

# Detection and recognition of veterinary drug residues in beef using hyperspectral discrete wavelet transform and deep learning

Rongchang Jiang<sup>1,3</sup>, Jingxin Shen<sup>1,2</sup>, Xinran Li<sup>1</sup>, Rui Gao<sup>1</sup>, Qinghe Zhao<sup>1</sup>, Zhongbin Su<sup>1\*</sup>

(1. College of Electrical and Information, Northeast Agricultural University, Harbin 150030, China;

2. Shandong Academy of Agricultural Machinery Sciences, Jinan 250100, China;

3. Harbin Big Data Center, Harbin 150030, China)

**Abstract:** A fast, non-destructive recognition method for veterinary drug residues in beef was proposed to mitigate the laborious sample preparation and long detection times associated with conventional chemical detection techniques. Control beef samples free of veterinary drug residues and four groups of beef sprayed with relevant concentrations of metronidazole, ofloxacin, salbutamol, and dexamethasone under ambient conditions were analyzed by 400-1000 nm hyperspectral imaging followed by multiplicative scatter correction preprocessing. Data dimension reduction was performed using Competitive Adaptive Reweighted Sampling (CARS), Principal Component Analysis (PCA), and Discrete Wavelet Transform (DWT) based on Haar, db3, bior1.5, sym5, and rbio1.3 wavelet basis functions. Treated data were subjected to Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM) modelling. CNN, MLP, SVM, and RF algorithms achieved overall accuracies of 91.6%, 88.6%, 87.6%, and 86.2%, respectively, when combined with DWT (wavelet basis functions and numbers of transform layers being Haar-4, db3-2, bior1.5-4, and sym5-3, respectively). The algorithm Kappa coefficients (0.89, 0.86, 0.85, and 0.83, respectively) and time consumption for prediction (140.60 ms, 57.85 ms, 70.67 ms, and 87.16 ms, respectively) were also superior to models based on CARS and PCA. DWT combined with deep learning can shorten prediction times, considerably improve the accuracy of classification and recognition, and alleviate the Hughes phenomenon, thus providing a new method for the fast, non-destructive detection and recognition of veterinary drug residues in beef.

**Keywords:** hyperspectral, beef, veterinary drug residues, discrete wavelet transform, convolutional neural network, deep learning

**DOI:** 10.25165/j.ijabe.20221501.6459

**Citation:** Jiang R C, Shen J X, Li X R, Gao R, Zhao Q H, Su Z B. Detection and recognition of veterinary drug residues in beef using hyperspectral discrete wavelet transform and deep learning. *Int J Agric & Biol Eng*, 2022; 15(1): 224–232.

## 1 Introduction

Beef is an indispensable part of the diet of the population and, unavoidably, veterinary drugs will be used to prevent and treat livestock diseases during the production process. However, non-standard use and abuse of veterinary drugs have become increasingly severe due to the pursuit of economic benefits and a lack of professional practices among some producers. Consequently, veterinary drugs can accumulate in beef for human consumption. Ingesting meat containing drug residues above safe limits can lead to drug resistance, poisoning, and alteration of intestinal flora in adults, while children can also suffer from abnormal physiological precocity, obesity, malformation, and even

genetic mutations<sup>[1]</sup>. Therefore, it is essential to rapidly, efficiently, and accurately detect and identify veterinary drug residues in beef.

Physicochemical and immunoassay methods are commonly used to detect veterinary drug residues, the former<sup>[2-6]</sup> involving mainly gas chromatography, high-performance liquid chromatography, and chromatography-mass spectrometry methods, which often require derivatization steps and compound separation before the detection of the analytes. Despite offering high accuracy and favorable sensitivity and stability, these processes can be complicated and time-consuming. The latter techniques<sup>[7-9]</sup> include enzyme immunoassay, fluorescence immunoassay, and colloidal gold immunoassay which are means of detecting drug residues by virtue of specific binding of antigens and antibodies. Immunoassays combine the merits of high selectivity and simple sample processing but can be subject to background interferences. They generally have low sensitivity and accuracy and are prone to false-negative or false-positive results<sup>[10]</sup>.

