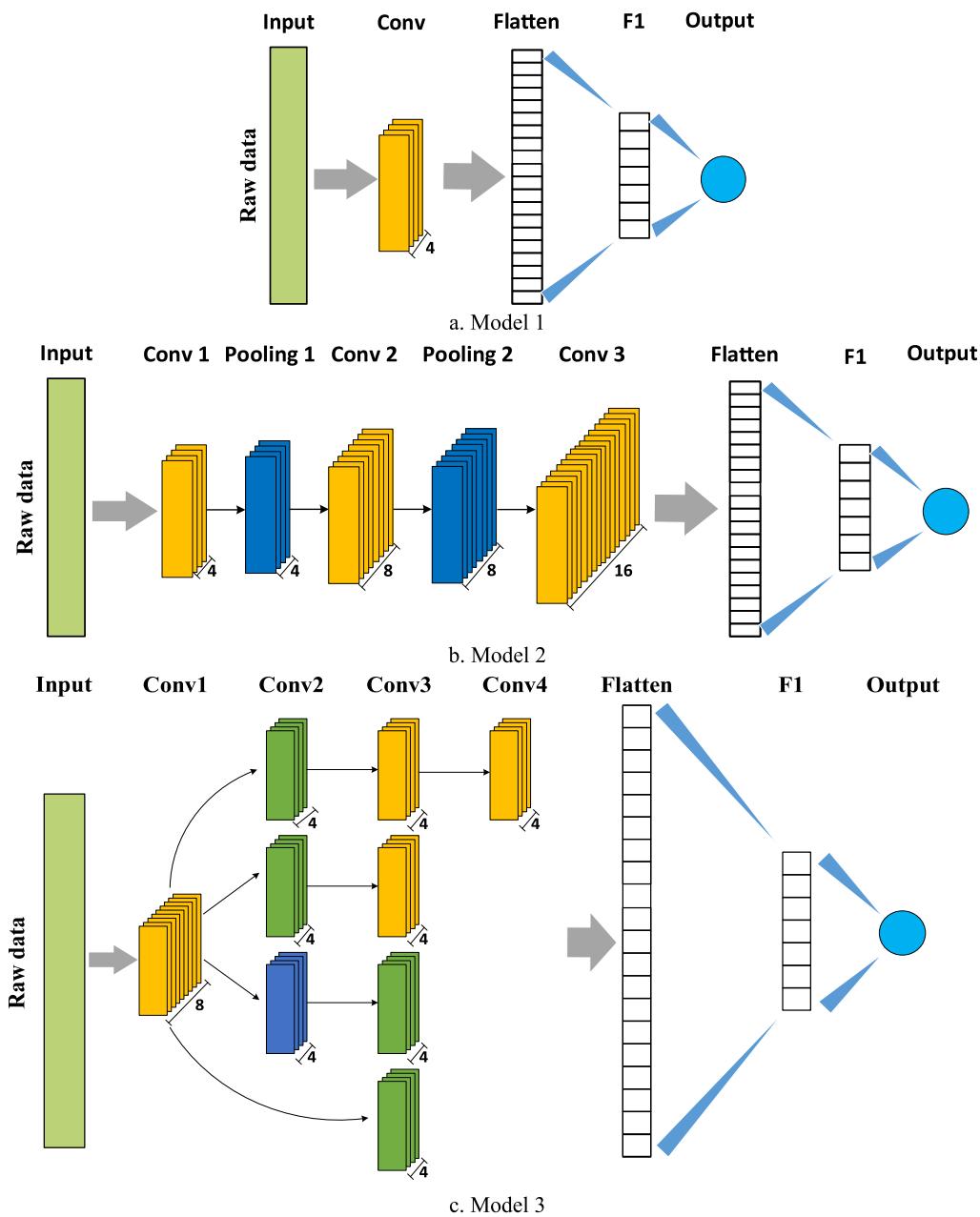
The most critical component of this model is Mark layer, which is used to obtain discrete distribution of characteristic wavelengths and avoid the effect of collinearity. Different from visual recognition tasks, characteristic wavelengths are not able to be picked directly. Thus, an ordinary least square estimation is introduced to evaluate the effect of each wavelength. In this process, different wavelengths are entered as input and the coefficient of determination ( $R^2$ , Eq. (6)) is used to evaluate the importance of different wavelengths. Note that in the Equation (6),  $R^2$  can vary from  $-\infty$  to 1 (i.e., it can yield negative values).

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

where  $y_i$  and  $\hat{y}_i$  are measured values and predicted values respectively;  $n$  is the number of samples in the training set;  $\bar{y}$  is the arithmetic mean of  $y_i$ . This way, Mark layer can capture limited number of discrete characteristic wavelengths to reduce computational resource and time, which are ordered by the importance of wavelengths.



**Fig. 5.** The structure of Mark-Spectra model. Light-green module represents the raw spectral data, and it is the input of Mark-Spectra. The importance of different wavelengths is computed and these wavelengths satisfied conditions are marked in brown, which are extracted as the output of the mark layer and other layers will follow it. Yellow modules represent convolution (Conv1, Conv2, and Conv3), and blue modules represent max-pooling (Pooling1 and Pooling2). Different convolutional layers have different numbers of filters (i.e., 4, 8, and 16); pooling layers have the same filters as the previous convolutional layer.



**Fig. 6.** Architectures of three CNNs. The yellow modules represent the general convolution, the blue modules represent max-pooling operations. In model 3, the dark-green modules represent  $1 \times 1$  convolution.

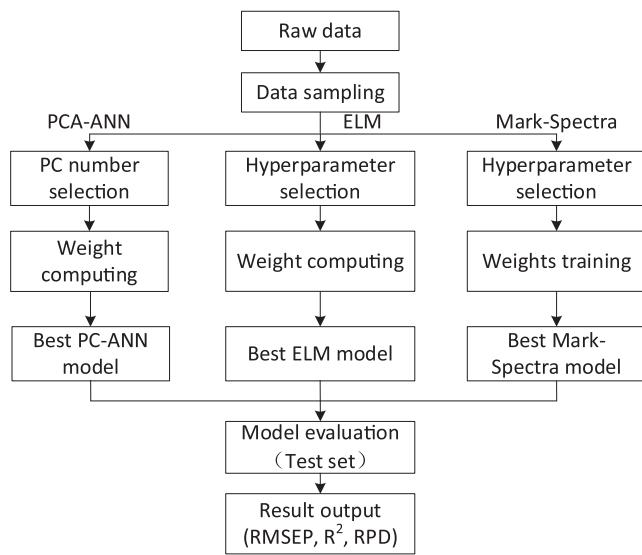
### 2.3. Design of experiments

#### 2.3.1. Comparison of Mark-Spectra with other CNN models

Different structures will extract different features. The structure of CNN is one of the most important influencing factors for performance, and it is also one of the most extensively studied techniques for improving prediction accuracy. Therefore, there is a need to compare the performance of Mark-Spectra with the other models based on CNNs. Three convolutional neural networks are built (Fig. 6), which separately compare from depth of model, numbers of filters and width of models. The first CNN model (Model 1) is the simplest one with only one convolutional layer (Conv1) with 4 same sized filters to extract features from the raw spectral data, one flatten layer (Flatten), one fully connected layer (F1), and one output layer. The kernel size of convolutional layer and the stride are same as kernel size 1 and stride 1 in Table 1. The second CNN model (Model 2) only lacks the mark layer in contrast to

Mark-Spectra, and the numbers of filters with convolutional and pooling layers are the same. During the course of the experiment, the hyperparameters of Model 2 are the same with Mark-Spectra.

