JUSTC

Article

http://justc.ustc.edu.cn

Received: June 28, 2024; Accepted: November 04, 2024

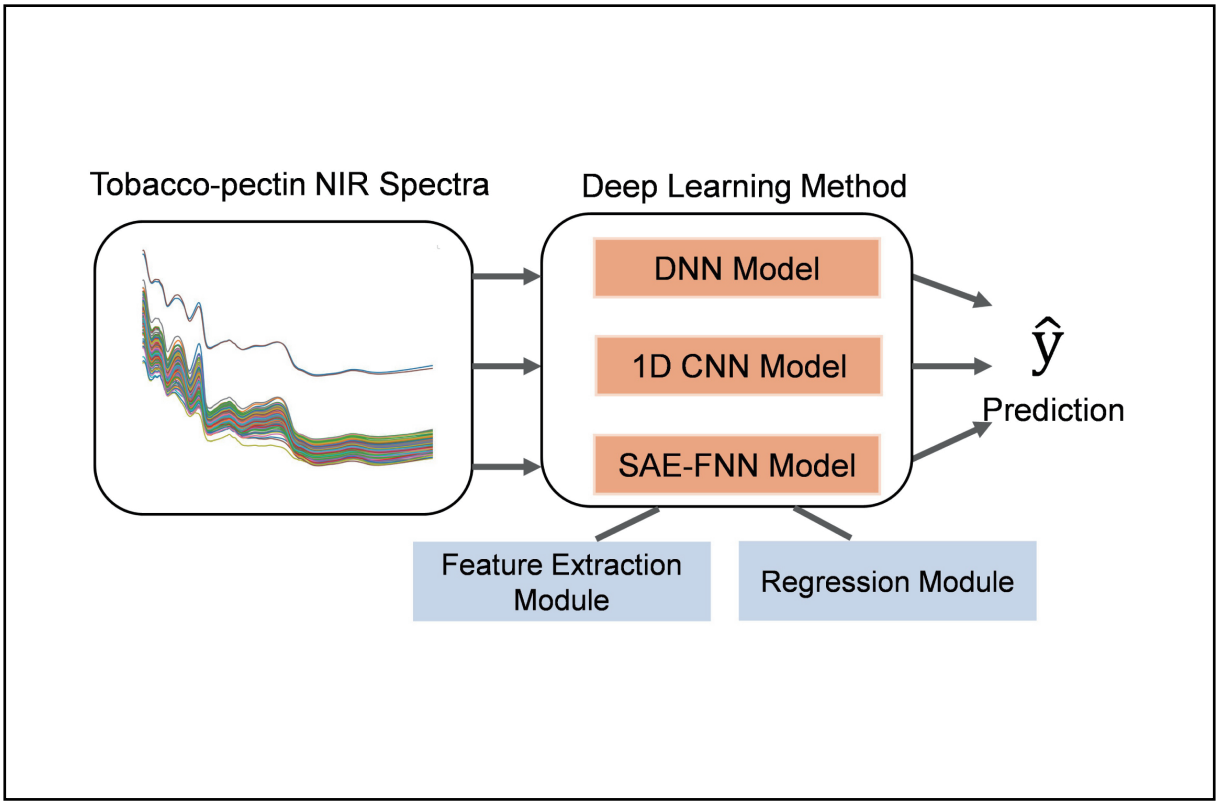
Using deep learning to reduce nonlinearity effects in near-infrared spectroscopy for accurate quantification of tobacco leaf pectin concentrations

Wenhui Yang, and Limin Shao ✉

Department of Chemistry, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Limin Shao, E-mail: lshao@ustc.edu.cn

Graphical abstract



Using three deep learning models to analyze NIR data of tobacco-pectin, aiming to extract features for predictions of pectin concentrations.

Public summary

- Applying deep learning (DL) to address the complex nonlinearity in NIR data has overcome the limitations of traditional linear modeling methods like partial least squares (PLS), offering fresh approaches to analytical problems.
- Using DL techniques for in-depth quantitative analysis of key factors like pectin concentration in tobacco leaf enhances our understanding of tobacco and provides references for future analyses of similar complex systems.
- Our research serves as another example for the effectiveness of DL models to yield accurate results in applications.

Citation: Yang W H, Shao L M. Using deep learning to reduce nonlinearity effects in near-infrared spectroscopy for accurate quantification of tobacco leaf pectin concentrations. *JUSTC*, **2025**, 55(6): 0607. CSTR: 32290.14.JUSTC-2024-0095. DOI: 10.52396/JUSTC-2024-0095

JUSTC

Article

http://justc.ustc.edu.cn

Received: June 28, 2024; Accepted: November 04, 2024

Using deep learning to reduce nonlinearity effects in near-infrared spectroscopy for accurate quantification of tobacco leaf pectin concentrations

Wenhui Yang, and Limin Shao ✉

Department of Chemistry, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Limin Shao, E-mail: lshao@ustc.edu.cn

Cite This: *JUSTC*, **2025**, 55(6): 0607 (10pp)

Read Online

Abstract: In the near-infrared (NIR) spectroscopic data of complex sample systems, such as tobacco leaves, nonlinearity is fairly significant between the absorbance and concentration. This nonlinearity severely degrades the quantitative results of traditional methods, such as partial least squares regression (PLS), which can be used to construct linear models. The problem was addressed in this study by using deep learning (DL). We employed three different DL models: a one-dimensional convolutional neural network (1D CNN), a deep neural network (DNN), and a stacked autoencoder with feedforward neural networks (SAE-FNNs). By carefully selecting and tuning the architectures and parameters of these models, we were able to find the most suitable model for dealing with such nonlinear relationships. Our experimental findings reveal that both the DNN and the SAE-FNN models excel in addressing the nonlinear issues of pectin concentration in tobacco, surpassing the performance of the classic linear model (PLS). Specifically, the DNN model stands out for its low average root mean squared error of prediction (RMSEP) value and small standard deviation (SD) of RMSEPs, leading to a tighter and more centered distribution of residuals in the prediction set. These DL models not only proficiently identify complex patterns within NIR data but also boast high prediction accuracy and fast implementation, demonstrating their effectiveness in analytical applications.

Keywords: quantitative regression; nonlinearity; deep learning methods; near-infrared spectroscopy

CLC number: O657.33 **Document code:** A

1 Introduction

Near-infrared (NIR) spectroscopy features fast measurement, high efficiency and low cost[1,2]. Especially when combined with chemometric analysis, NIR spectroscopy plays a significant role in the quantitative analysis of complex sample systems in many fields, such as tobacco, petroleum, pharmaceuticals, and food[3–5]. For quantitative modeling of NIR data, classical chemometric methods, which are mostly based on principal component analysis (PCA) and partial least squares (PLS) methods in conjunction with spectral preprocessing, have been widely adopted with positive results[6,7]. However, the classical chemometric methods still have certain limitations, including the inability to fully capture the complex nonlinearities inherent in datasets, potential overfitting issues, reliance on preprocessing techniques, limited scalability, and reduced adaptability. Despite preprocessing advancements, they struggle to fully simulate nonlinearities in complex samples.