Spectroscopic techniques are becoming more common for the detection of veterinary drug residues, particularly Raman spectroscopy and infrared (IR) spectroscopy. For example, Tao et al.<sup>[11]</sup> demonstrated the rapid detection of testosterone propionate in duck using surface-enhanced Raman spectroscopy with an  $R^2$  coefficient between true and predicted values of 0.9996, root mean square error of prediction of 0.45 mg/L, and an average recovery of  $\geq 91.0\%$ . Guo<sup>[12]</sup> determined the structural characteristics of molecularly imprinted polymers using IR spectroscopy and

**Received date:** 2021-01-19 **Accepted date:** 2021-11-08

**Biographies:** **Rongchang Jiang**, PhD candidate, Senior Engineer, research interest: agricultural engineering and UAV remote sensing, Email: [jake\\_jrc@neau.edu.cn](mailto:jake_jrc@neau.edu.cn); **Jingxin Shen**, PhD candidate, Senior Engineer, research interest: agricultural electrification and automation, Email: [420270637@qq.com](mailto:420270637@qq.com); **Xinran Li**, PhD candidate, Senior Engineer, research interest: agricultural electrification and automation, Email: [827623386@qq.com](mailto:827623386@qq.com); **Rui Gao**, PhD, Lecturer, research interest: agricultural engineering and machine vision, Email: [415730327@qq.com](mailto:415730327@qq.com); **Qinghe Zhao**, PhD candidate, senior engineer, research interest: agricultural electrification and automation, Email: [zhaoqinghe@neau.edu.cn](mailto:zhaoqinghe@neau.edu.cn).

**\*Corresponding author:** **Zhongbin Su**, PhD, Professor, research interest: agricultural informatization and agricultural artificial intelligence research, Northeast Agricultural University, Harbin 150030, China. Tel: +86-18304623668, Email: [suzb001@163.com](mailto:suzb001@163.com).

established a method of detecting sulfadimethoxine residues in pork (standard recovery 65%-85%, relative standard deviation 5.0%). However, while Raman and IR spectroscopy are rapid and non-destructive techniques, their accuracy and sensitivity can be impacted by other light sources.

Hyperspectral technologies have been applied frequently in recent years to the detection of agricultural products. They are superior to spectral technologies like Raman and IR spectroscopy due to their ability to combine images and spectra. Hyperspectral technologies not only generate spatial distribution information from spectra but can also accurately identify chemical components of the object being observed using continuous and dense spectral signals. Sun et al.<sup>[13]</sup> established a model for lettuce leaves using hyperspectral analysis. Their RF-RFE-SPA-LSSVR model identified mixed pesticide residues with a prediction accuracy of 93.8%. Ji et al.<sup>[14]</sup> acquired spectral data (900-1700 nm) from spinach leaves using a hyperspectral imager, identified eight characteristic wavelengths via the Chi-square test-based feature selection algorithm and established a recognition model using SVM. Model prediction accuracy was 99.3%, indicating that pesticide residues in leaves could be accurately identified using this approach. Gao et al.<sup>[15]</sup> used a portable hyperspectral imaging system to collect field strawberry sample data in real-time. After dimension reduction by principal component analysis, established a maturity recognition model using Convolutional Neural Network (CNN). The results showed that the model prediction accuracy was 98.6%.

Given these advantages of non-destructive and fast hyperspectral detection techniques, the limitations of laborious sample preparation and the long detection times of existing methods for veterinary drug residues, a detection and recognition algorithm for such residues in beef is presented based on hyperspectral Discrete Wavelet Transform (DWT) and deep learning. This algorithm was applied to hyperspectral datasets from control beef samples and samples containing the veterinary drug residues metronidazole, ofloxacin, salbutamol, and dexamethasone, and the reliability and accuracy of the proposed algorithm were verified.

## 2 Materials and methods

### 2.1 Sample preparation

All beef used in this study was purchased from a large supermarket in Harbin City, Heilongjiang Province, China. Stock solutions of the following veterinary drugs were purchased: metronidazole (500 mg/L), ofloxacin (300 mg/L), salbutamol (150 mg/L), and dexamethasone (500 mg/L). Working solutions of metronidazole (0.75  $\mu\text{g}/\text{kg}$ ), ofloxacin (100.00  $\mu\text{g}/\text{kg}$ ), and dexamethasone (0.75  $\mu\text{g}/\text{kg}$ ) were prepared in distilled water in accordance with No. 235 Bulletin - Maximum Residue Limit of Veterinary Drugs in Animal Foodstuff issued by the Chinese Ministry of Agriculture and Rural Affairs. As a banned drug, salbutamol was prepared at a lower concentration (0.30  $\mu\text{g}/\text{kg}$ ). The working solutions were placed in watering cans with a jet distance of 8-10 cm. Each solution was used to evenly spray ten beef samples at room temperature (20°C) and relative humidity (RH) of 65% and allowed to absorb naturally for 3 h. Samples were then transferred to the preservative layer of a refrigerator for 9 h (5°C, 40% RH) to allow further diffusion of the drug solutions through the beef samples. Surplus liquid on the surface was then removed with absorbent paper containing no chemical additives before analysis by hyperspectral imaging (a total of 50 samples,

including ten controls containing no veterinary drug residues).

