Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2022.28884

274 (2022) 130–139
October

Concentration prediction of imidacloprid in water through the combination of Fourier transform infrared spectral data and 1DCNN with multilevel feature fusion

Xin Liu[a], Xiaojiang Tang[a], Junwei Guo[b], Lianfeng Lin[a], Feng Huang[b,c,d,*], Eric Robert[c]

[a]*College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China, emails: liuxin166@163.com (X. Liu), tangxiaojiang@cau.edu.cn (X. Tang), linlianfeng@cau.edu.cn (L. Lin)*
[b]*College of Science, China Agricultural University, Beijing 100083, China, emails: huangfeng@cau.edu.cn (F. Huang), gjw020@cau.edu.cn (J. Guo)*
[c]*GREMI UMR7344 CNRS/Université d'Orléans, 45067 Orleans Cedex, France, email: eric.robert@univ-orleans.fr (E. Robert)*
[d]*LE STUDIUM Loire Valley Institute for Advanced Studies, Centre-Val de Loire region, France*

Received 5 March 2022; Accepted 12 August 2022

ABSTRACT

The Fourier transform infrared (FTIR) spectra combined with one-dimensional convolutional neural network (1DCNN) based on multi-level feature fusion, that is, MLF-1DCNN, were used to determine the concentration of imidacloprid in water. The FTIR spectra of imidacloprid water solutions with different concentrations (0–0.41 g/L) in 700–4,000 cm^{-1} were measured and the corresponding dataset was constructed, and the concentrations were predicted by the MLF-1DCNN. The effect of the spectral data preprocessing by multivariate scattering correction (MSC) and standard normal variate (SNV) transformation on improving the concentration prediction accuracy was studied. The result shows that the SNV preprocessing has the better prediction effect. The comparison of our model with partial least squares (PLS), support vector regression (SVR) and multiple linear regression (MLR) shows that our model can effectively predict the imidacloprid concentrations with a higher prediction accuracy than the other comparative models. The results obtained in this study demonstrate the analytical potential of applying this method to rapidly predict imidacloprid concentration in water.

Keywords: Concentration prediction; Convolutional neural network; Fourier transform infrared (FTIR) spectra

1. Introduction

Correct use of pesticides plays an important role in preventing crop diseases, pests, rodents, etc., and promoting high-quality and high-yield crops. With the increase in the production and consumption of pesticides year by year, the release of pesticide to the environment will increase, which will inevitably lead to an increase in the harm risk of human beings by the pesticides with high toxicity, strong stability, and leading to biological aggregation [1,2]. Usually, the utilization rate of pesticides is low, that is, most of the pesticides will remain in the environment, for example entering the soil and water through precipitation, surface runoff and soil leaching, thereby destroying the ecosystems [3] and posing a great threat to human health [4,5]. Imidacloprid is a kind of insecticide with high efficiency, long residual effect, easy absorption and potent contact killing [67], and if it enters the environment in

* Corresponding author.

large quantities, it will cause great harm on environment, including aquatic organisms, industrial and agricultural production water [8]. Therefore, the research on imidacloprid residues in the water is an essential part of water quality testing.

There are some available methods for the detection of imidacloprid, such as gas chromatography-mass spectrometry (GC-MS) [9], high performance liquid chromatography (HPLC) [10], HPLC-mass spectrometry (HPLC-MS) [11], nano-capsule probe [12] and nano-plasmonic biochips [13]. In addition, researches show that the specific fluorescence response of Tyr-MoO_3 quantum dots toward imidacloprid [14] or the colorimetric reaction via the aggregation of functionalized metal nanoparticles induced by pesticides as colorimetric sensors [15] can also be used for the detection of imidacloprid. These detection methods mentioned above are generally time consuming with many steps and costly. With the development of pesticide residue analysis technology, the simplified, miniaturized and high-sensitivity detection methods are constantly emerging. The infrared spectrum can be used in the quantitative analysis of pesticide residual, which uses Lambert–Beer law, that is, the absorption strength of a substance to light at a certain wavelength is proportional to the concentration and the thickness of the absorbing substance. Accordingly, the sample to be tested is put into the sample cup of the infrared spectrometer, and the corresponding absorbance is obtained by scanning and the transmittance at the selected wavelength band is measured to obtain the content of the tested sample. Quantitative detection can also be performed by combining spectral data and modeling including partial least squares (PLS), multiple linear regression (MLR) and artificial neural network (ANN) [16–22]. For example, Soyeurt et al. [16] used the infrared spectroscopy to analyze the content of several elements in milk. Basalekou et al. [21] combined infrared spectroscopy and least squares discrimination to accurately identify wines. In recent years, convolutional neural networks (CNNs) have been applied to spectral analysis. For example, Ng et al. [22] used a CNN with visible/near-infrared, mid-infrared and their combined spectra to predict soil properties.

So far, for the concentration prediction of imidacloprid in water, there is the lack of FTIR spectral dataset of imidacloprid in water with different concentrations and its efficient detection model. In this paper, the corresponding dataset was constructed and a one-dimensional CNN model based on multilevel feature fusion (MLF-1DCNN) was proposed to predict the concentration of imidacloprid in water.

2. Materials and methods

2.1. Experimental materials

The tested imidacloprid water dispersible granules, with an active ingredient content of 70%, were purchased from Yi Nong Plant Protection Mall. In our experiment, the imidacloprid in water with 42 different concentrations from 0 to 0.41 g/L with the gradient of 0.01 g/L were used. All the imidacloprid solutions with different concentrations were prepared under the same environmental conditions with the same water sample.

2.2. Experimental equipment

In this experiment, the infrared spectra of different concentrations of imidacloprid in water were detected by the FTIR spectrometer (Perkin Elmer), which includes a light source, an interferometer, a sample room, a detector and a computer (convert the interference signal into spectrum), as shown in Fig. 1. The detection principle is that when light source passing through the interferometer and sample room, all wavelengths of light source enter the detector at the same time, and the total light signal is split through Fourier transform (FT) calculation and the intensity of different wavelength is extracted to form the entire FTIR spectrum.