The Lambert–Beer law serves as the foundation for quantitative NIR spectroscopy, offering a straightforward relationship between the substance concentration and light absorption intensity. However, real-world conditions often deviate from the idealized assumption of this law[8]. Factors such as sample heterogeneity[9], nonlinear absorption characteristics[10], stray light[11], and variations in temperature and pressure[12] can introduce nonlinearities into NIR spectroscopy. Indeed, nonlinearity is a “fact of life” in NIR analysis, and nonlinearity between the sample spectrum and the reference value is inevitable[13]. Traditional linear methods, e.g., PLS, may not be able to adequately capture the complex underlying nonlinear effects in data, leading to unsatisfactory prediction results. Therefore, developing new methods that can address these nonlinearities is highly practical.

Recently, deep learning (DL) has emerged as a powerful approach for addressing nonlinearities in quantitative analysis, offering both modeling and feature extraction capabilities. Currently, there are two DL approaches for spectral predictive modeling. One is the supervised approach using convolutional neural networks (CNNs)[14,15] and deep neural networks (DNNs)[16,17]. The other approach involves extracting deep spectral features in an unsupervised manner, which are subsequently used to predict target chemical concentrations in a supervised manner. A typical method of this approach is the use of stacked autoencoders with feedforward neural networks (SAE-FNNs)[18,19]. These nonlinear methods are not only capable of more accurately capturing complex data relationships but also minimize the need for manual feature

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607

selection[20], thus offering a more objective and efficient framework for data analysis.

In this study, the 1D CNN, which is appropriate for one-dimensional spectral analysis, the DNN, and the SAE-FNN were employed for the quantitative analysis of complex sample systems combined with spectroscopy. The DL methods were compared with the PLS regression model, which was preprocessed via multiplicative scatter correction (MSC) and standard normal variate (SNV) techniques. The preprocessing steps MSC and SNV were applied to the PLS model to mitigate issues related to baseline shifts and scattering effects in the NIR data, ensuring a fair comparison of model performance in handling nonlinearity and improving the robustness of the analysis. We have adapted appropriate architectures and hyperparameters for these three DL models, which are the prerequisites for comparison. The findings highlight that DNNs excel in prediction, offering robust performance and a remarkably narrow range of distributions of residuals. While the SAE-FNN does not match the performance of the DNN, it has notable strengths and serves as a viable alternative. With respect to the 1D CNN, its performance is less desirable for modeling nonlinearity and is not recommended for quantitative analyses of the NIR spectra of tobacco leaves.

2 Materials and methods

2.1 Data collection

The dataset comprises the NIR spectra of tobacco leaves and their pectin concentrations from laboratory measurements. There were 201 tobacco leaves from Yunnan, Guizhou, Hubei, Chongqing, Fujian, and Anhui production areas in 2014–2015 provided by China Tobacco Chongqing Industrial Co., Ltd., and China Tobacco Anhui Industrial Co., Ltd.

The tobacco leaves were maintained at 22 °C and 60% humidity for 48 h. The tobacco leaves were then dried strictly according to the standard operating procedures of YC/T31-1996, ensuring a precise moisture measurement. Following drying, the samples were ground and passed through a 40-mesh sieve. At least 150 g of the processed tobacco powder was compressed via a press, resulting in a smooth, crack-free surface. The 3-minute compression process guaranteed a consistent thickness of at least 10 mm.

The spectra of the tobacco leaves were collected on a Thermo Antaris II FT-NIR spectrometer (Thermo Scientific Thermo Electron Inc., USA) according to the method specified in DB/T497-2013. As shown in Fig. 1, the wavenumbers of the tobacco leaves range from 10000 cm^{-1} to 4000 cm^{-1}, with an interval of 4 cm^{-1}, and each individual has 1557 absorbances.

The pectin mass percentage of the tobacco leaves was measured via the standard m-hydroxydiphenyl method.

2.2 PLS method

PLS is known for its ability to establish a definitive relationship between the spectra and the response variables[21]. PLS achieves this by constructing new predictor variables, known as latent variables, which are linear combinations of the original predictor variables. It is therefore a linear chemometric regression model.

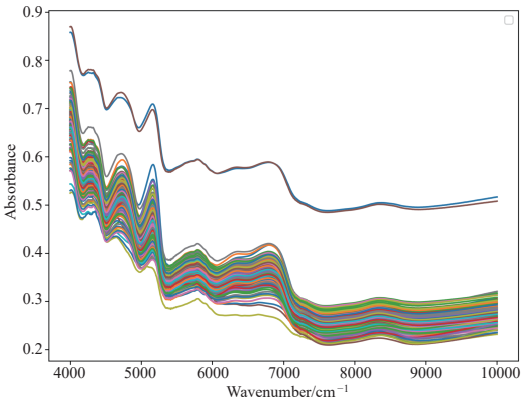


Fig. 1. NIR spectra of tobacco.

2.3 DNN method

A deep neural network (DNN) is a feedforward, artificial neural network with more than one layer of hidden units between its inputs and outputs[22,23]. In the neural network, each neuron performs a nonlinear transformation on the input signal through an activation function. In the connection relationship, the connection between every two neurons indicates that the passed data are weighted.

Compared with general artificial neural network (ANN) models, DNNs, which are characterized by numerous hidden layers and a large number of units per layer, are very flexible models with a vast array of parameters. This makes them capable of modeling highly complex and highly nonlinear relationships between inputs and outputs[24]. Additionally, DNNs include dropout layers to prevent overfitting, whereas ANNs generally have fewer layers and neurons and may lack dropout layers for controlling overfitting.

2.4 1D CNN method

Currently, CNNs are among the most popular types of DL tools applied in analytical chemistry. Unlike traditional DNN architectures with fully connected layers, the CNN[25] relies on convolution kernels to mutualize weights among a given layer, leading to a substantial decrease in the number of trainable parameters and a short learning time. Traditional CNNs designed for high-dimensional data cannot be used directly for the input data of one-dimensional sequences such as NIR spectra. Thus, an improved 1D CNN is needed.

The 1D CNN model[26] is composed of three types of layers[27], namely, the convolutional, pooling, and fully connected layers. The convolutional layer extracts features from the inputs, the pooling layer reduces the dimensionality of the input feature, and the fully connected layer connects the outputs from previous layers to the desired target outputs.

Neural networks represent a powerful nonlinear modeling technique that uses neurons as an architectural structure. Each neuron consists of a linear combination of all inputs, or neurons from a previous layer, and transforms them via a nonlinear activation function, for example, a rectified linear unit (ReLU)[21,28].

2.5 SAE-FNN method

In this study, the DL method is composed of a stacked

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607

autoencoder (SAE) and a feedforward fully connected neural network (FNN)[29]. The key steps of the developed SAE-FNN model can be described in two stages: SAE pretraining and FNN fine-tuning. In the first stage, an SAE is used to extract deep spectral features from the NIR image. In the second stage, an FNN layer is added to the last encoding layer of the pretrained SAE. Finally, an SAE-FNN regression model is established on the basis of the deep spectral features.