### 2.2 Hyperspectral imaging system

The hyperspectral imaging system (American Headwall Photonics Inc.) comprised an area array CCD, grating spectrometer, hyperspectral imaging camera lens, uniform light source, 1D electric displacement platform, USB1394 high-speed image acquisition card, high-performance computer, and related acquisition control software (Figure 1). The camera resolution was 1000 pixels $\times$ 164 pixels and the bit depth was 24 bits. Linear array scanning imaging mode was used with the spectral range of the grating spectrometer being 40-1000 nm (composed of 800 wavebands), and the light source consisted of two 200 W bromine tungsten lamps. Light sources were located on two sides of the electric displacement platform with an angle of incidence of 45°.

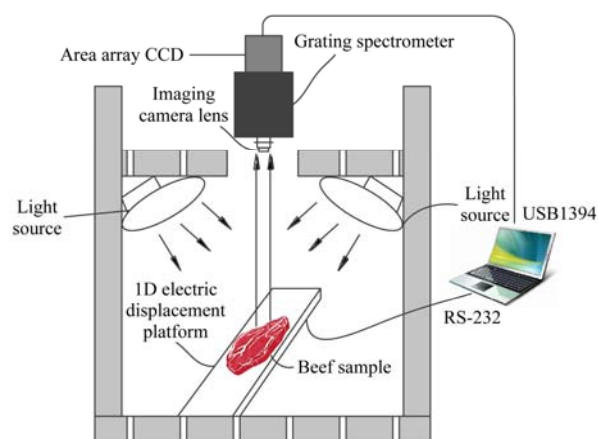


Figure 1 Schematic of the hyperspectral imaging system

### 2.3 Calibration of hyperspectral imaging system

Diffuse reflection of light is caused by the different shapes and uneven surfaces on a sample. Significant image noise is also generated under weak light intensity due to dark currents and power harmonics in the camera. It is necessary, therefore, to set the scanning parameters and calibrate the imaging system before use. After multiple debugging and effect comparisons, the exposure time was set at 0.03 s, the speed of the electric displacement platform at 3.0 mm/s, and the lens, which faced vertically downward, at 450 mm from the platform. The built-in Hyperspec 2.0 (Headwall Inc., Boston, USA) was used to calibrate the blackboard and whiteboard images as in Equation (1).

$$I = \frac{I_0 - I_{\text{dark}}}{I_{\text{white}} - I_{\text{dark}}} \quad (1)$$

where,  $I$  is the calibrated hyperspectral image;  $I_0$  is the original hyperspectral image;  $I_{\text{dark}}$  denotes the calibrated image under all-black conditions after the lid of the area array CCD is closed;  $I_{\text{white}}$  is the calibrated image under all-white conditions.

Data analysis and processing software included ENVI 5.2 (ITT Visual Information Solutions, Boulder, Co., USA), OriginPro 8 (OriginLab Co., MA, USA), Unscrambler X 10.4, Python 3.6, and TensorFlow 2.0.

### 2.4 Selection of sample regions of interest

To facilitate data acquisition, each beef sample was cut to 16 cm $\times$ 8 cm, and regions of interest (ROIs) were manually selected via ENVI software. To achieve objective and accurate data, 40 ROIs were selected for each sample (Figure 2), each being 30 pixels $\times$ 30 pixels. As the uneven surface of the samples and exudation of the drug solutions caused diffuse reflections, preference was given to dark regions with no bright spots during this selection. The mean spectral value of these 900 points was

calculated and served as a single spectral record for that sample. In total there were 2000 ( $5 \times 400$ ) spectral records.

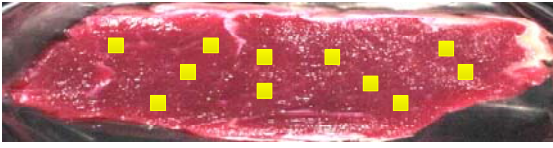


Figure 2 Schematic of representation of ROI on beef sample selection

## 2.5 Spectra preprocessing and rejection of abnormal data

Hyperspectral image acquisition can be disturbed by both sample characteristics, such as sample flatness, chromatic aberrations, and moisture content, and environmental factors. Multiple Scattering Correction (MSC) preprocessing was applied to the hyperspectral data via Unscrambler. Figure 3 illustrates 200 spectra randomly selected from the five sample groups before (Figure 3a) and after (Figure 3b) MSC preprocessing.

As shown in Figure 3b, the fluctuation and noise of the marginal spectral regions were large, which influenced subsequent modeling classification capability. Therefore, the marginal wavebands 400–414 nm and 924–1000 nm were eliminated and the 415–923 nm range (corresponding to 680 of the original 800 wavebands, specifically bands 20–700) was used for modeling.