2.3. Data description

In our experiment, during detecting the imidacloprid concentration in water by FTIR, the wavenumber range is 700~4,000 cm^{-1} and each spectral curve is obtained by averaging eight scans. For each concentration of imidacloprid in water, 50 spectral curves were collected through repeated measurement, that is, a total of 2,100 spectra curves were collected from the 42 concentration samples of the imidacloprid water solutions. The obtained original 2,100 spectra curves with different concentrations are shown in Fig. 2. It can be seen that there are obvious absorption peaks in these samples in 1,500–1,700 cm^{-1} and 3,000–3,700 cm^{-1} with the strongest peaks mainly at 1,636 and 3,310 cm^{-1}.

Because the 2,100 original spectral curves in Fig. 2 are close together, the spectral characteristics of different concentrations cannot be seen intuitively. By averaging 50 spectral curves at each concentration to obtain the average spectrum, the variation of spectrum with imidacloprid concentration can be analyzed. Fig. 3a shows the example of average spectral curves at four different concentrations. It can be seen that the spectral intensity increases as imidacloprid concentration, especially at 1,636 and 3,310 cm^{-1}. In order to show the contribution of imidacloprid to the spectral peaks more clearly, the spectra curves at different concentrations by deducting the influence of water from the original curves were shown in Fig. 3b. It can be seen that imidacloprid mainly contributes the wider peak at 2,700–3,700 cm^{-1} and the peaks of about 1,636 and 1,065 cm^{-1}. The areas of these peaks mainly increase as the concentration and there is a relationship between the

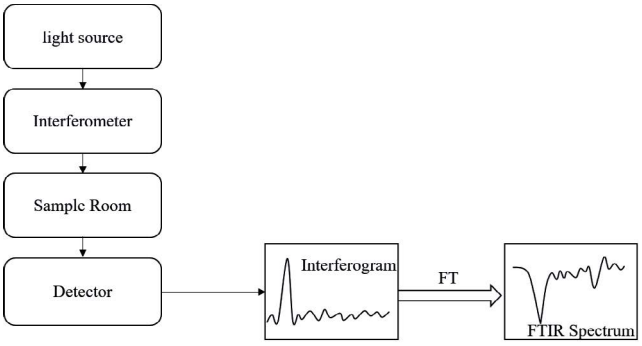

Fig. 1. Structure diagram of Fourier transform infrared spectrometer.

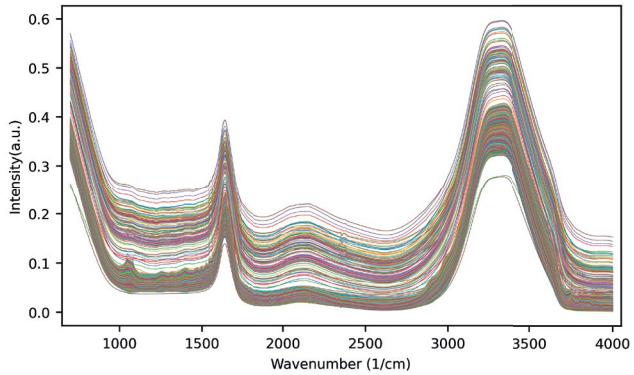

Fig. 2. Original 2100 FTIR spectra of imidacloprid in water with the different concentrations from 0 to 0.41 g/L.

areas of these peaks and the imidacloprid concentration. Thus, the method based on the calculation of the characteristic peak areas of imidacloprid in FTIR spectra is commonly used. However, with this method, it is needed to find the characteristic peaks first and calculate the peak areas, then the concentration of imidacloprid can be obtained, which is time consuming and manual. Since the spectra at different concentrations differ in overall and indetailed shapes, we propose the MLF-1DCNN model to quickly identify the concentration of imidacloprid water solution without calculating the areas of the peaks and without calculating the deduction of the spectra of imidacloprid water solutions and the spectra of water.

2.4. Spectra data preprocessing

The original spectral data can be affected by different noise interferences from the environmental background (for example, illumination) and from the instrument performance (for example, random error), causing baseline drift, light scattering, and sample inhomogeneity, which may influence the accuracy of the predicated concentration. In order to make the extracted spectral data more accurate and more effective, so that it can be used for the fast and accurate identification of our proposed model. It is necessary to preprocess the original spectra data to eliminate or reduce the noise influences. In this paper, we will compare the effectiveness of two preprocessing methods, that is, multivariate scatter correction (MSC) and standard normal variable (SNV) transformation.

MSC is one of the commonly used algorithms to eliminate the influence of light scattering, baseline translation and offset. The MSC algorithm assumes that the effect of light scattering on each sample at each wavelength is linear, and the algorithm is as follows.

(i) Calculate the average spectrum $\bar{x}_i$ of all spectra from Eq. (1), and use the average spectrum as the standard spectrum.

$$\bar{x}_i = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

(ii) Perform a linear regression operation on the spectrum of each sample and the standard spectrum by Eq. (2) to obtain the regression constant a_i and regression coefficient b_i of each spectrum relative to the standard spectrum.

$$x_i = a_i + b_i \bar{x}_i \tag{2}$$

(iii) The corrected spectrum x_{MSC} is obtained by performing MSC by Eq. (3).

$$x_{\text{MSC}} = \frac{x_i - a_i}{b_i} \tag{3}$$

In MSC preprocessing, the relative baseline inclination of each spectrum is corrected by subtracting the linear shift amount from the original spectrum of each sample and dividing by the regression coefficient, so that under the reference of the standard spectrum the baseline shift and offset of each spectrum are corrected, which is different for the imidacloprid samples of each concentration. The corresponding spectral absorption information has no influence in the whole process of data processing, and thus the MSC preprocessing improves the spectral signal-to-noise ratio.