An SAE[30] is a deep neural network superimposed from a simple autoencoder (AE) structure. The depth feature reduction extraction of the SAE is essentially an encoding process, which nonlinearly maps a training input into hidden layers through a mapping function. The mapping function in SAE transforms input data into lower-dimensional feature representations through nonlinear transformations. The SAE stacks AE layers to progressively reduce the dimensions, eliminating redundancy while preserving essential information. It optimizes the reconstruction error to retain key information, especially related to chemical compositions. Its deep structure facilitates hierarchical learning, capturing multilevel critical information.

2.6 Hyperparameters of the DL models

In machine learning, hyperparameters are parameters whose values are set before starting the learning process. Setting hyperparameters usually depends on the experiences and trial-and-error strategies of researchers.

The number of epochs defines the number of times the complete training dataset will be propagated through the neural network. The learning rate (LR) is one of the most important hyperparameters and controls the degree to which the model weights are adjusted in response to the estimated error each time. The batch size determines the number of training samples used in each parameter update. A large batch size may oversimplify the model, hindering its ability to capture the complexity of the data and affecting its generalization ability. To better capture the inherent relationships within the data, small batch sizes were selected for this study. RMSProp, as an optimizer, was first proposed by Tieleman and Hinton[31] and is able to converge well in the presence of an unstable objective function. The Adam[32] optimizer has significant advantages, such as computational efficiency and few requirements for fine-tuning.

For the three DL methods used in this study, the outputs of each hidden layer are transformed by the activation function to capture nonlinearity. The ReLU introduces nonlinearity in the neural network with its simple, piecewise-linear function. This property allows the network to capture complex, nonlinear patterns in the data. The PreLU introduces learnable parameters that allow the neural network to learn more complex nonlinear features. The Leaky rectified linear unit (Leaky ReLU) introduces a slight negative slope, enhancing the model's capacity to manage negative inputs and enabling the network to learn a broader spectrum of nonlinear patterns. The exponential function of the exponential linear unit (ELU) renders it nonlinear in the negative region, a crucial property that enables the neural network to effectively learn and capture the intricate patterns and structures inherent in the data. The continuously differentiable exponential linear unit (CELU), an extension of the ELU, incorporates learnable parameters for fine-tuning the shape of the negative part. This adaptability allows CELU to accommodate diverse data distributions, enhancing the network's ability to model nonlinearity more effectively.

2.7 Evaluation metrics

In this study, the accuracy of the models was evaluated with the root mean squared error of calibration (RMSEC) and coefficient of correction determination (R_c^2). To evaluate the generalization performance of the models, we used the root mean squared error of prediction (RMSEP) and coefficient of predictive determination (R_p^2). We repeated the process of training the DL models ten times and took the standard deviation (SD) of ten RMSEPs as an indicator of model robustness. Generally, larger values of R^2 and smaller bias and RMSE values indicate better model performance. The definitions are shown in Eqs. (1), (2), (3), (4), and (5).

$$R_c^2 = 1 - \frac{\sum_{i=1}^{n_c} (\hat{y}_{ci} - y_{ci})^2}{\sum_{i=1}^{n_c} (y_{ci} - \bar{y}_c)^2}, \tag{1}$$

$$\mathrm{RMSEC} = \sqrt{\frac{\sum_{i=1}^{n_c} (\hat{y}_{ci} - y_{ci})^2}{n_c}}, \tag{2}$$

where n_c is the number of calibration spectra, i refers to the ith measurement spectrum, $\hat{y}_{ci}$ is the predicted value for calibration, y_{ci} is the actual calibration value, and $\bar{y}_c$ is the mean of the calibration values.

$$R_p^2 = 1 - \frac{\sum_{k=1}^{n_p} (\hat{y}_{pk} - y_{pk})^2}{\sum_{k=1}^{n_p} (y_{pk} - \bar{y}_p)^2}, \tag{3}$$

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{k=1}^{n_p} (\hat{y}_{pk} - y_{pk})^2}{n_p}}, \tag{4}$$

where n_p is the number of prediction spectra, k refers to the kth measurement spectrum, $\hat{y}_{pk}$ is the predicted value for prediction, y_{pk} is the actual prediction value, and $\bar{y}_p$ is the mean of the prediction values.

$$\mathrm{SD} = \sqrt{\frac{\sum_{j}^{M} \left(\mathrm{RMSEP}_j - \overline{\mathrm{RMSEP}}\right)^2}{M - 1}}, \tag{5}$$

where RMSEP_j is the jth RMSEP, $\overline{\mathrm{RMSEP}}$ is the average value of all RMSEPs and M is the number of RMSEPs, which is 10 in this study.

The PRD[33] is the ratio of the standard deviation of the measured value to the RMSEP. The PRD is commonly employed to evaluate prediction errors and is frequently utilized

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607

in analyzing a predictive model's performance on a prediction set. A PRD value less than 1.0 indicates a model with very poor predictive capability, whereas a value between 1.0 and 1.4 suggests a model with poor predictive capability. A range of 1.4–1.8 indicates limited predictive capability, whereas a range of 1.8–2.0 suggests an average predictive capability. A PRD value between 2.0 and 2.5 indicates good predictive capability, and a value above 2.5 indicates excellent predictive capability. The definition is shown in Eq. (6).

$$\mathrm{PRD} = \sqrt{\frac{\frac{1}{n_\mathrm{p}}\sum_{k=1}^{n_\mathrm{p}}(y_{\mathrm{p}k}-\bar{y})^2}{\frac{1}{n_\mathrm{p}}\sum_{k=1}^{n_\mathrm{p}}(\hat{y}_{\mathrm{p}k}-y_{\mathrm{p}k})^2}}, \tag{6}$$

where $\bar{y}$ is the average actual concentration of the prediction set.

The models were trained and evaluated on Google Colab (version "0.0.1a2"). PLS modeling was conducted with the scikit-learn Python library (version "1.2.2"). For the DNN, PyTorch (version "2.1.0") was utilized and optimized for performance on systems with CUDA 11.8 support. The 1D CNN and SAE-FNN architectures were implemented through TensorFlow (version "2.14.0").

3 Results

The dataset contains the NIR spectra of 201 tobacco leaves scanned by the FT-NIR spectrometer, each containing a total of 1557 absorbances, alongside its corresponding pectin concentration obtained from chemical measurements. The dataset was divided into calibration and prediction sets according to the SPXY[34] method on a 3∶1 basis. The number of samples in the calibration set was 151, and the number of samples in the prediction sets was 50. Given the limitations of the number of spectra, we did not use cross-validation in this study to assess whether the models exhibited overfitting. Instead, we relied on the performance of the calibration and prediction sets to make a combined judgment regarding the model's generalization ability.