The third model (Model 3) is the most structurally complex among three models. It adopts the Inception structure (Zhang et al., 2019a,b,c) in layer Conv2, Conv3 and Conv4, which is capable of improving the predicted performance. Conv1 is the general convolutional layer (in yellow) with 8 same sized filters, its kernel size and stride size are the same as kernel size 1 and stride 1 in Table 1. Conv2 includes two  $1 \times 1$  convolutions (in dark-green) and one max-pooling (in blue, pooling size is 2), Conv3 includes two  $1 \times 1$  convolutions (in dark-green) and two general convolutions (in yellow, the size of kernel and stride are setting as kernel size 2 and stride 2 in Table 1), and Conv4 only has one general convolutional layer (in yellow, the size of kernel and stride are setting as kernel size 3 and stride 3 in Table 1). The increase of the width of the network by stacking of parallel  $1 \times 1$ ,  $m \times 1$  and  $n \times 1$  convolutions in



**Fig. 7.** Flowchart for the comparison of Mark-Spectra with conventional approaches. Mark-Spectra model as well as the PCA-ANN and ELM models are implemented on raw spectral data. The PC numbers are optimized through 2 to 50, and the hidden nodes ranged from 2 to 1000 in ELM.

layer Conv2, Conv3, and Conv4 improves the adaptability of extracting the features with different scales of raw spectra variables. One neuron, which is entered into the F1 layer, covers more features of original spectral variables by feature extraction of Model 3.

### 2.3.2. Comparison of Mark-Spectra with conventional modeling approaches

In conventional calibration approaches for spectral analysis, neural network-based algorithms have achieved tremendous success (Zheng et al., 2008), which is because of generalization and fitting. The other classical algorithms (e.g., partial least square (PLS)) always assume that the content of target element is linearly related to wavelengths. This makes the generalization ability of the model poor, and the application value is greatly reduced (Zhang et al., 2019a,b,c). By introducing activation functions, neural network-based algorithms have nonlinear fitting ability and it is crucial for network performance. Neural network-based algorithms can be classified in many ways according to feature extraction, parameters updating, and other standards. Principal component analysis – artificial neural network (PCA-ANN) and extreme learning machine (ELM) have been applied to spectral analysis with quite good results (Wang et al., 2020; Zhou et al., 2017). Thus, they are selected to compare the performance with Mark-Spectra, and the flowchart of the experimental designed is shown in Fig. 7.

During the trial, firstly, the raw spectral data is read-in and data sampling is carried out to cleanse and segment the dataset. Secondly, through each modeling process, the best model is gained from each specific method. Finally, the performance of different models is evaluated. Note that the number of principal components (PCs) is selected from 2 to 50 and it has three hidden layers (the numbers of hidden layer nodes are 64, 32, 16) in PCA-ANN model. The other hyperparameters are the same with Mark-Spectra. ReLU is selected as activation function and the number of hidden nodes ranged from 2 to 1000 in ELM.

### 2.4. Dataset description

Three datasets have been investigated in this study. All of samples in each dataset are collected from either wide geographical origin, long time horizons, various production conditions, or different spectrometers. In addition, each dataset has different size of samples and wavelengths, which allows us to investigate the performance of each prediction model more comprehensive.

The first dataset consists of 80 samples of corn measured on 3 different NIR spectrometers (M5, MP5, and MP6) (Zhang et al., 2019a,b,c). The dataset covers a range from 1100 to 2498 nm with a resolution of 2 nm (700 channels). Raw spectral data collected by M5 is selected to predict the protein content of corn. The second dataset is used to predict the protein content of wheat with 523 samples (Pedersen et al., 2002). The wheat spectra have a range from 850 to 1048 nm with a resolution of 2 nm and total 100 channels. The third dataset is soil spectroscopic data, LUCAS topsoil dataset, which is one of the largest open accessed soil datasets in the world (Orgiazzi et al., 2018; Tóth et al., 2013). The dataset consists of soil properties (clay, silt and sand content, pH, organic carbon content, CaCO<sub>3</sub>, nitrogen, etc.) and Vis-NIRS absorbance. The spectra covered a range from 400 to 2499.5 nm with a resolution of 0.5 nm. The dataset includes 19,036 samples of topsoil, but 17 samples with soil total nitrogen content = 0 g kg<sup>-1</sup> need to be removed. Ultimately, 19,019 samples are obtained for research.

The second dataset has been divided into different subsets, and the first and third datasets are not divided. In order to segment the datasets more uniformly and reasonably, these datasets are divided as described below. At first, the samples are sorted in descending order according to target elements. Since there are too few samples to train the deep learning-based approach with the dataset of corn, each dataset was divided by 8:2 in order to use the most possible samples to train the corn model. Accordingly, a set of data was extracted at regular intervals as the testing subset (20% of the raw spectral data), and the remaining 80% are used as the training subset (shown in Table 2). In addition, the statistical properties of target elements are calculated including minimum values (Min), maximum values (Max), mean values (Mean), and standard deviation values (SD).

### 2.5. Model evaluation

The results of selecting characteristic wavelengths and constructing features are evaluated by the coefficient of determination ( $R^2$ ), root mean squared error of prediction (RMSEP), and the residual prediction deviation (RPD), and their calculations are shown in Equations (6)–(9).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (7)$$