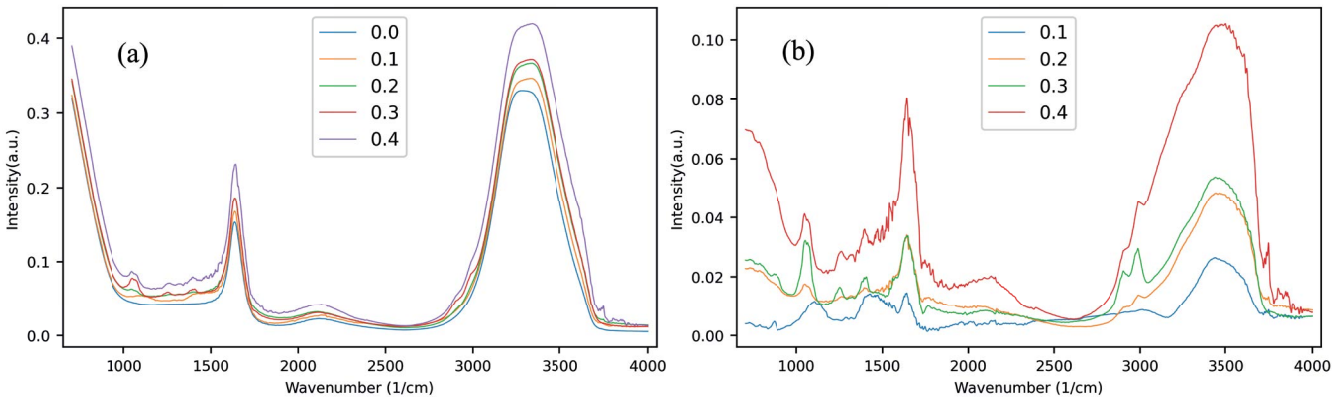

Fig. 3. Average spectra of the imidacloprid water solutions with four concentrations (a) and the corresponding spectra after deducting water spectrum (b).

For the method of SNV, the purpose of subtracting the linear translation from the original spectrum of each sample is very similar to the multivariate scatter correction, both to eliminate the influence of scattering effects. SNV considers that the scattering of the spectra satisfies a certain distribution law (i.e., standard normal distribution), and the spectral correction on each detection data of a spectrum curve according to the mean values of individual spectrum curve can be performed. Therefore, in most cases, the correction effect of SNV is better than MSC. The correction formula of SNV is shown in Eq. (4).

$$x_{\mathrm{SNV}} = \frac{x - \bar{x}}{\sqrt{\frac{\sum_{k=1}^{m}\left(x_k - \bar{x}\right)^2}{\left(m-1\right)}}} \tag{4}$$

where $\bar{x} = \frac{1}{m}\sum_{k=1}^{m} x_k$ and m is the number of detected wavelengths.

In this paper, both MSC and SNV are carried for the FTIR data processing of the imidacloprid water solutions with different concentrations. The preprocessed results by MSC and SNV are shown in Fig. 4a and b, respectively. In our experiment, the average spectrum was taken as the standard or ideal spectrum for MSC correction, which is practically acceptable and feasible. Through our MSC corrections, the spectral curve error is improved, the spectral characteristic absorption information related to imidacloprid content was enhanced, and the intensity range of all spectra becomes less dispersed decreasing from 0–0.6 to 0–0.4. Thus, the spectra curves become smoother and the characteristic information of all the spectra is increased, as shown in Fig. 4a. For the SNV preprocessing, the spectral curves after correction are shown in Fig. 4b. From each spectrum curve, it is seen that each detected data of individual spectrum curve was corrected according to Eq. (4), which makes the overall curve shape different from MSC. In the later model training and testing, these corrected spectral curves by MSC and SNV will correspond to the concentration labels.

3. Model framework

In this paper, based on the preprocessed FTIR spectra, the MLF-1DCNN is proposed to detect the concentrations of the imidacloprid water solutions, as shown in Fig. 5. The model consists of five convolution layers, two max pooling layers and five full connection layers. Meanwhile, two skip connection branches (C3-FC1 and C4-FC2) are also included to fuse low-level and high-level features. In the multi-layer feature fusion module, for FC1, FC2 and FC3, each contains three linear layers shown in Fig. 5b. For FC4 and FC, each contains one linear layer. The low-level local features of the spectra are extracted from the input 1D spectral data through two convolution layers and two pooling layers, and then the features of C3, C4 and C5 layers are fused by the skip connection method. Finally, the prediction results of the data are deduced through FC4 and the final FC layer. The multi-level feature fusion module can directly transfer the feature mapping of C3, C4 and C5 layers to FC4, which can fuse the low-level local features and high-level global features, realize the comprehensive utilization of multi-level data features, improve the overall feature extraction ability of the model, and improve the prediction ability of the model. Its basic parameters as shown in Table 1.

4. Results and discussion

4.1. Experimental environment

Experiments are performed on an NVIDIA GeForce GTX 1050 Ti GPU. In addition, it adopts Intel (R), Core (TM) i7-9700k (3.00 GHz) processor and 32 GB memory. The operating system is Windows 10 (64-bit). The implementation of the model uses Pytorch. To achieve faster graphics computation and less storage overhead, the Compute Unified Device Architecture (CUDA) toolkit 10.0 is applied in the experiments.

4.2. Performance evaluation index

In order to verify the performance of the proposed model, the determination coefficient R-square (R^2), the root mean squared error (RMSE), the mean absolute error (MAE), the maximum error (Max Error) and median absolute error (MedAE) are performed and documented. The above five evaluation indicators are defined as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2} \tag{5}$$

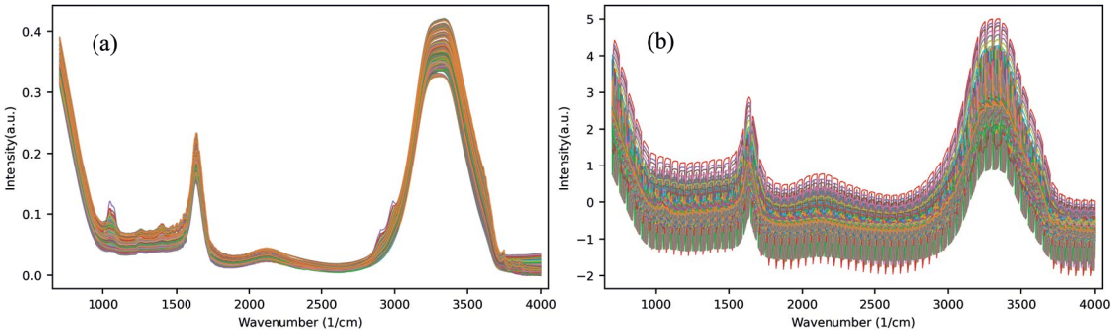

Fig. 4. Comparison of the preprocessed results by (a) MSC and (b) SNV on the FTIR spectra of the imidacloprid water solutions with different concentrations from 0 to 0.41 g/L.