3.1 Modeling data

The prerequisite for comparing the modeling effects of different models is to select the optimal hyperparameters for the corresponding models.

3.1.1 PLS method

The number of latent variables is a key parameter of PLS, which concerns the prediction capability of the PLS model. We plotted the variation in the prediction residual error sum of squares (PRESS) values against the number of latent variables in the PLS model and finalized 9 latent variables.

3.1.2 DL architectures

The DNN spectral model structure, illustrated in Fig. 2a, consists of three fully connected layers. The input is a 1-dimensional raw spectral segment, and the output is the predicted concentration of pectin. The first hidden layer is followed by a dropout layer to minimize overfitting. The dropout ratio is 0.5. Fig. 2b shows the network structure of the 1D CNN model. The network consists of three convolutional layers, each followed by an activation function and a pooling layer (or a flatten layer). The number of neurons in the convolutional layers increases sequentially to 8, 16, and 32, all with a kernel size of 3. The network ends with two fully connected layers. As shown in Fig. 2c, the SAE-FNN encoder contains two fully connected layers, and each hidden layer employs a sigmoid activation function. The decoder also has two fully connected layers, but different activation functions are used.

3.1.3 Selection of hyperparameters for the DNN Model

When adjusting and correcting errors of the DNN structure, a comprehensive approach is indispensable. In addition to optimizing the model architecture by strategically adjusting the number of layers and neurons, meticulous tuning of hyperparameters is crucial. In addition to these structural considerations, selecting the appropriate optimization algorithm and fine-tuning its internal parameters are vital for enhancing the efficiency and stability of the model during the training phase. These algorithms govern how the model updates its weights and biases on the basis of the error gradients, thereby guiding the learning process toward minimizing prediction errors.

In Fig. 3a and b, the epoch-RMSEC curves for ReLU are notably higher than those for the other activation functions, including ELU, PReLU, leaky ReLU and CELU. This finding indicates that the ReLU is unsuitable for the DNN model. The curves for LR=0.001 and LR=0.0005 sharply fluctuate compared with those for LR=0.0001. The curve for LR=0.0001 no longer decreases after 2500 epochs. Hence, we determine the model's initial learning rate to be 0.0001 and the number of epochs to be 2500, resulting in more stable performance.

The epoch RMSEC curves obtained by both the CELU and ELU methods highly overlap, as shown in Fig. 3c. Further investigations revealed that CELU consistently produces more stable results. Therefore, we select the CELU as the activation function. For the DNN model, we choose a small training batch size of 8. The model uses the mean squared error (MSE) and the Adam optimizer, which increase the prediction accuracy and training efficiency.

3.1.4 Selection of hyperparameters for the 1D CNN Model

In Fig. 4c, the overall performance of the epoch-RMSEC curves for LR=0.0001 is deemed unstable and therefore not selected. Additionally, for Fig. 4a and b, the PReLU curves indicate low RMSEC values, suggesting a good model fit. In Fig. 4b, the PReLU curve exhibits minimal vibrational significance, leading to the selection of LR=0.0005 as the initial learning rate and PReLU as the activation function. In Fig. 4b, the PReLU curve does not exhibit a decreasing trend after 2500 epochs. Therefore, we opt for 2500 epochs. Our final model configuration includes a batch size of 8, the Adam optimizer, and MSE as the loss function.

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607

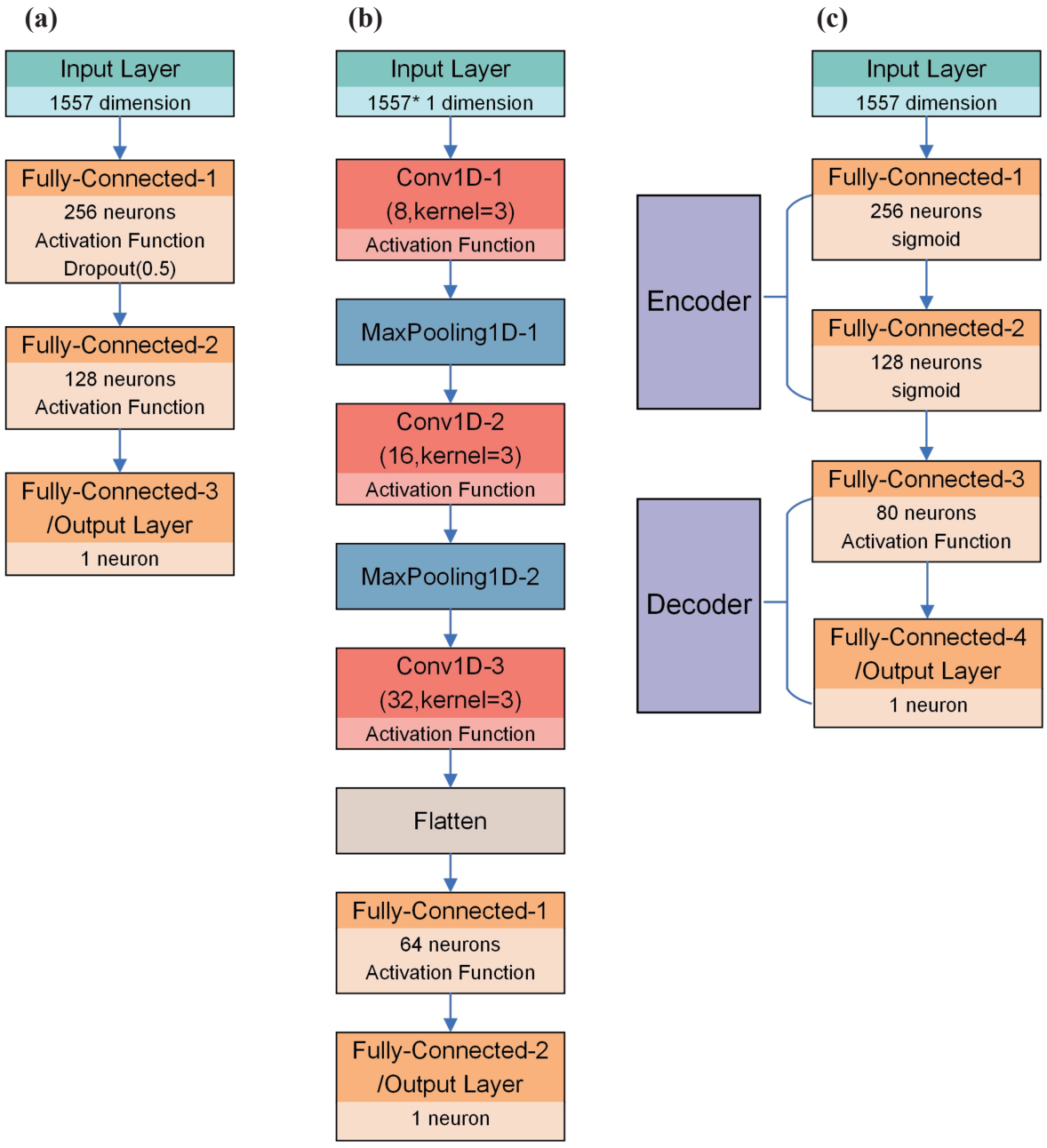


Fig. 2. Schematic structure of the three deep learning models. (a) Structure of the DNN model. (b) Structure of the 1D CNN model. (c) Structure of the SAE-FNN model.