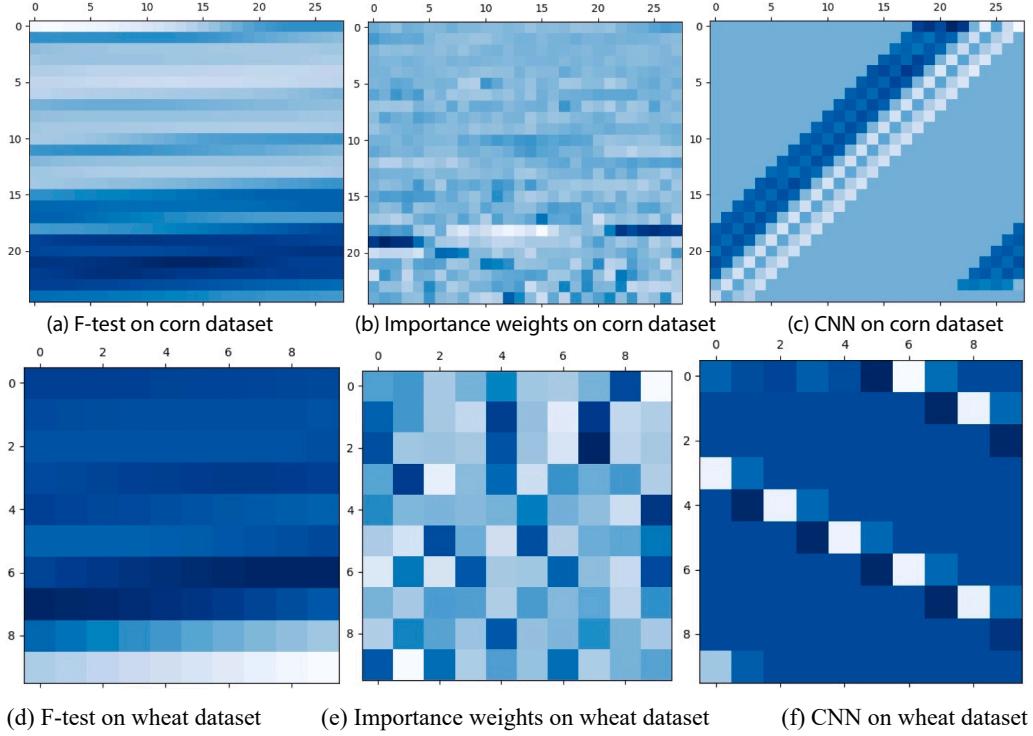
$$RPD = \frac{s_y}{RMSEP} \quad (8)$$

$$s_y = \sqrt{\frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N}} \quad (9)$$

where  $y_i$  and  $\hat{y}_i$  are measured values and predicted values respectively;  $n$  is the number of samples in the training set;  $s_y$  is the standard deviation

**Table 2**  
Description of the datasets.

Datasets	Total samples	Training samples	Testing samples	Channels	Min	Max	Mean	SD
Corn protein content (%)	80	64	16	700	7.65	9.71	8.67	0.50
Wheat protein content (%)	523	418	105	100	6.77	16.95	9.96	1.60
Soil total nitrogen (g kg <sup>-1</sup> )	19,019	15,215	3804	4200	0.2	38.6	2.93	3.76



**Fig. 8.** Distributions of characteristic wavelengths with different methods based on different datasets.

of the observed values, which is calculated using Eq. (9). In which,  $\bar{y}$  is the arithmetic mean of  $y_i$ ;  $N$  is the number of samples in the test set. Of these, RPD is used to observe and evaluate the usability of different models. If RPD is below 1.5, the model performance cannot be used for prediction due to poor performance. If it is between 1.5 and 1.8, the model needs to be improved to get better results. For values of RPD in the range of 1.8–2, the prediction model is considered to be practical. And if it is higher than 2.0, the model performance is considered to be very good (Morellos et al., 2016).

Note that, all the algorithms and models are implemented on Python platform using Pytorch and Scikit-learn library, which are run on a Windows laptop (Windows 10) with 16 GB of RAM, and an Nvidia Geforce GTX 1650 graphics card with 4 GB of RAM.

### 3. Results and discussion

#### 3.1. Distributions of characteristic wavelengths marked by different approaches

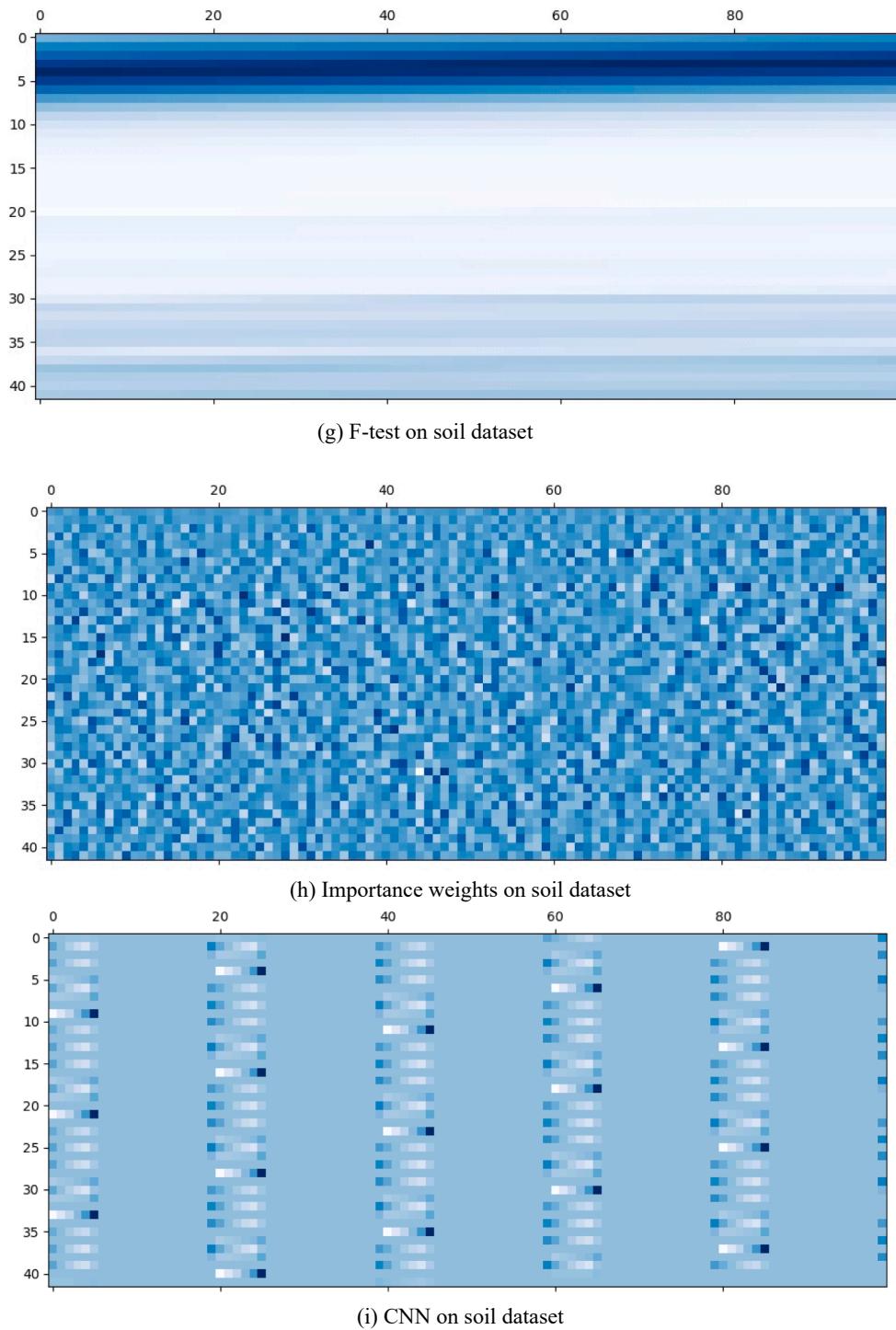
As mentioned in Section 2.1, there are many approaches to mark the characteristic wavelengths of target elements. For better illustration, the distributions of characteristic wavelengths are visualized, which are extracted by different approaches (shown in Fig. 8). These indices of outputted spectroscopic wavelengths from different screening methods are transformed to 2D image, a cell in the image represents a resolution of different dataset, and the size of image is the channels of datasets. Fig. 8.a, Fig. 8.b and Fig. 8.c are the distributions of characteristic wavelengths marked by F-test, importance weights and convolutional neural networks of the corn dataset, and the image size is  $25 \times 28$  (total 700 channels). Fig. 8.d, Fig. 8.e and Fig. 8.f are the distributions of characteristic wavelengths with wheat dataset, the size of image is  $10 \times 10$  (total 100 channels). Fig. 8.g, Fig. 8.h and Fig. 8.i are obtained according to soil dataset, the size of image is  $42 \times 100$  (total 4200 channels). The indices of wavelengths can be computed according to the indices of points in the subgraph (column-coordinate multiplied by the

max scale value of row-coordinate, and plus the row-coordinate in the last row). These weights of wavelengths are plotted as dark pixels, darker colors meaning larger weights. For example, the index of the cell in image is  $15 \times 10$  in subgraph a, it means the 385th wavelengths in corn dataset, which color shows the weights of full wavelengths.