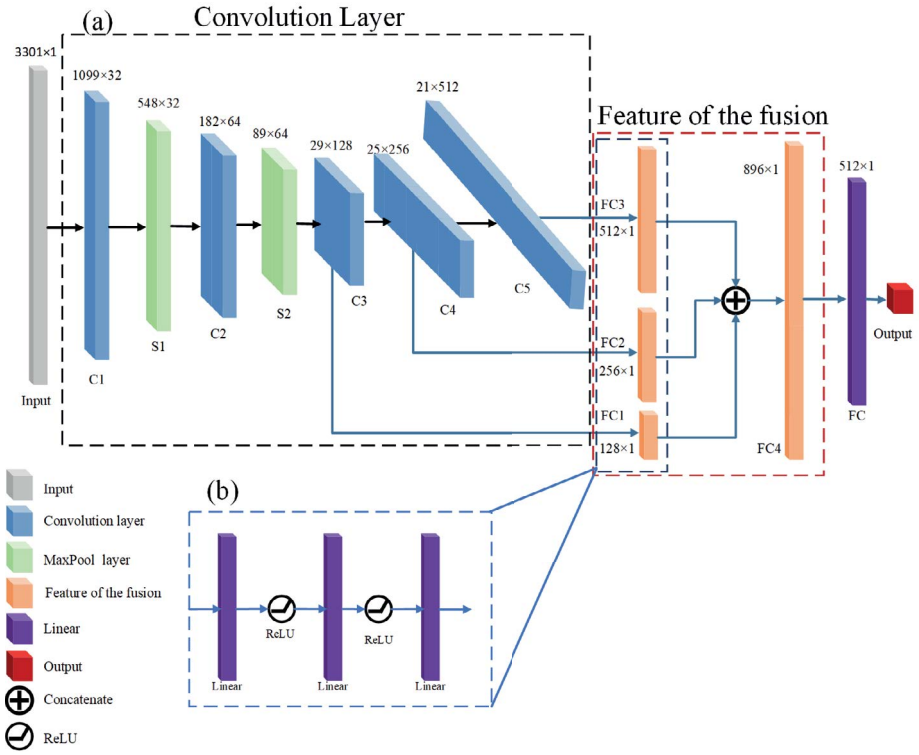

Fig. 5. The architecture of MLF-1DCNN.

Table 1
The structure and parameters of the MLF-1DCNN model

Network layer	Parameters
Input layer	Spectral data
Convolutional layer C1	Convolution kernel 5, stride 3, activation function ReLU
Pooling layer S1	Max pooling, size 5, stride 2
Convolutional layer C2	Convolution kernel 5, stride 3, activation function ReLU
Pooling layer S2	Max pooling, size 5, stride 2
Convolutional layer C3	Convolution kernel 5, stride 3, activation function ReLU
Convolutional layer C4	Convolution kernel 5, stride 1, activation function ReLU
Convolutional layer C5	Convolution kernel 5, stride 1, activation function ReLU
Fully connected layer	Linear activation function
Output layer	Output a neuron

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{6}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{7}$$

$$\text{Max Error} = \max\left(y_i - \hat{y}_i\right) \tag{8}$$

$$\text{MedAE} = \text{median}\left(\left|y_1 - \hat{y}_1\right|,\ldots,\left|y_n - \hat{y}_n\right|\right) \tag{9}$$

where n is the number of samples, y_i is the true value, and $\hat{y}_i$ is the predicted value. When the R^2 value is larger and closer to 1, it indicates that the predicted value is closer to the true value and that the prediction accuracy of the model is higher. RMSE and MAE indicate the error between the true value and the predicted value, and the smaller it is, the higher the prediction accuracy of the model is. Max Error can show the tolerance of the model for edge data and the robustness and generalization ability of the model. The smaller the value, the stronger the robustness of the model. MedAE value is used to measure the deviation between the predicted value and the true value. The smaller the deviation, the higher the reliability of the model.

4.3. Experimental results

Before the model training starts, the dataset of the 2,100 preprocessed FTIR spectra curves of the imidacloprid water solutions with different concentrations was constructed, in which 80% data in the dataset is randomly selected as the training set and the remaining 20% of the dataset is used as the validation set. The hyperparameter settings during the training process are shown in Table 2.

In model training, mean square error (MSE) is used as the loss function, which is defined in Eq. (10).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{10}$$

where n is the number of samples, y_i is the true value, and $\hat{y}_i$ is the predicted value.

Fig. 6 compares the prediction results of MLF-1DCNN (using the parameters in Table 2) without and with the preprocessing methods of MSC and SNV. In Fig. 6a–c each figure has a black line with the slope of 1.0, which indicates that if the predicted concentration values are on the black lines, it means that the predicted concentrations are equal to the real concentrations, that is, the prediction results on the lines are the most accurate. The deviation degree of the predicted concentrations from the lines and the number of deviation points can show the prediction effect of the method. The closer the predicted value is to the line, the more accurate the predicted value is. From the comparison of Fig. 6a–c, it is seen that more data points deviate from the black line in Fig. 6a than in Fig. 6b and c, which shows that without data preprocessing, the predicted concentrations exhibit larger deviation from the real concentrations than with data preprocessing. In addition, different data preprocessing methods also influence the prediction results. The comparison of Fig. 6b and c shows the predicted concentration by SNV preprocessing is closer to the real concentration than by MSC.

In order to further present the influence of data preprocessing method on the prediction effect of MLF-1DCNN, this paper also conducts a detailed evaluation through the five evaluation indicators. In comparing the evaluation indicators of different methods, the largest determination coefficient R^2 and the smallest error indicators are highlighted in bold. From these specific indicators in Table 3, it is seen that the MLF-1DCNN model with SNV preprocessing has the highest R^2 value and all the lowest error indicators (in bold), showing the best performance on the prediction of the concentration of the imidacloprid water solutions. Therefore, in the following part, the dataset preprocessed by SNV are selected for the optimization of MLF-1DCNN.

In machine learning training, the learning rate (LR) has an important influence on whether the training process can converge and the convergence speed. In order to obtain the best learning rate in MLF-1DCNN training process, the prediction accuracy of the model under different learning rates was compared. Fig. 7 shows the prediction effect of the MLF-1DCNN model under different learning rates (In this paper, the meaning of the black lines in Figs. 7–10 is same to that in Fig. 6, that is, the points on these lines indicate that the predicted concentrations are equal to the real concentrations.). From Fig. 7 it can be clearly seen that the prediction effect is the worst at LR = 0.005, but at LR = 0.001, 0.0005 and 0.0001, the prediction effect of MLF-1DCNN is difficult to distinguish by the naked eyes. So, it is necessary to compare the five evaluation indexes, shown in Table 4. As can be seen, at LR = 0.0005, R^2 is the highest and the error evaluations are the lowest (except for the Max Error ranking second), showing the best performance of the model at LR = 0.0005. Therefore, LR = 0.0005 is selected for MLF-1DCNN to make concentration prediction.