3.1.5 Selection of hyperparameters for the SAE-FNN model

As illustrated in Fig. 5, the curves at LR=0.0001 appear to be stable and relatively smooth. Therefore, 0.0001 is chosen as the initial learning rate. At LR=0.0001, the ReLU curve has the smallest RMSEC values, indicating good model fit. Therefore, we opt for ReLU. Additionally, at LR=0.0001, after 10000 epochs, the ReLU curve no longer shows a decreasing trend; hence, the number of epochs is set to 10000. The batch size is set at 256, and in the SAE-FNN model, we employ RMSProp as the optimizer and MSE as the loss function.

3.2 Model performance

3.2.1 Evaluation of model performance

From Table 1, the RMSEP values for the PLS model, after preprocessing with MSC or SNV, are unexpectedly worse than those for the unprocessed model. This finding indicates that MSC and SNV preprocessing, intended to correct baseline shifts and scattering, actually degrades the PLS model's performance. These results suggest that these methods may introduce noise or artifacts, reducing the model's accuracy. In contrast, the unprocessed PLS model performed better, highlighting that MSC and SNV may have been detrimental for this dataset. This underscores the need to carefully

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607

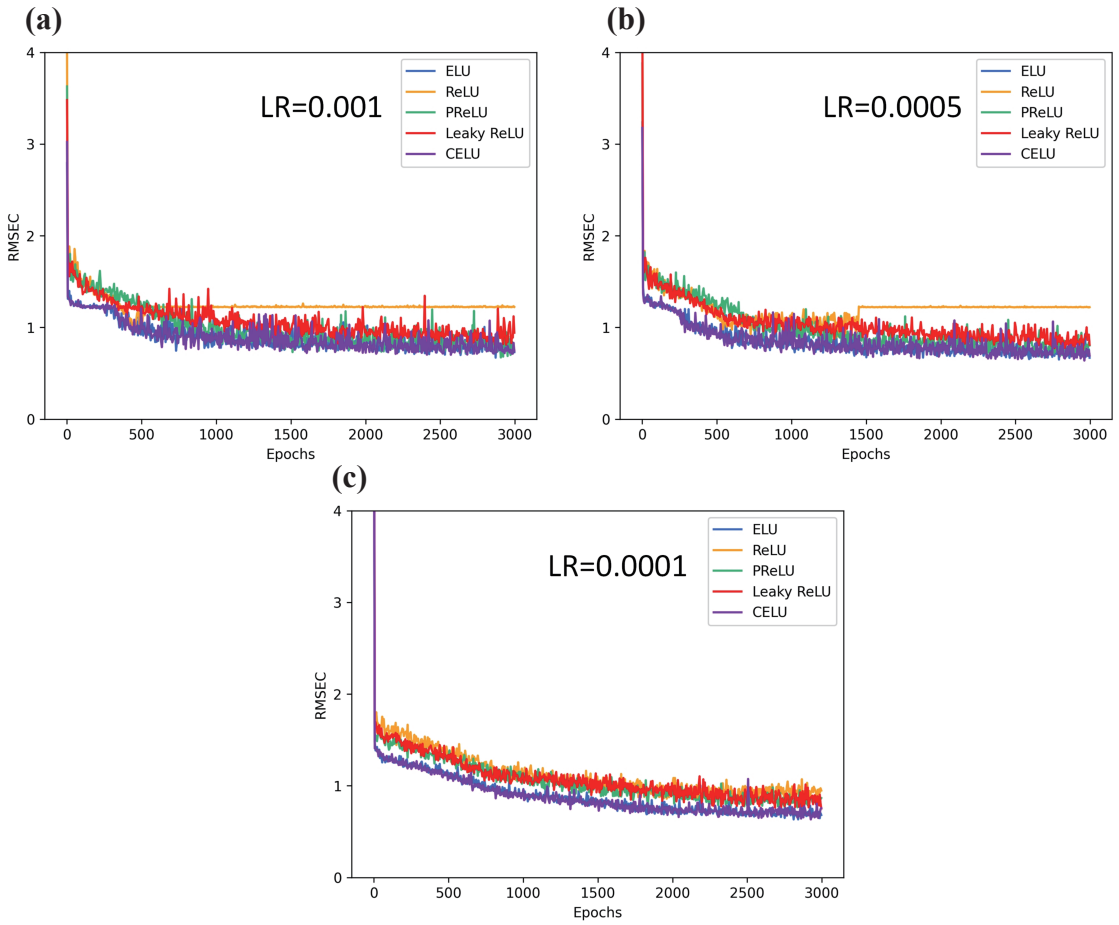


Fig. 3. Epoch-RMSEC relationship for different activation functions in the DNN model with varying learning rates. (a) LR=0.001. (b) LR=0.0005. (c) LR=0.0001.

evaluate preprocessing techniques to ensure that they improve, rather than impair, model performance. The subsequent sections perform comparative analysis using the unprocessed PLS model, which demonstrated better predictive performance.

Compared with the classical linear model PLS approach, we evaluated the performance of three models: the 1D CNN model, the DNN model, and the SAE-FNN model. The results for the calibration and prediction sets are shown in Table 1. The PRD values of all four models are between 2.0 and 2.5, which means that they are very good predictive models. Table 1 also shows that the 1D CNN has the longest runtime, which is up to 20 min, whereas the DNN and SAE models have runtimes of 3 min and 2 min, respectively.

Compared with the PLS model, the DNN model has a smaller value of RMSEC and a larger R_c^2, indicating that it is a good model for the calibration set. Its performance is also good for the prediction model. The average RMSEP for the DNN is 0.43, indicating that it is one of the most effective methods. Additionally, its low SD value of just 0.01 signifies the model's commendable robustness and consistency in performance. The results show that the DNN model has a very good ability to generalize, and the likelihood of overfitting is low.

The performance of the 1D CNN method on the calibration set is better than that of the PLS method, with a larger R_c^2 and a smaller RMSEC. However, its performance on the prediction set is inferior to that of the PLS method. The 1D CNN results in an average RMSE value of 0.48, which is among the largest values of all the models. Coupled with an SD of 0.02, there is a notable fluctuation in the range of results obtained each time, indicating poor robustness in performance.

In comparison, the SAE-FNN method yields better results than does the PLS method for the calibration set. However, in the prediction set, the performance of the SAE-FNN is comparable to that of PLS. Given that PLS establishes a clear quantitative relationship, it consistently yields definitive RMSE results. In contrast, the SAE-FNN, with an SD of 0.02, exhibits a certain degree of variability in its outcomes.