We can see that the distributions of characteristic wavelengths with convolutional neural network are discrete with different range of wavelengths (in size of convolutional kernel at the first convolutional layer), especially in corn and soil datasets which have more spectral variables in the raw data. By the principle of convolutional neural network, one neuron will cover many original spectral wavelengths by convolutional operations (Fig. 3.a), and max-pooling will extract these neurons for the next layer. This suggests that the characteristic wavelengths marked by convolutional neural network are rough (not in units of one wavelength), and it cannot overcome the collinearity of with adjacent wavelengths. To improve the performance of models, it needs more granularly screening process before convolutional operations.

The distributions of the characteristic wavelengths show strip-shaped with different datasets, which are extracted based on F-test. With the previous theoretical analysis, F-test use the correlation of feature with the target variable alone as the measure of relevance, and there is possibility of presence of intercorrelation within the extracted features (Ng et al., 2019). Strip-shaped distributions of the characteristic wavelengths illustrate the correlation of adjacent wavelengths with little difference, and characteristic wavelengths will be selected from a certain range of bands. Therefore, it can also not overcome the collinearity of raw spectral data and not fit to mark features for convolutional neural network. Marking the characteristic wavelengths by importance weights obtains a discrete distribution with different datasets. According to the process of picking characteristic wavelengths, the intercorrelation within the wavelengths is reduced with updating the subsets, and it will provide granular data for convolutional neural network. Therefore, importance weights are selected to mark features in Mark layer with Mark-Spectra.

In addition, it should be noted that the locations of the characteristic

**Fig. 8. (continued).**

wavelengths are not similar when using different approach to select characteristic wavelengths, and the trend is not identical for the results obtained from the F-test, importance weights, and the CNN approach. The reason for the different characteristic wavelength locations is because different method uses different matrix to evaluate the contribution of different wavelengths. As mentioned in Section 2.1, F-test focuses only on the F-values of different wavelengths. Importance weight selects characteristic wavelengths according to prediction results, and it ignores the process of feature extraction prior to modelling. As for CNN approach, it can capture more features (e.g., higher order derivatives) by itself, and this is one reason why it is widely used in

spectral analysis.

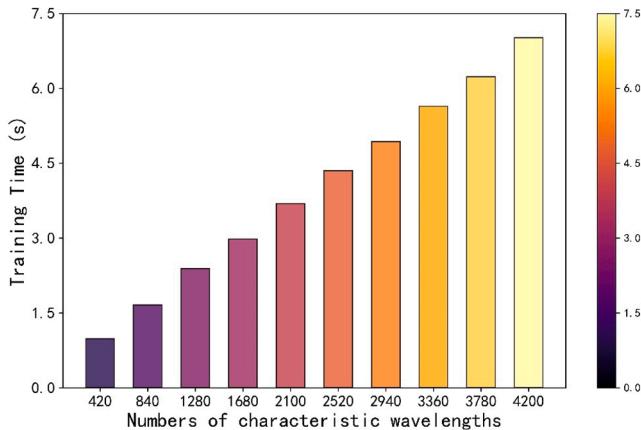
### 3.2. Performance of Mark-Spectra for quantitative spectral analysis with different numbers of features

Mark-Spectra can select different wavelengths to predict the target elements by Mark layer, and different number of wavelengths will lead to different performance. In order to analyze the effect of different set of characteristic wavelengths, this experiment is carried out to compare the performance of Mark-Spectra with different sizes of characteristic wavelengths in the following. The size of characteristic wavelengths

**Table 3**

Predictive results obtained by Mark-Spectra on raw data for three datasets.

Corn (%)				Wheat (%)				Soil ( $\text{g kg}^{-1}$ )			
Features	R <sup>2</sup>	RMSEP	RPD	Features	R <sup>2</sup>	RMSEP	RPD	Features	R <sup>2</sup>	RMSEP	RPD
70	0.88	0.03	16.39	10	—	—	—	420	0.88	1.65	2.26
140	0.83	0.04	11.79	20	0.53	1.19	1.34	840	0.89	1.55	2.41
210	0.84	0.04	12.40	30	0.76	0.61	2.62	1280	0.88	1.65	2.27
280	0.89	0.03	18.16	40	0.58	1.06	1.50	1680	0.88	1.75	2.14
350	0.84	0.04	12.29	50	0.66	0.86	1.86	2100	0.89	1.61	2.33
420	0.87	0.03	15.49	60	0.84	0.39	4.04	2520	0.89	1.59	2.35
490	0.89	0.03	17.52	70	0.72	0.7	2.28	2940	0.88	1.71	2.20
560	0.85	0.04	13.02	80	0.73	0.68	2.33	3360	0.88	1.74	2.15
630	0.75	0.06	8.05	90	0.76	0.6	2.66	3780	0.88	1.66	2.25
700	0.71	0.07	7.04	100	0.86	0.37	4.35	4200	0.88	1.66	2.26

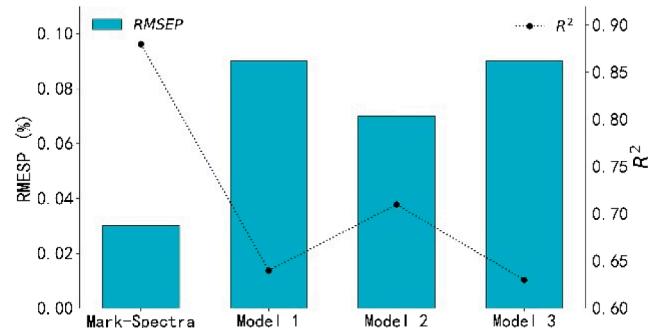
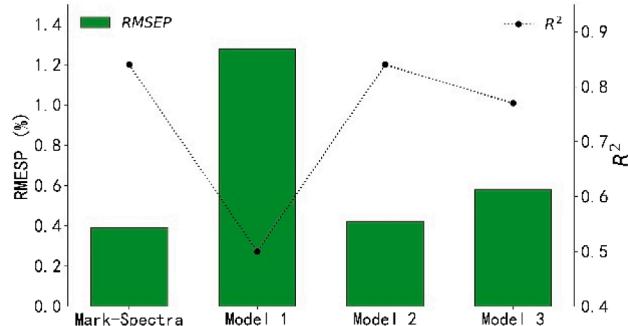
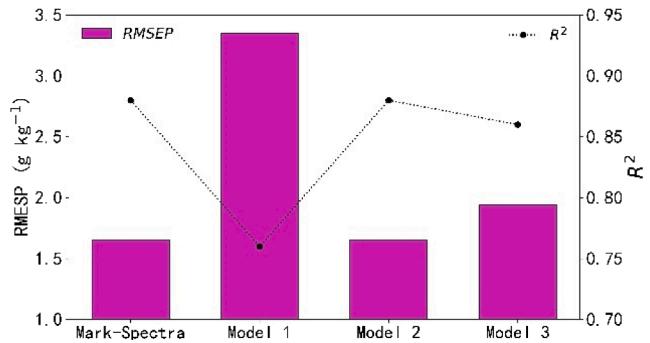
**Fig. 9.** Training time of Mark-Spectra on Soil dataset.