Similarly, different batch size will also have impact on the prediction result of imidacloprid concentration, as

Table 2
MLF-1DCNN model training hyperparameter settings

Hyperparameters	Numerical value
Learning rate (LR)	0.0001
Minimum sample of input batch	16
The maximum number of iterations	5,000
Optimization	Adam

Table 3
Comparison of evaluation indicators for using MLF-1DCNN with and without data preprocessing

Dataset	R^2	RMSE	MAE	Max Error	MedAE
Without data preprocessing	0.9797	0.0165	0.008	0.1273	0.0041
With MSC preprocessing	0.9969	0.0066	0.0037	0.0598	0.0025
With SNV preprocessing	**0.9984**	**0.0048**	**0.0031**	**0.0271**	**0.0019**

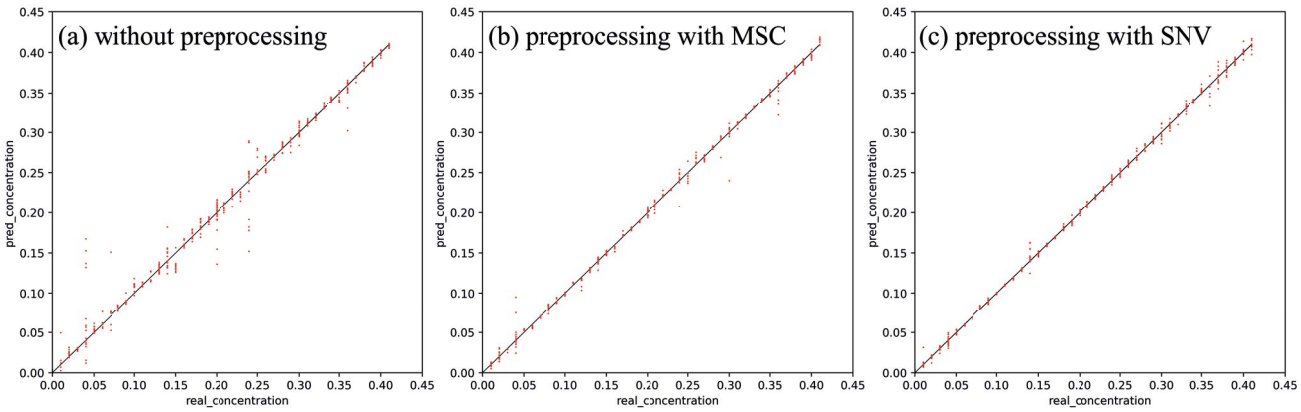

Fig. 6. Comparison of the predicted concentrations of the imidacloprid in water of the MLF-1DCNN model of without data preprocessing (a) and with the preprocessing by MSC (b) and by SNV (c).

shown in Fig. 8. It is seen that when batch size is set to 16, the prediction effect is the best. At the same time, when the batch size is 16, the results of the five evaluation indicators (i.e., the highest R^2 and the lowest errors) also show the best performance on the imidacloprid concentration prediction, which is shown in Table 5. Considering with the results of Fig. 8 and Table 5, the batch size is finally selected as 16.

Table 4
Prediction and evaluation indicators of our model (MLF-1DCNN) with different learning rates (LRs)

LR	R^2	RMSE	MAE	Max error	MedAE
0.005	0.3153	0.0981	0.0572	0.6737	0.0391
0.001	0.9966	0.007	0.0042	0.0639	0.0031
0.0005	**0.9988**	**0.004**	**0.0024**	0.0274	**0.0016**
0.0001	0.9984	0.0048	0.0031	**0.0271**	0.0019

In order to verify the impact of multi-level feature fusion (MLF) module on the performance of 1DCNN, we compared the predicted results by 1DCNN with and without MLF module, which are shown in Fig. 9 and Table 6. As can be seen from Fig. 9a, the predicted concentrations of our model (MLF-1DCNN) are closer to the real concentrations. In addition, in Table 6, the results of our model with the highest R^2 and lowest error indicators show that the better prediction performance. If the MLF module is removed, that is, the model becomes the 1DCNN (Fig. 9b and Table 6), R^2 is decreased from 0.9988 to 0.9911 and all the error indicators are significantly