3.2.2 Distribution of residuals

The residual denotes the difference between the model's estimation and the actual measurement. The presence of residuals of approximately 0 signifies strong predictive ability. Therefore, we calculated the 10% and 90% quantiles of the distribution of residuals for each model, as listed in Table 2. The upper quantile (90%) signifies that 90% of the residuals are less than or equal to this upper quantile, thereby representing the upper boundary of the distribution of residuals. Conversely, the lower quartile (10%) denotes the lower boundary. The narrower the gap is between the upper quartile (90%) and the lower quartile (10%), the more concentrated

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607

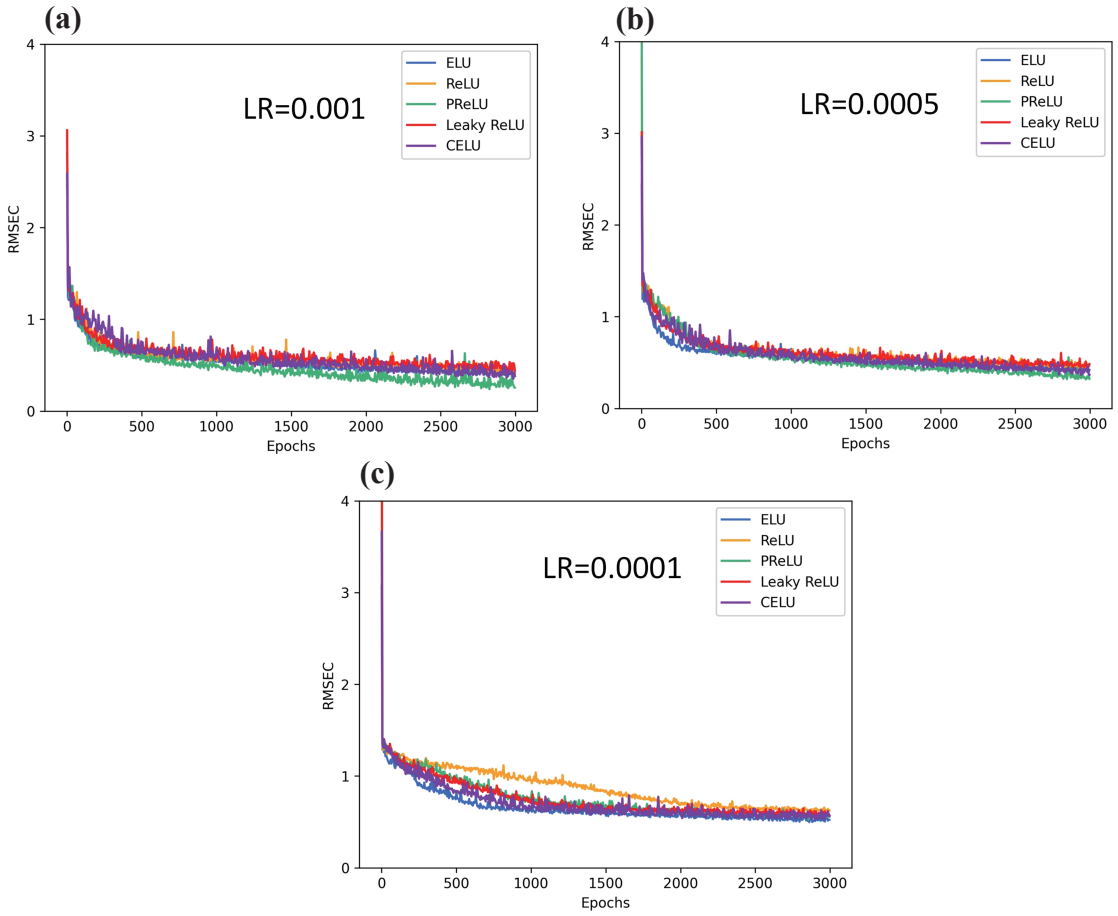


Fig. 4. Epoch-RMSEC relationship for different activation functions in the 1D CNN model with varying learning rates. (a) LR=0.001. (b) LR=0.0005. (c) LR=0.0001.

the distribution of the residuals becomes. This indicates that the model makes more accurate predictions, resulting in a smaller range of prediction errors. We also draw scatter plots of the distributions of the residuals for the various models on the prediction set, as shown in Fig. 6.

A narrow range of residuals and closeness to 0 for the SAE-FNN and DNN models indicate that the predicted values are close to the actual values, implying that the results predicted by these models are more accurate, especially in the case of the DNN model. Moreover, the residuals of the SAE-FNN and DNN models being more aggregated at approximately 0 typically suggest that these models are better than the PLS model at fitting the data in terms of capturing the primary trends and patterns in the data. A small and aggregated distribution of residuals indicates that the model neither overfits the noise of the training data (overfitting) nor ignores important features in the data (underfitting).

4 Discussion

On the basis of the above results, we realized that the quantitative results of the 1D CNN are unsatisfactory, and we identified possible explanations from related research[35]. The inputs of nonlinear functions are one-dimensional vectors, and convolution is a local operator rather than a global operator. This means that the convolution kernel affects only a portion of the sequence rather than the entire series. This contradicts the aim of nonlinear regression. Therefore, the original architecture of the 1D CNN is not suitable for nonlinear regression.

Better quantitative outcomes were achieved through modeling with the SAE-FNN method. SAE-FNN models often have strong generalizability because they can learn a generalized feature representation of the data. For SAE-FNN, the stacking of self-encoders usually involves layer-by-layer pretraining, which helps the model initialize the weights without labels. Afterwards, a small amount of labeled data can be used to fine-tune the entire network, which allows the knowledge gained from unsupervised learning to be utilized in supervised learning tasks. Self-encoders are usually robust to noise and outliers in the input data during the feature learning process because of their ability to reconstruct key features from noisy data. During the SAE training process, after artificially adding noise to the input data, the SAE reconstructs the original and clean data from the noisy input, with a focus on learning the key features. These features are progressively abstracted in the hidden layers, removing noise while preserving the information necessary for accurate reconstruction. The FNN receives the key features extracted by the SAE and performs precise feature mapping and regression. As a result, SAE-FNN integration significantly enhances the system's robustness to noise and outliers.