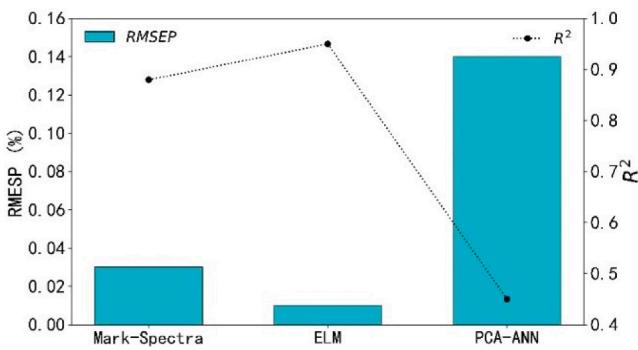
increases from 10% to 90% with a step of 10% on three datasets, and these characteristic wavelengths are used for modeling and the results are shown in **Table 3**.

For corn protein analysis, the best performance ( $R^2 = 0.89$ , RMSEP = 0.03) is obtained with 280 and 490 feature wavelengths and it does not differ much from 70 wavelengths ( $R^2 = 0.88$ , RMSEP = 0.03). The others show relatively poor performances, especially full spectra data. For the dataset of wheat, full data obtains the best performance, the  $R^2$  is 0.86 and RMSEP is 0.37. Note that there are no results given with 10 features due to having too few features. For the dataset of soil, the results do not differ substantially with different characteristic wavelengths of different elements.

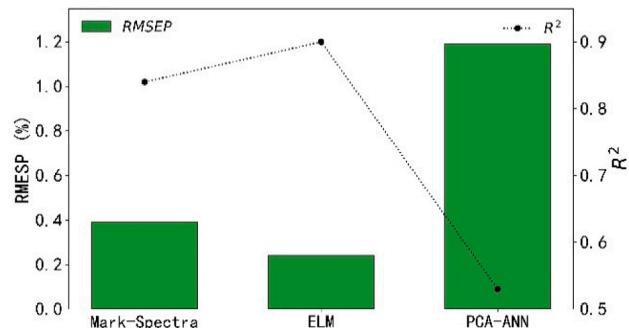
According to the dataset description (**Table 2**), the dataset of corn has the minimum sample size and the second-highest number of features among the three datasets. More samples should be used for relatively large number of features in model training, but the dataset of corn cannot meet this condition. It causes underfitting of the CNN model, so full spectra data in turn leads to the worst results. Mark-Spectra can select the characteristic wavelengths of protein, which reduced the number of features, and improved the performance of CNN model. The dataset of wheat has the minimum feature size and 523 samples, in these circumstances reducing the number of features only decreases the accuracy of predicting model. Therefore, Mark-Spectra is not efficiently effective for datasets with small number of features. According to adequate features and samples, the datasets of soil can obtain stable results with different size of characteristic wavelengths. According to the results with different datasets, the optimal values of predictive accuracy don't find in the case of full channels (features). This also illustrates that moderate or high collinearities exist among the independent spectral variables.

In addition, we paid attention to the singular points of  $R^2$  and RMSEP in datasets of corn and wheat. The local optima of prediction are

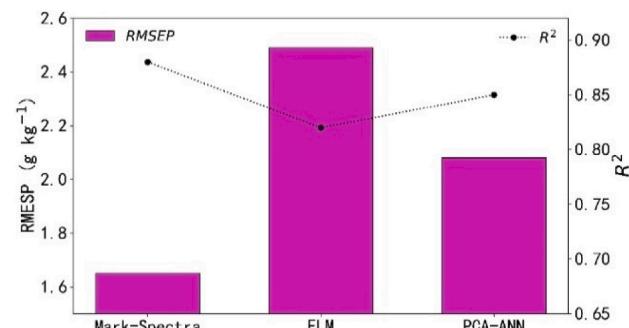
**a. The dataset of corn****b. The dataset of wheat****c. The dataset of soil****Fig. 10.** Predictive results obtained by four convolutional neural networks on three datasets.



a. The dataset of corn



b. The dataset of wheat



c. The dataset of soil

**Fig. 11.** Predictive results obtained by Mark-Spectra and two conventional approaches on three datasets.

obtained with 70, 280, 490 samples on corn dataset and 30, 60, 90 samples on wheat dataset. Based on our analysis, this is caused by the spatial distribution of features. Redundant information increases with increasing characteristic wavelengths, although it is mixed with useful information. The approaches based on CNN maintain the underlying spatial relationships, and it can't remove the redundant information assessed in each wavelength (Fig. 3.a, Fig. 8). As the feature space

increases, CNN is able to remove the redundant information according to the convolutional kernel, and it will regain high prediction accuracy.

Due to mark the characteristic wavelengths of target elements, Mark-Spectra model will save training time on the premise of ensuring the accuracy. To manifest this, the average of training time per 100 epochs with different number of characteristic wavelengths is counted with the dataset of soil, and the results are shown in Fig. 9. During the test, only the number of characteristic wavelengths is changed, and the other parameters with possible impact to training time (e.g., sample size, batch size and others) are the same. The minimum of training time is 0.98 s epoch<sup>-1</sup> with 420 wavelengths, and the max is 7.01 s epoch<sup>-1</sup> with full data. Note that they are not linearly related between training time and number of characteristic wavelengths.