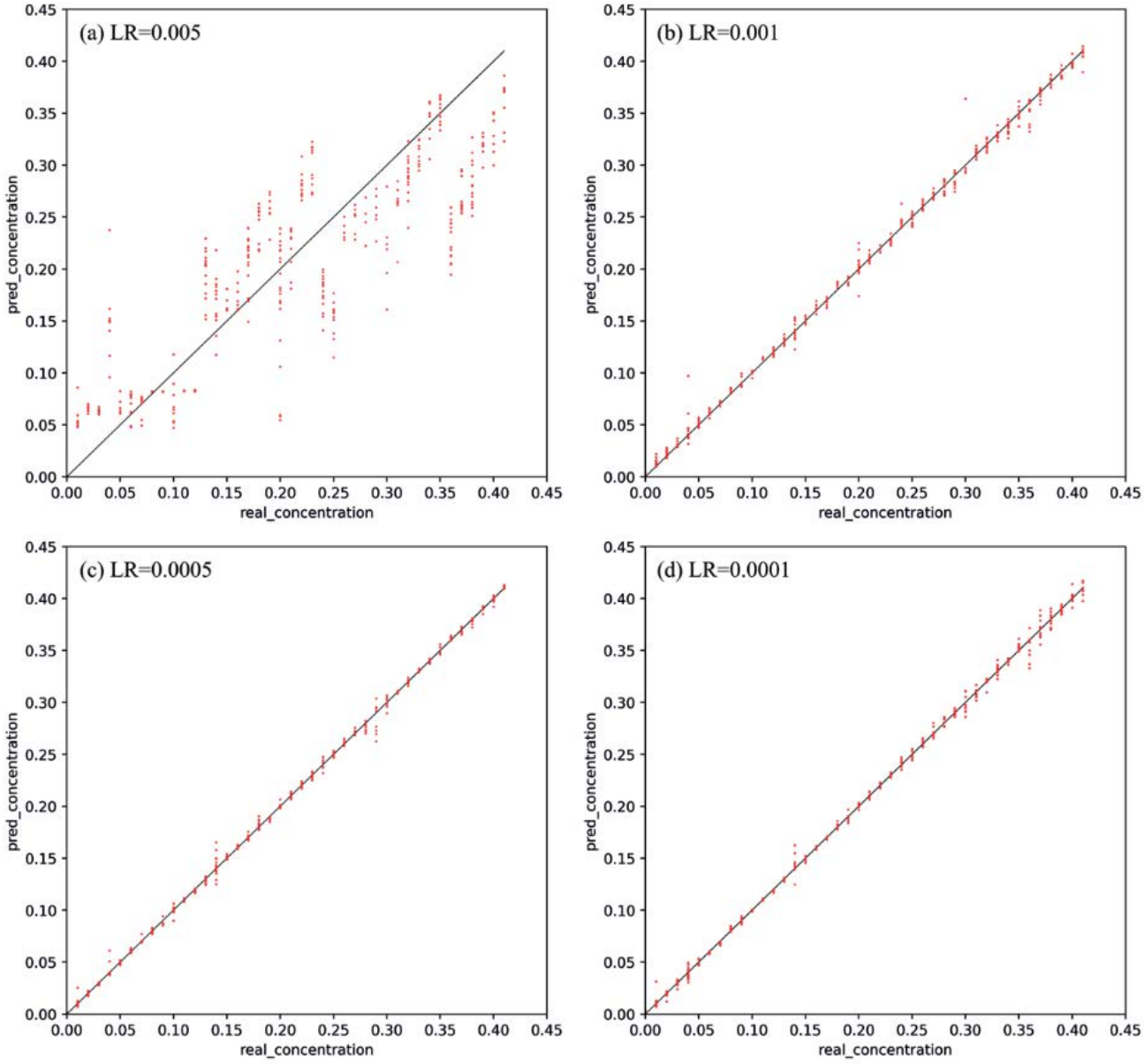

Fig. 7. Comparison of the prediction results by MLF-1DCNN model with different learning rates.

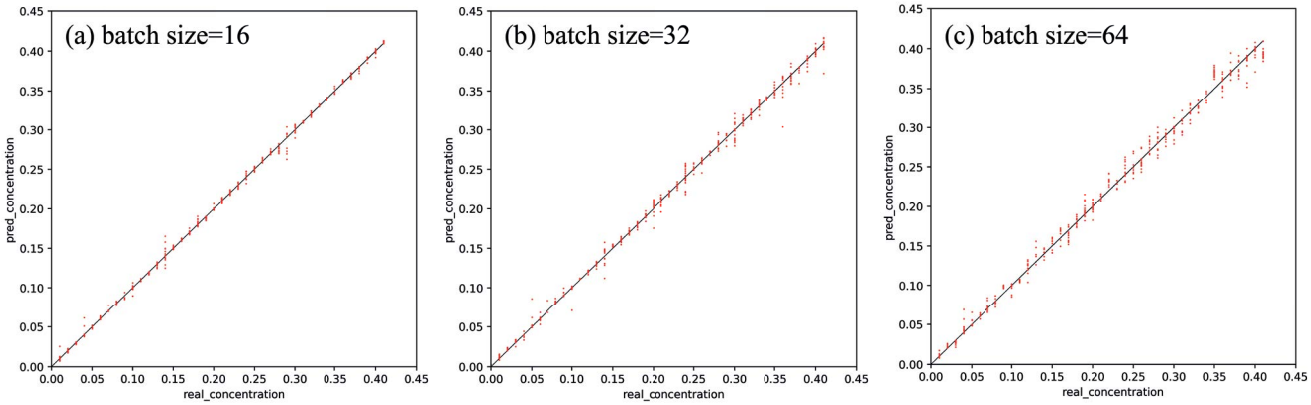

Fig. 8. Prediction comparison of imidacloprid concentration by MLF-1DCNN with different batch sizes.

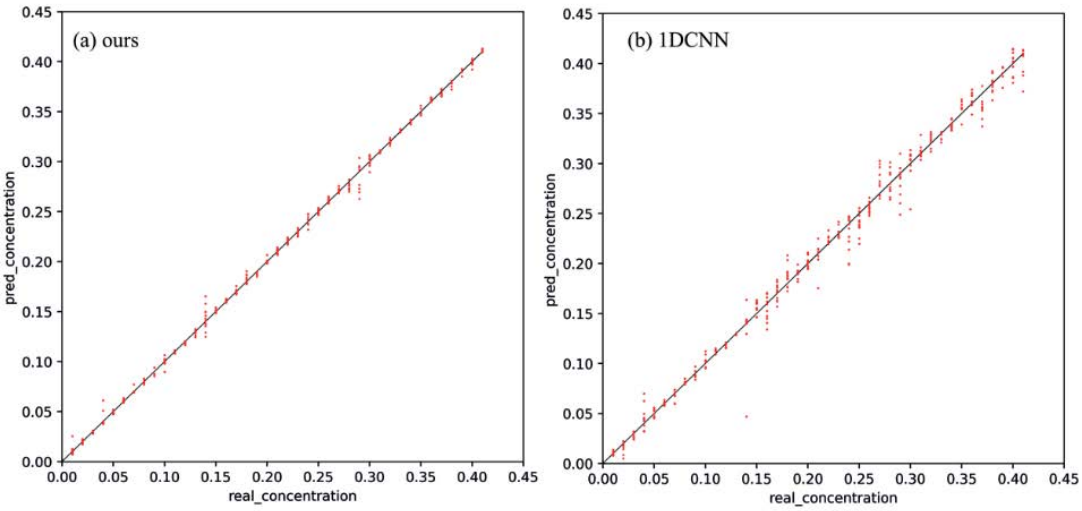

Fig. 9. Comparison of the prediction results by our model (MLF-1DCNN) and 1DCNN model.

increased. These results indicate that the MLF module can improve the prediction performance of the 1DCNN model with making the prediction results more accurate.

In this paper, the comparison on the predicted imidacloprid concentrations by MLF-1DCNN (with the selected parameters in Table 2) with the methods of Support Vector Regression (SVR), Multiple Linear Regression (MLR) and Partial Least Squares (PLS) are also conducted. Fig. 10 shows the comparison of the predicted imidacloprid concentrations by different models. From Fig. 10 it is seen that compared with other models, our model (MLF-1DCNN) has fewer deviation points and weaker deviation from the real concentration, which indicates that our model has better prediction performance, that is, the predicted concentration by our model fits the real concentration best, followed by PLS, and the worst is SVR.