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607

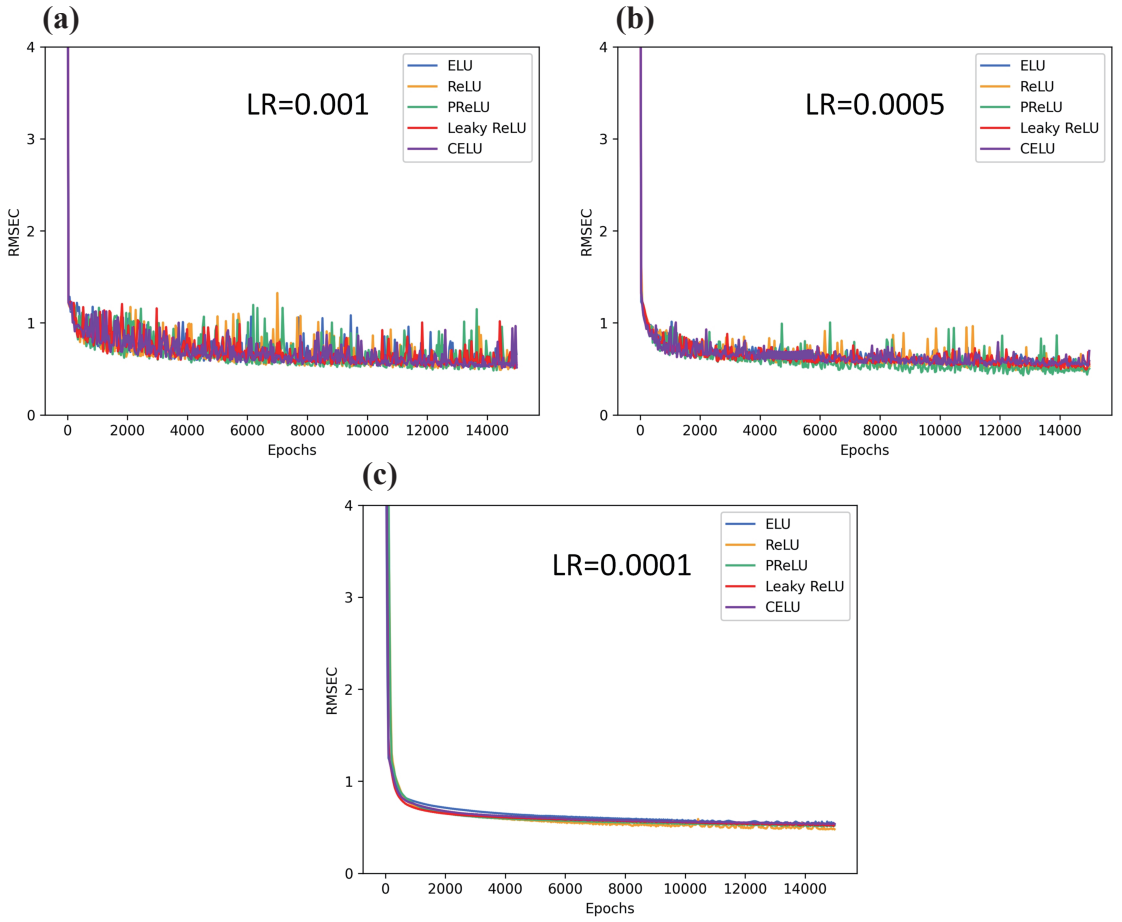


Fig. 5. Epoch-RMSEC relationships for different activation functions in the SAE-FNN model with varying learning rates. (a) LR=0.001. (b) LR=0.0005. (c) LR=0.0001

Table 1. Regression parameters of the three DL models and the PLS model.

Method	Calibration		Prediction			Average time for modeling operations (s)
	R_c^2	RMSEC	R_p^2	RMSEP (Mean ± SD)	PRD	
None-PLS	0.77	0.58	0.76	0.44	2.1	3
SNV-PLS	0.79	0.56	0.73	0.50	1.8	3
MSC-PLS	0.79	0.56	0.51	1.00	0.9	3
DNN	0.74	0.63	0.78	0.43±0.01	2.1	180
1D CNN	0.81	0.53	0.76	0.48±0.02	2.0	1200
SAE-FNN	0.81	0.53	0.78	0.44±0.02	2.1	120

Table 2. Lower and upper quantiles (10% and 90%) for the scaled distribution of residuals.

Model	Lower quantile (10%)	Upper quantile (90%)
PLS	−0.95	1.14
1D CNN	−0.85	1.33
DNN	−0.74	0.85
SAE-FNN	−0.98	1.02

On the basis of the above findings, employing the DNN method for nonlinear quantitative modeling yields impressive outcomes. For the fully connected layer of the DNN model, each neuron is connected to all the neurons in the previous layer, and each connection has a weight. This indicates that neural networks have the ability to learn various combinations of weights, including nonlinearity. Each neuron can weigh and combine inputs, introducing nonlinearity through an activation function. Furthermore, by stacking multiple fully connected layers, neural networks can learn multiple levels of feature representation, allowing them to better capture complex relationships in the data. The layer-by-layer structure of neural networks allows them to gradually extract and combine features, thus better adapting to nonlinearity.

Overall, the DNN model trained with large numbers of NIR spectra has stronger nonlinear fitting ability and generalization ability. It adeptly extracts features from NIR spectra, thereby enhancing the accuracy of predictive analys-

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607

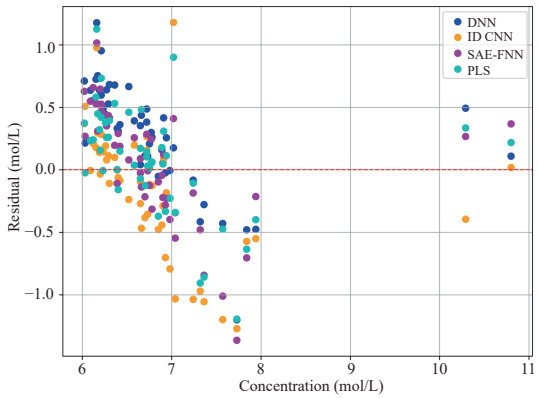


Fig. 6. Prediction residual plots for the three DL models and the PLS model.

is. The SAE-FNN model, developed through unsupervised learning, effectively manages nonlinearity in NIR spectra and shows a commendable generalization ability. However, in terms of the predictive power and robustness of the model, the SAE-FNN falls short of the DNN. Therefore, for nonlinear modeling based on NIR spectra, the DNN model is highly recommended for its efficacy, with the SAE-FNN as a viable alternative. Conversely, the 1D CNN, owing to its limited generalization ability and prolonged analysis duration, is not a choice for the quantitative regression of nonlinearity in NIR spectra.

5 Conclusions

This study explores three DL methods for modeling and analyzing nonlinearity in NIR spectra. Traditional linear methods are often ineffective in capturing complex and multidimensional relationships presented in chemical datasets. DL models, such as the DNN and the SAE-FNN, are able to learn and recognize these complex data structures through their multilayer and nonlinear processing capabilities. The DNN and SAE-FNN methods not only improve the accuracy and reliability of the analysis but also broaden our understanding of the intrinsic structure of the data. In particular, the DNN demonstrates outstanding predictive capabilities, supported by a significantly narrower range of distributions of residuals. They also exhibit a high level of generalizability and robustness, making them the choice for quantitative modeling of nonlinearity. These findings provide new perspectives on nonlinear quantitative regression analysis, especially its potential and application in dealing with complex chemical data.

Acknowledgements

This work was supported by a joint project with SINOPEC (Dalian) Research Institute of Petroleum and Petrochemicals Co., Ltd. (Contract No. 323061).