### 3.3. Comparisons between Mark-Spectra and other approaches

#### 3.3.1. The comparisons between Mark-Spectra and three CNN models

Mark-Spectra model with partial features (10% in the dataset of corn and soil, 60% in the dataset of wheat) of raw data is compared with three CNN models (with full sample), and the results are shown in Fig. 10. Bars are presented as RMSEP of different models, and lines indicate R<sup>2</sup> of four models. For corn protein analysis, Mark-Spectra outperforms the other three CNNs in terms of a lower RMSEP (0.03), higher R<sup>2</sup> (0.88) and stable effects, which RPD is 16.39. The worst results are obtained by Model 3 (R<sup>2</sup> is 0.63, RMSEP is 0.09, and RPD is 5.49), which is the most structurally complex of the four models. This illustrates that the model with too many parameters needs much more samples to train, which are introduced by model complexity. For wheat protein analysis, Mark-Spectra obtains the best results (R<sup>2</sup> is 0.84, RMSEP is 0.39 and RPD is 4.04) and Model 2 obtained the second-best performance (R<sup>2</sup> is 0.84, RMSEP is 0.42 and RPD is 3.82), which only lacks the mark layer in contrast to Mark-Spectra. According to the results in Section 3.2, Mark-Spectra has potential to reach better performance by increasing the number of characteristic wavelengths. For soil total nitrogen analysis, the Mark-Spectra model provides a highest R<sup>2</sup> of 0.88 and a lowest RMSEP of 1.65 g kg<sup>-1</sup> among four calibration approaches, which is not significantly different from Model 2 (R<sup>2</sup> is 0.88 and RMSEP is 1.65 g kg<sup>-1</sup>) and Model 3 (R<sup>2</sup> is 0.86 and RMSEP is 1.94 g kg<sup>-1</sup>).

According to the performance of Model 1, Model 2 and Model 3, we can know that structurally complex models are able to capture more features, but it also needs more samples to train the parameters introduced by more convolutional and pooling layers. With same structure of convolutional and pooling layers (e.g., Mark-Spectra and Model 2), Mark-Spectra can achieve good results with small number of features according to Mark layer, which is used to overcome the collinearities among raw spectral data. Mark-Spectra combines convolutional operations and importance weights to extract the characteristic wavelengths of target elements. On the one hand, it overcomes the spatial relationships among raw spectral data, improves the performance of modeling and saves training time. On the other hand, it allows flexible controllability of characteristic wavelengths for convolutional operations according to raw spectral data properties, which does not lead to reduced accuracy.

#### 3.3.2. The comparisons between Mark-Spectra and conventional methods

Mark-Spectra model based on partial features (10% in the dataset of corn and soil, 60% in the dataset of wheat) of raw data is compared with

**Table 4**

Predictive results obtained by Mark-Spectra and three conventional approaches on three datasets.

Models	Corn (%)			Wheat (%)			Soil (g kg <sup>-1</sup> )		
	R <sup>2</sup>	RMSEP	RPD	R <sup>2</sup>	RMSEP	RPD	R <sup>2</sup>	RMSEP	RPD
Mark-Spectra	0.88	0.03	16.39	0.84	0.39	4.04	0.88	1.65	2.26
ELM	0.95	0.01	44.25	0.90	0.25	6.43	0.82	2.49	1.50
PCA-ANN	0.45	0.14	3.63	0.53	1.19	1.34	0.85	2.08	1.80

**Table 5**Description of subsets with soil nitrogen content ( $\text{g kg}^{-1}$ ).

Dataset No.	Total samples	Training samples	Test samples	Min	Max	Mean	SD
1	1903	1522	381	0.2	30.2	2.81	3.73
2	3801	3040	761	0.2	34.4	2.87	3.71
3	7609	6087	1522	0.2	34.2	2.96	3.74
4	9508	7606	1902	0.2	38.6	2.92	3.70
5	13,314	10,651	2663	0.2	38.6	2.93	3.74

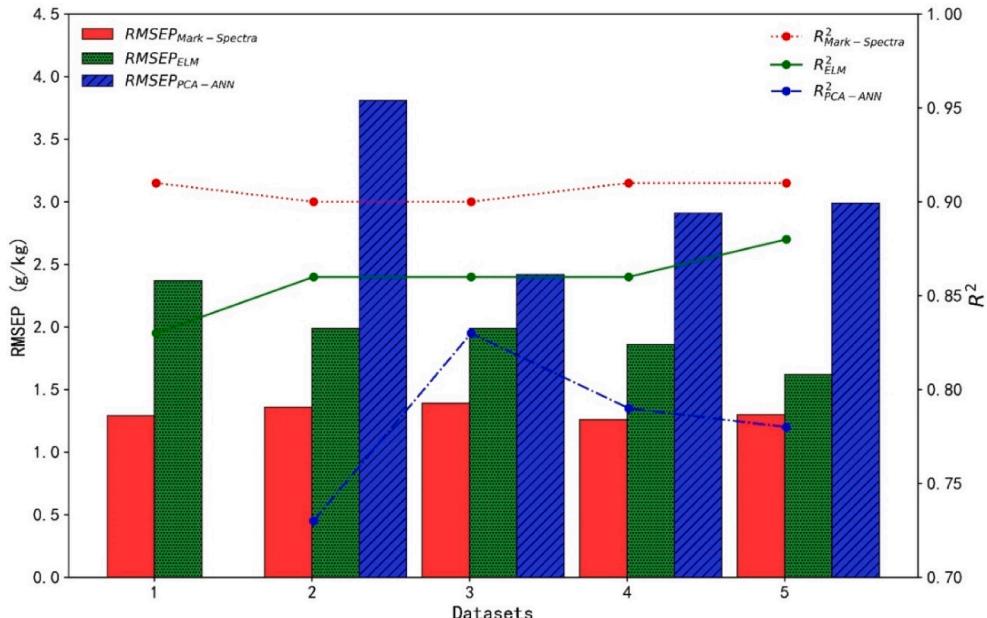
PCA-ANN and ELM, and the results are shown in Fig. 11 and Table 4. Different datasets, however, would have their preferred prediction methods. The ELM method provides the highest  $R^2$  and the lowest RMSEP for corn and wheat datasets. Mark-Spectra model is best suited for soil total nitrogen, respectively. Note that the number of samples is smaller than the number of features on corn dataset, and PCA is unsuitable for this condition. Therefore, the range of PCs on corn dataset is 2 to 16.

As stated earlier, the characteristics of raw data determines the effect of prediction models. Corn and wheat datasets have smaller standard deviation (Table 2), this implies that different samples have small gap so that they are easy to fit. ELM performs better than other methods on the datasets of corn and wheat, and each of them has a small number of

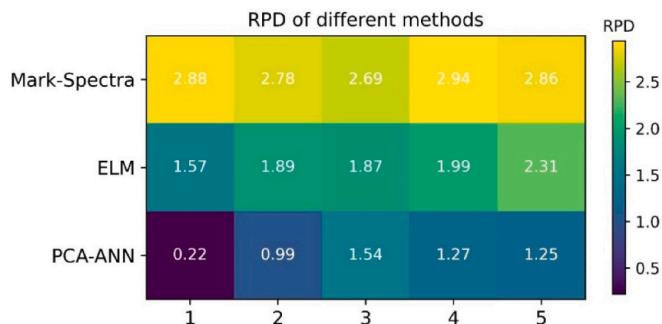
samples to train. Since there is only single layer feedforward neuron network in ELM, this makes it very suitable for training with small amount of data. For datasets with large sample size (e.g., dataset of soil), Mark-Spectra and PCA-ANN can be trained with sufficient samples. This makes Mark-Spectra and PCA-ANN obtain better results than ELM. PCA-ANN is not free from limitations of feature extraction (mainly caused by principal component analysis) and number of samples, and it still has room for improvement. Although Mark-Spectra still needs a certain number of samples and features to train, but it has improved the performance than traditional CNN models (Section 3.3.1). Meanwhile, Mark-Spectra retains better generalization with different datasets.