Table 7 shows the evaluation indicators of imidacloprid concentration prediction by different models. It is seen that in these comparative models, our model has the largest coefficient of determination R^2 and the smallest error indicators. The R^2 value of our model (0.9988) is very close to 1, which is 11.4% higher than that of PLS and 57.1% higher than SVR. The results of five evaluation indicators in Table 7 show that our model has the best prediction performance for the determination of the concentration of imidacloprid in water.

Compared with some traditional detection methods such as chromatography, mass spectrometry and their combination, the proposed method in this study does not need to obtain the detailed information of the spectral peaks, and it makes the detection of imidacloprid pesticides simpler and faster. In addition, the proposed method extracts the spectra features through a 1DCNN with the multi-level feature fusion module, which can correlate and fuse the features of low levels and high levels of the spectra, and thus improves the feature extraction and generalization ability of the model and effectively improves the accuracy of model. The ablation experiment (in Fig. 9 and Table 6) proves the effectiveness of the multi-level feature fusion module. Compared with traditional machine learning methods, such as SVR, MLR and PLS, the proposed model shows the better performance, indicating the potential advantages of this method in the pesticide detection.

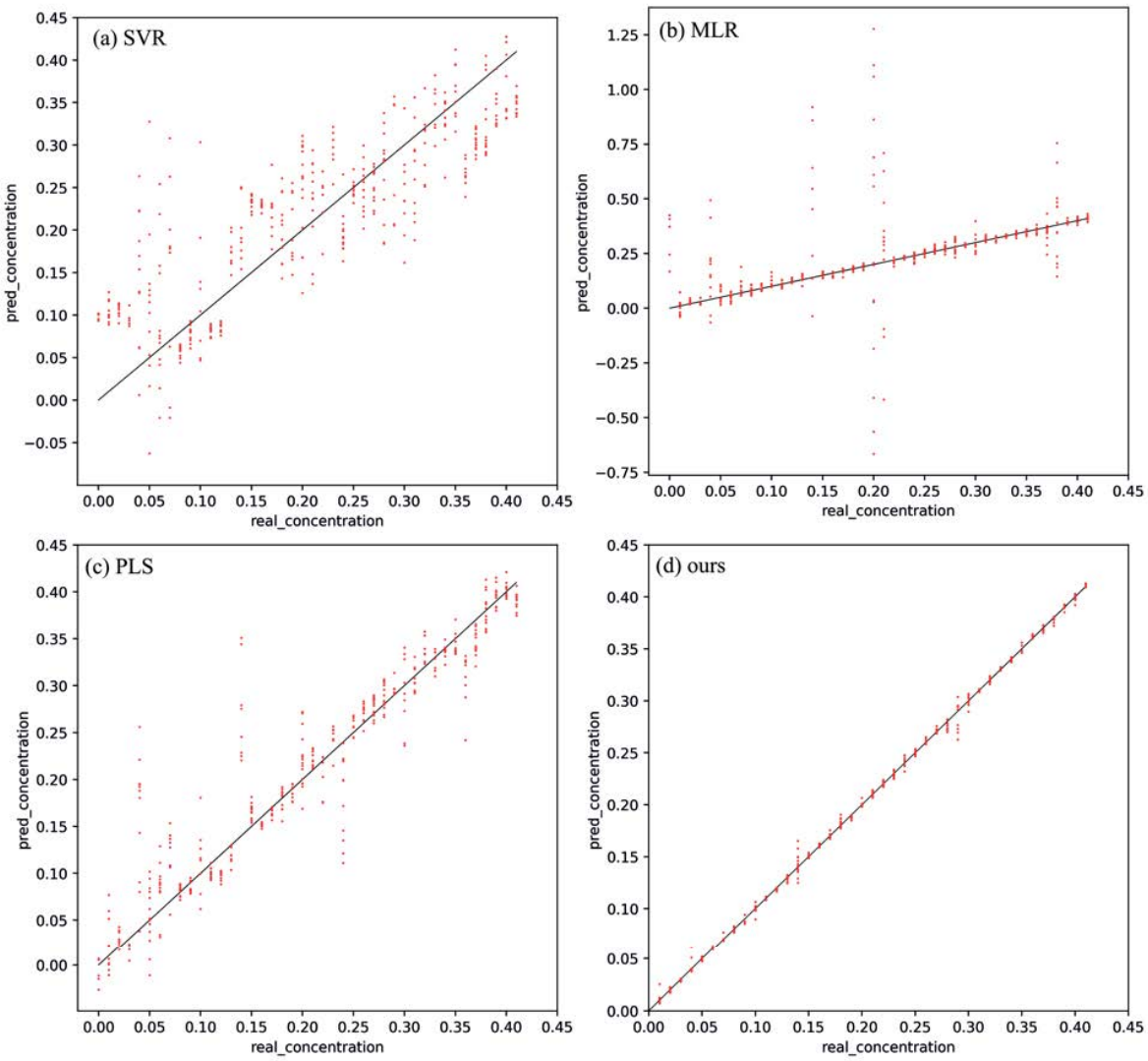

Fig. 10. Comparison of the predicted imidacloprid concentrations by different models.

Table 5
Prediction and evaluation indicators by MLF-1DCNN with different batch size

Batch size	R^2	RMSE	MAE	Max error	MedAE
16	**0.9988**	**0.004**	**0.0024**	**0.0274**	**0.0016**
32	0.9965	0.0071	0.0044	0.0563	0.0028
64	0.993	0.0101	0.0076	0.038	0.0056

Table 6
Evaluation indicators of imidacloprid concentration prediction by MLF-1DCNN (ours) and 1DCNN

Methods	R^2	RMSE	MAE	Max error	MedAE
Ours	**0.9988**	**0.004**	**0.0024**	**0.0274**	**0.0016**
1DCNN	0.9911	0.011	0.007	0.0931	0.0043

5. Conclusion

In order to accurately predict the concentration of imidacloprid in water, the FTIR dataset of imidacloprid water solution was constructed in this paper, and the SNV correction was used to preprocess the spectra data. A 1DCNN based on multi-level feature fusion is proposed, which fuses low-level features and high-level features in the FTIR spectra by adding two skip-connection branches, which

Table 7
Evaluation indicators of imidacloprid concentration prediction by different models

	R^2	RMSE	MAE	Max error	MedAE
SVR	0.6358	0.0709	0.0555	0.2937	0.0475
MLR	0.8117	0.1581	0.0524	0.1587	0.0121
PLS	0.8963	0.0386	0.0226	0.216	0.0135
Ours	**0.9988**	**0.004**	**0.0024**	**0.0274**	**0.0016**

improves the ability to extract spectral features. The results show that the model can accurately predict the concentration of imidacloprid in water, and the prediction effect is better than other comparative models. At the same time, the effectiveness and performance of the proposed model is better than the traditional detection or prediction methods, which provides a new method for pesticide detection.