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Wenhui Yang received her Master's degree from the University of Science and Technology of China in 2024, under the supervision of Associate Professor Limin Shao. Her research mainly focuses on the quantitative analysis of near-infrared spectroscopy using deep learning method.

Limin Shao is currently an Associate professor of Analytical Chemistry in the Department of Chemistry at the University of Science and Technology of China (USTC). He received his B.S., M.S., and Ph.D. degrees from USTC, and completed postdoctoral research at the University of Idaho under the supervision of Professor Peter R. Griffiths. His research focuses on chemometrics and Fourier transform infrared spectrometry.

References

[1] Yun Y H, Li H D, Deng B C, et al. An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends in Analytical Chemistry*, **2019**, *113*: 102–115.
[2] Yu C, Liang D, Yang C, et al. Research progress and the application of near-infrared spectroscopy in protein structure and molecular interaction analysis. *Vibrational Spectroscopy*, **2022**, *121*: 103390.
[3] Zhang X L, Yang J, Lin T, et al. Food and agro-product quality evaluation based on spectroscopy and deep learning: A review. *Trends in Food Science & Technology*, **2021**, *112*: 431–441.
[4] Bian X H, Li S J, Fan M R, et al. Spectral quantitative analysis of complex samples based on the extreme learning machine. *Analytical Methods*, **2016**, *8* (23): 4674–4679.
[5] Rocha W F D C, Prado C B D, Blonder N. Comparison of chemometric problems in food analysis using non-linear methods. *Molecules*, **2020**, *25* (13): 3025.
[6] Mishra P, Passos D, Marini F, et al. Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends in Analytical Chemistry*, **2022**, *157*: 116804.
[7] Zhang W, Kasun L C, Wang Q J, et al. A review of machine learning for near-infrared spectroscopy. *Sensors*, **2022**, *22* (24): 9764.
[8] Mamouei M, Budidha K, Baishya N, et al. An empirical investigation of deviations from the Beer–Lambert law in optical estimation of lactate. *Scientific Reports*, **2021**, *11* (1): 13734.
[9] Blanco M, Coello J, Iturriaga H, et al. NIR calibration in non-linear systems: different PLS approaches and artificial neural networks. *Chemometrics and Intelligent Laboratory Systems*, **2000**, *50* (1): 75–82.
[10] Liu A, Li G, Fu Z G, et al. Non-linearity correction in NIR absorption spectra by grouping modeling according to the content of analyte. *Scientific Reports*, **2018**, *8* (1): 8564.
[11] Miller C E. Sources of non-linearity in near infrared methods. *NIR News*, **1993**, *4* (6): 3–5.
[12] Chauchard F, Cogdill R, Roussel S, et al. Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics and Intelligent Laboratory Systems*, **2004**, *71* (2): 141–150.
[13] Du Z, Tian W, Tilley M, et al. Quantitative assessment of wheat quality using near-infrared spectroscopy: A comprehensive review. *Comprehensive Reviews in Food Science and Food Safety*, **2022**, *21* (3): 2956–3009.
[14] Gan F, Luo J F. Simple dilated convolutional neural network for quantitative modeling based on near infrared spectroscopy techniques. *Chemometrics and Intelligent Laboratory Systems*, **2023**, *232*: 104710.
[15] Wang D, Zhao F Y, Wang R, et al. A lightweight convolutional neural network for nicotine prediction in tobacco by near-infrared spectroscopy. *Frontiers in Plant Science*, **2023**, *14*: 1138693.
[16] Wang Z X, Li J G, Zhang Z Q, et al. SBS content detection for modified asphalt using deep neural network. *Advances in Materials Science and Engineering*, **2020**, *2020*: 2513147.
[17] Shi G W, Gao J, Zhang X Y, et al. Quantitative detection of multicomponent SF_6 decomposition products based on Fourier

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607

transform infrared spectroscopy combined with SCARS-DNN. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **2024**, *311*: 123989.

[18] Yu X J, Yu X, Wen S T, et al. Using deep learning and hyperspectral imaging to predict total viable count (TVC) in peeled Pacific white shrimp. *Journal of Food Measurement and Characterization*, **2019**, *13* (3): 2082–2094.

[19] Yu X J, Lu H D, Wu D. Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging. *Postharvest Biology and Technology*, **2018**, *141*: 39–49.

[20] Ng W, Minasny B, Mendes W D S, et al. The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. *SOIL*, **2020**, *6* (2): 565–578.

[21] Einarson K A, Baum A, Olsen T B, et al. Predicting pectin performance strength using near-infrared spectroscopic data: A comparative evaluation of 1-D convolutional neural network, partial least squares, and ridge regression modeling. *Journal of Chemometrics*, **2022**, *36* (2): e3348.

[22] Du J, Xu Y. Hierarchical deep neural network for multivariate regression. *Pattern Recognition*, **2017**, *63*: 149–157.

[23] Chu Y W, Luo Y, Chen F, et al. Visualization and accuracy improvement of soil classification using laser-induced breakdown spectroscopy with deep learning. *iScience*, **2023**, *26* (3): 106173.

[24] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, **2012**, *29* (6): 82–97.

[25] Zhang C, Wu W Y, Zhou L, et al. Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (Lycium ruthenicum Murr.) using near-infrared hyperspectral imaging. *Food Chemistry*, **2020**, *319*: 126536.

[26] Mishra P, Passos D. A synergistic use of chemometrics and deep learning improved the predictive performance of near-infrared spectroscopy models for dry matter prediction in mango fruit. *Chemometrics and Intelligent Laboratory Systems*, **2021**, *212*: 104287.

[27] Cui C, Fearn T. Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. *Chemometrics and Intelligent Laboratory Systems*, **2018**, *182*: 9–20.

[28] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, **2017**, *60* (6): 84–90.

[29] Biganzoli E, Boracchi P, Mariani L, et al. Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. *Statistics in Medicine*, **1998**, *17* (10): 1169–1186.

[30] Zhou X, Zhao C J, Sun J, et al. Detection of lead content in oilseed rape leaves and roots based on deep transfer learning and hyperspectral imaging technology. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **2023**, *290*: 122288.

[31] Tieleman T, Hinton G. Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, **2012**, 4: 26–31.

[32] Kingma D, Ba J. Adam: A Method for Stochastic Optimization. arXiv: 1412.6980, **2014**.

[33] Viscarra Rossel R A, McGlynn R N, McBratney A B. Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy. *Geoderma*, **2006**, *137* (1-2): 70–82.

[34] Yang Z F, Xiao H, Zhang L, et al. Fast determination of oxide content in cement raw meal using NIR spectroscopy with the SPXY algorithm. *Analytical Methods*, **2019**, *11* (31): 3936–3942.

[35] Chen D, Hu F, Nian G, et al. Deep residual learning for nonlinear regression. *Entropy*, **2020**, *22* (2): 193.

CSTR: 32290.14.JUSTC-2024-0095
DOI: 10.52396/JUSTC-2024-0095
JUSTC, **2025**, 55(6): 0607