For further illustration the advantage of Mark-Spectra, five subsets are selected from the dataset of soil ordered by content of soil nitrogen and details are shown in Table 5. These subsets with different numbers of samples, and all standard deviations values are greater than 3.70 in five subsets. Mark-Spectra model based on 10% of raw features (420 characteristic wavelengths) is compared with PCA-ANN and ELM, and results are shown in Fig. 12. The exceptional is that the results of PCA-ANN with the first subset are not shown due to the poor outcome ( $R^2 = -0.23$ , RMSEP is  $16.99 \text{ g kg}^{-1}$ , RPD is 0.22).

Mark-Spectra performs best among these three models in terms of accuracy and stability with different subsets of soil dataset. The reason for this result is the ability of generalization. ELM can obtain good performance with the dataset of corn and wheat, but it cannot fit



a. Accuracy of different predicted models with different subsets



b. Credibility of different predicted models with different subsets

Fig. 12. Performance of different predicted models with different subsets of soil dataset.

datasets with high dispersion degree (e.g., subsets of soil datasets), which happens a lot in agricultural production. PCA-ANN is limited by the feature extraction and sample size. Although PCA is the classical algorithm for dimension reduction, it neglects the interpretability of dimensions and reduces precision of prediction (Engel et al., 2013). As for ANN, it needs more samples to train for updating parameters. Comprehensively, PCA-ANN is not suitable for those datasets. Mark-Spectra inherits the ability of feature extraction from the traditional convolutional operations, and max-pooling operations can remove redundant information. This makes good results of Mark-Spectra.

#### 4. Conclusions

Spectroscopic technique is one of the most important methods for detecting target element content in the field of agriculture. However, the large number of spectral wavelengths contributes to collinearity and redundancies rather than relevant information. This paper proposed a novel convolutional neural network, named Mark-Spectra, for quantitative spectral analysis overcoming spatial correlations among adjacent wavelengths. The following conclusions are accordingly achieved.

- (1) Traditional convolutional neural network retains the spatial relationships among adjacent wavelengths with the operations of convolution and pooling, and it contributes to collinearity and redundancies rather than relevant effective information. Mark layer has been introduced as part of Mark-Spectra, which could extract characteristic wavelengths and obtain the indices of discrete wavelengths according to importance weights. Experimentally, Mark-Spectra has better performance than other traditional CNN models.
- (2) Mark-Spectra introduces characteristic wavelength labeling method into deep learning-based spectral analysis. Typically, in order to obtain good performance, complex models are built to capture more features and large number of samples are required to train those models. Mark-Spectra can reduce the dependence on sample size and overcome spatial relationships by characteristic wavelength labeling method.
- (3) Mark-Spectra is still limited by the characteristics of raw spectral data (e.g., the number of samples and features), but it performed better than the other traditional CNN models. Mark-Spectra can be applied to more scenarios owing to its generalization and fitting, since it reduces the need from prior knowledge and human effort on quantitative spectral analysis techniques.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgement

This research was funded by National Key Research and Development Program (grant number 2019YFE0125500 and 2017YFD0201501-02) and Chinese Universities Scientific Fund (Grant No. 2021TC111).

#### References

- Abulaiti, Y., Sawut, M., Maimaitiali, B., Chunyue, M., 2020. A possible fractional order derivative and optimized spectral indices for assessing total nitrogen content in cotton. *Comput. Electron. Agric.* 171, 105275. <https://doi.org/10.1016/j.compag.2020.105275>.
- Aikawa, K., 1992. Phoneme recognition using time-warping neural networks. *J. Acoust. Soc. Japan* 13, 395–402. <https://doi.org/10.1250/ast.13.395>.
- An, X., Li, M., Zheng, L., 2012. Estimation of soil total nitrogen and soil moisture based on NIRS technology. *IFIP Adv. Inf. Commun. Technol.* 369 AICT, 639–647. [https://doi.org/10.1007/978-3-642-27278-3\\_66](https://doi.org/10.1007/978-3-642-27278-3_66).
- An, X., Li, M., Zheng, L., Sun, H., 2015. Eliminating the interference of soil moisture and particle size on predicting soil total nitrogen content using a NIRS-based portable detector. *Comput. Electron. Agric.* 112, 47–53. <https://doi.org/10.1016/j.compag.2014.11.003>.
- Binetti, G., Coco, L. Del, Ragone, R., Zelasco, S., Perri, E., Montemurro, C., Valentini, R., Naso, D., Fanizzi, F.P., Schena, F.P., 2016. Cultivar classification of Apulian olive oils: use of artificial neural networks for comparing NMR, NIR and merceological data. *Food Chem.* <https://doi.org/10.1016/j.foodchem.2016.09.041>.
- Bjerrum, E.J., Glaisher, M., Skov, T., 2017. Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics. *arXiv* 1–10.
- Cang, Z., Wei, G.W., 2017. TopologyNet: Topology-based deep convolutional neural networks for biomolecular property predictions. *arXiv* 1–27.
- Engel, J., Gerretzen, J., Szymańska, E., Jansen, J.J., Downey, G., Blanchet, L., Buydens, L. M.C., 2013. Breaking with trends in pre-processing? *TrAC Trends. Anal. Chem.* 50, 96–106. <https://doi.org/10.1016/j.trac.2013.04.015>.
- Fernández-Ugalde, O., Ballabio, C., Lugato, E., Scarpa, S., Jones, A., 2020. Assessment of changes in topsoil properties in LUCAS samples between 2009/2012 and 2015 surveys. <https://doi.org/10.2760/5503>.
- Fukuhara, M., Fujiwara, K., Maruyama, Y., Itoh, H., 2019. Feature visualization of Raman spectrum analysis with deep convolutional neural network. *Anal. Chim. Acta* 1087, 11–19. <https://doi.org/10.1016/j.aca.2019.08.064>.
- Gerretzen, J., Szymańska, E., Bart, J., Davies, A.N., van Manen, H.J., van den Heuvel, E. R., Jansen, J.J., Buydens, L.M.C., 2016. Boosting model performance and interpretation by entangling preprocessing selection and variable selection. *Anal. Chim. Acta* 938, 44–52. <https://doi.org/10.1016/j.aca.2016.08.022>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2016-Decem, 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- Hetrick, E.M., Shi, Z., Barnes, L.E., Garrett, A.W., Rupard, R.G., Kramer, T.T., Cooper, T. M., Myers, D.P., Castle, B.C., 2017. Development of Near Infrared Spectroscopy-based Process Monitoring Methodology for Pharmaceutical Continuous Manufacturing Using an Offline Calibration Approach. *Anal. Chem.* 89, 9175–9183. <https://doi.org/10.1021/acs.analchem.7b01907>.
- Jansen, J.J., Downey, G., Blanchet, L., Engel, J., Gerretzen, J., Szymańska, E., 2013. Trends in Analytical chemistry Breaking with trends in pre-processing ? 50, 96–106. <https://doi.org/10.1016/j.trac.2013.04.015>.
- Jia, S., Li, H., Wang, Y., Tong, R., Li, Q., 2017. Hyperspectral imaging analysis for the classification of soil types and the determination of soil total nitrogen. *Sensors (Switzerland)* 17, 1–14. <https://doi.org/10.3390/s17102252>.
- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Li, F.F., 2014. Large-scale video classification with convolutional neural networks. In: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 1725–1732. <https://doi.org/10.1109/CVPR.2014.223>.
- Li, P., Zhang, X., Li, S., Du, G., Jiang, L., Liu, X., Ding, S., Shan, Y., 2020. A rapid and nondestructive approach for the classification of different-age citri reticulatae pericarpium using portable near infrared spectroscopy. *Sensors (Switzerland)* 20. <https://doi.org/10.3390/s20061586>.
- Malek, S., Melgani, F., Bazi, Y., 2018. One-dimensional convolutional neural networks for spectroscopic signal regression. *J. Chemom.* 32, 1–17. <https://doi.org/10.1002/cem.2977>.
- Minasny, B., McBratney, A.B., Bellon-Maurel, V., Roger, J.M., Gobrecht, A., Ferrand, L., Joalland, S., 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma* 167–168, 118–124. <https://doi.org/10.1016/j.geoderma.2011.09.008>.
- Morellos, A., Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., Mouazen, A.M., 2016. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosyst. Eng.* 152, 104–116. <https://doi.org/10.1016/j.biosystemseng.2016.04.018>.
- Ng, W., Minasny, B., Malone, B.P., Sarathjith, M.C., Das, B.S., 2019. Optimizing wavelength selection by using informative vectors for parsimonious infrared spectra modelling. *Comput. Electron. Agric.* 158, 201–210. <https://doi.org/10.1016/j.compag.2019.02.003>.
- Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS Soil, the largest expandable soil dataset for Europe: a review. *Eur. J. Soil Sci.* 69, 140–153. <https://doi.org/10.1111/ejss.12499>.
- Palou, A., Miró, A., Blanco, M., Larraz, R., Gómez, J.F., Martínez, T., González, J.M., Alcalá, M., 2017. Calibration sets selection strategy for the construction of robust PLS models for prediction of biodiesel/diesel blends physico-chemical properties using NIR spectroscopy. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 180, 119–126. <https://doi.org/10.1016/j.saa.2017.03.008>.
- Pedersen, D.K., Martens, H., Nielsen, J.P., Engelsen, S.B., 2002. Near-infrared absorption and scattering separated by extended inverted signal correction (EISC): Analysis of near-infrared transmittance spectra of single wheat seeds. *Appl. Spectrosc.* 56, 1206–1214. <https://doi.org/10.1366/000370202760295467>.
- Riese, F.M., Keller, S., 2019. Soil texture classification with 1D convolutional neural networks based on hyperspectral data. *ISPRS Ann. Photogram. Remote Sens. Spatial Inform. Sci.* 615–621. <https://doi.org/10.5194/isprs-annals-IV-2-W5-615-2019>.
- Shi, W., Caballero, J., Theis, L., Huszar, F., Aitken, A., Ledig, C., Wang, Z., 2016. Is the deconvolution layer the same as a convolutional layer?
- Tóth, G., Jones, A., Montanarella, L., 2013. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. *Environ. Monit. Assess.* <https://doi.org/10.1007/s10661-013-3109-3>.