Acknowledgments

This work was supported by the horizontal project (No. 202105511011054 and 202005511011203).

References

[1] X.M. Hua, X.L. Jiang, Characteristics and control countermeasures of pesticide pollution and its damage on environment in China, Res. Environ. Sci., 13 (2000) 40–43.
[2] J. Fenik, M. Tankiewicz, M. Biziuk, Properties and determination of pesticides in fruits and vegetables, TrAC, Trends Anal. Chem., 30 (2011) 814–826.
[3] X. He, Y.F. Ma, H.X. Zhao, X.J. Nie, Simultaneous determination of 24 pesticide residues in environmental water using solid-phase extraction and high-performance liquid chromatography-tandem mass spectrometry, J. Inst. Anal., 36 (2017) 1487–1493.
[4] O. Golge, B. Kabak, Determination of 115 pesticide residues in oranges by high-performance liquid chromatography–triple-quadrupole mass spectrometry in combination with QuEChERS method, J. Food Compos. Anal., 41 (2015) 86–97.
[5] P.N. Patil, S.D. Bote, P.R. Gogate, Degradation of imidacloprid using combined advanced oxidation processes based on hydrodynamic cavitation, Ultrason. Sonochem., 21 (2014) 1770–1777.
[6] M.P. Halm, A.Rortais, G. Arnold, J.N. Tasei, S. Rault, New risk assessment approach for systemic insecticides: the case of honey bees and imidacloprid (Gaucho), Environ. Sci. Technol., 40 (2006) 2448–2454.
[7] C. Segura, C. Zaror, H.D. Mansilla, M.A. Mondaca, Imidacloprid oxidation by photo-Fenton reaction, J. Hazard. Mater., 150 (2008) 679–686.
[8] T. Tisler, A. Jemec, B. Mozetic, P. Trebse, Hazard identification of imidacloprid to aquatic environment, Chemosphere, 76 (2009) 907–914.
[9] S. Rossi, A.G. Sabatini, R. Cenciarini, S. Ghini, S. Girotti, Use of high-performance liquid chromatography-UV and gas chromatography-mass spectrometry for determination of the imidacloprid content of honeybees, pollen, paper filters, grass, and flowers, Chromatographia, 61 (2005) 189–195.
[10] M. Rancan, A.G. Sabatini, G. Achilli, G.C. Galletti, Determination of Imidacloprid and metabolites by liquid chromatography with an electrochemical detector and post column photochemical reactor, Anal. Chim. Acta, 555 (2006) 20–24.
[11] M.J. Hengel, M. Miller, Analysis of pesticides in dried hops by liquid chromatography-tandem mass spectrometry, J. Agric. Food Chem., 56 (2008) 6851–6856.
[12] X.D. Yu, Y.S. Li, X.L. Liu, B. Qiao, Y.Y. Yang, Y.Y. Zhang, P. Hu, S.Y. Lu, H.L. Ren, Z.S. Liu, M.Y. Liu, Y. Zhou, Polyelectrolyte nanocapsule probe for the determination of imidacloprid in agricultural food samples, Food Agric. Immunol., 30 (2019) 432–445.
[13] K.L. Lee, M.L. You, C.H. Tsai, E.H. Lin, S.Y. Hsieh, M.H. Ho, J.C. Hsu, P.K. Wei, Nanoplasmonic biochips for rapid label-free detection of imidacloprid pesticides with a smartphone, Biosens. Bioelectron., 75 (2016) 88–95.
[14] M.R. Kateshiya, N.I. Malek, S.K. Kailasa, Facile synthesis of highly blue fluorescent tyrosine coated molybdenum oxide quantum dots for the detection of imidacloprid pesticide, J. Mol. Liq., 319 (2020) 114329.
[15] V.N. Mehta, N. Ghinaiya, J.V. Rohit, R.K. Singhal, H. Basu, S.K. Kailasa, Ligand chemistry of gold, silver and copper nanoparticles for visual read-out assay of pesticides: a review, TrAC, Trends Anal. Chem., 153 (2022) 116607.
[16] H. Soyeurt, D. Bruwier, J.M. Romnee, N. Gengler, C. Bertozzi, D. Veselko, P. Dardenne, Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry, J. Dairy Sci., 92 (2009) 2444–2454.
[17] H.R. Xu, H.J. Wang, K. Huang, Y.B. Ying, C. Yang, H. Qian, J. Hu, Comparison of PLS and SMLR for nondestructive determination of sugar content in honey peach using NIRS, Spectrosc. Spectr. Anal., 28 (2008) 523–526.
[18] W.B. Zheng, X.P. Fu, Y.B. Ying, Spectroscopy-based food classification with extreme learning machine, Chemom. Intell. Lab. Syst., 139 (2014) 42–47.
[19] X.D. Li, W.J. Mao, W. Jiang, Extreme learning machine based transfer learning for data classification, Neurocomputing, 174 (2016) 203–210.
[20] A. Morellos, X.E. Pantazi, D. Moshou, T. Alexandridis, R. Whetton, G. Tziotzios, J. Wiebensohn, R. Bill, A.M. Mouazen, Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy, Biosyst. Eng., 152 (2016) 104–116.
[21] M. Basalekou, C. Pappas, P. Tarantilis, Y. Kotseridis, S. Kallithraka, Wine authentication with Fourier transform infrared spectroscopy: a feasibility study on variety, type of barrel wood and ageing time classification, Int. J. Food Sci. Technol., 52 (2017) 1307–1313.
[22] W. Ng, B. Minasny, M. Montazerolghaem, J. Padarian, R. Ferguson, S. Bailey, A.B. McBratney, Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra, Geoderma, 352 (2019) 251–267.