- Wang, Y., Li, M., Ji, R., Wang, M., Zhang, Y., Zheng, L., 2021a. Construction of complex features for predicting soil total nitrogen content based on convolution operations. *Soil Tillage Res.* 213, 105109. <https://doi.org/10.1016/j.still.2021.105109>.
- Wang, Y., Li, M., Ji, R., Wang, M., Zheng, L., 2021b. A deep learning-based method for screening soil total nitrogen characteristic wavelengths. *Comput. Electron. Agric.* 187, 106228. <https://doi.org/10.1016/j.compag.2021.106228>.
- Wang, Y., Li, M., Ji, R., Wang, M., Zheng, L., 2020. Comparison of soil total nitrogen content prediction models based on Vis-NIR spectroscopy. *Sensors (Switzerland)* 20, 1–20. <https://doi.org/10.3390/s20247078>.
- Weng, W., Zhu, X., 2021. INet: Convolutional Networks for Biomedical Image Segmentation. *IEEE Access* 9, 16591–16603. <https://doi.org/10.1109/ACCESS.2021.3053408>.
- Yang, J., Xu, J., Zhang, X., Wu, C., Lin, T., Ying, Y., 2019. Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Anal. Chim. Acta* 1081, 6–17. <https://doi.org/10.1016/j.aca.2019.06.012>.
- Zhang, X., Lin, T., Xu, J., Luo, X., Ying, Y., 2019a. DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis. *Anal. Chim. Acta* 1058, 48–57. <https://doi.org/10.1016/j.aca.2019.01.002>.
- Zhang, X., Xu, J., Yang, J., Chen, L., Zhou, H., Liu, X., Li, H., Lin, T., Ying, Y., 2020. Understanding the learning mechanism of convolutional neural networks in spectral analysis. *Anal. Chim. Acta* 1119, 41–51. <https://doi.org/10.1016/j.aca.2020.03.055>.
- Zhang, Y., Li, M., Zheng, L., Qin, Q., Lee, W.S., 2019b. Spectral features extraction for estimation of soil total nitrogen content based on modified ant colony optimization algorithm. *Geoderma* 333, 23–34. <https://doi.org/10.1016/j.geoderma.2018.07.004>.
- Zhang, Y., Li, M.Z., Zheng, L.H., Zhao, Y., Pei, X., 2016. Soil nitrogen content forecasting based on real-time NIR spectroscopy. *Comput. Electron. Agric.* 124, 29–36. <https://doi.org/10.1016/j.compag.2016.03.016>.
- Zhang, Y., Sui, B., Shen, H., Ouyang, L., 2019c. Mapping stocks of soil total nitrogen using remote sensing data : A comparison of random forest models with different predictors. *Comput. Electron. Agric.* 160, 23–30. <https://doi.org/10.1016/j.compag.2019.03.015>.
- Zheng, L.H., Li, M.Z., Pan, L., Sun, J.Y., Tang, N., 2008. Estimation of soil organic matter and soil total nitrogen based on NIR spectroscopy and BP neural network. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy Spectr. Anal.* 28, 1160–1164. <https://doi.org/10.3964/j.issn.1000-0593.2008.05.020>.
- Zhou, P., Yang, W., Li, M., Zheng, L., Chen, Y., 2017. Soil Total Nitrogen Content Prediction Based on Gray Correlation-extreme Learning Machine. *Nongye Jixie Xuebao/Trans. Chin. Soc. Agric. Mach.* 48, 271–276. <https://doi.org/10.6041/j.issn.1000-1298.2017.S0.041>.
- Zhou, P., Zhang, Y., Yang, W., Li, M., Liu, Z., Liu, X., 2019. Development and performance test of an in-situ soil total nitrogen-soil moisture detector based on near-infrared spectroscopy. *Comput. Electron. Agric.* 160, 51–58. <https://doi.org/10.1016/j.compag.2019.03.016>.