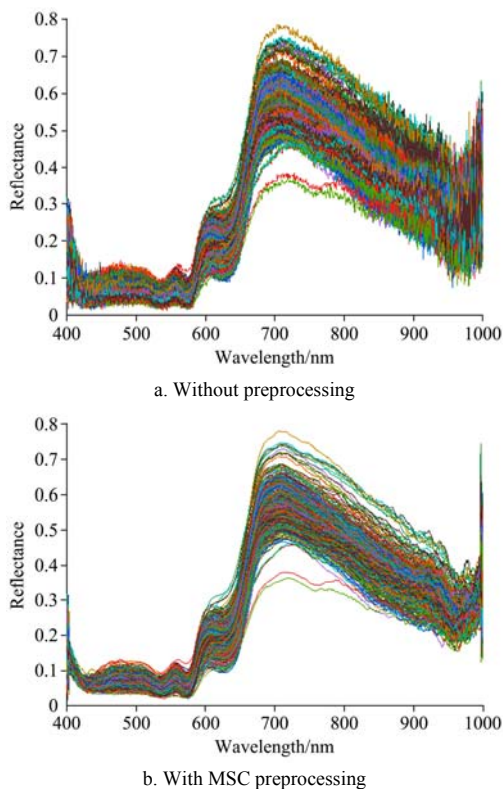


Figure 3 Spectral data preprocessing by Multiple Scattering Correction (MSC)

## 2.6 Dimension reduction of hyperspectral data

### 2.6.1 DWT dimension reduction algorithm

As a signal transform analysis method, wavelet transform (WT)<sup>[16]</sup> provides a time-frequency window that changes with frequency, enabling multiscale and multiresolution refinement of the signal (function) through the dilation and translation operations of wavelet basis function, and can extract precise (high-frequency) components and approximate (low-frequency) components simultaneously. The continuous wavelet transform is expressed as follows:

$$\begin{aligned} \text{WT}(j, k) &= \langle f(\lambda), \Psi_{j,k}(\lambda) \rangle \\ &= \frac{1}{\sqrt{|j|}} \int_{-\infty}^{+\infty} f(\lambda) \Psi\left(\frac{\lambda - k}{j}\right) d\lambda \end{aligned} \quad (2)$$

where,  $j$  is the scaling factor;  $k$  is the translation factor (both being continuous variables);  $\Psi(\lambda)$  is the wavelet basis function.

DWT<sup>[17]</sup>, a type of wavelet transform, enables signal decomposition at the discrete scale, and high- and low-frequency signal decomposition via high- and low-pass filters as Equation (3).

$$\text{DWT}(a, b) = \langle f(\lambda), \phi_{a,b}(\lambda) \rangle \quad (3)$$

where,  $a$  and  $b$  are the  $a$ -th layer decomposition and the  $b$ -th wavelet coefficient, respectively;  $\phi_{a,b}(\lambda)$  is discrete wavelet basis function.

The transformation of wavelength bands in the hyperspectral data corresponded to signal transformation in the time domain, and dimension reduction was achieved by the multiscale decomposition nature of DWT, namely, precise (high-frequency) components were omitted in the transformation of each layer and decomposed approximate (low-frequency) components were used as the modeling data. Similarly, each layer was re-transformed based on the approximate components obtained in the transformation of the previous layer, and the data dimension of each layer was successfully halved.

### 2.6.2 Dimension reduction via principal component analysis and competitive adaptive reweighted sampling

As a multivariate statistical analysis dimension reduction algorithm, Principal Component Analysis (PCA)<sup>[18]</sup> transforms the original variables into multiple independent new variables through orthogonal transformation, eliminates overlapping parts of the original data, and enables maximum characterization of the original information through the use of a few new variables.

Competitive Adaptive Reweighted Sampling (CARS)<sup>[19]</sup> selects the maximum regression coefficient via adaptive reweighted sampling<sup>[20]</sup>, exponential decay function<sup>[21]</sup>, and partial least squares (PLS) modelling<sup>[22]</sup> established through cycling. It then uses the cross-validation method to select the minimum root mean square error of cross-validation (RMSECV) subset from  $n$  PLS subset models as the optimal variable combination.

## 2.7 Classification and recognition modelling

### 2.7.1 Multilayer perceptron, support vector machine, and random forest modelling

Multilayer Perceptron (MLP)<sup>[23]</sup>, also called fully connected network or artificial neural network, is a neural network with several layers, where the neurons at different layers are mutually fully connected, and the nonlinear activation functions are added between layers to strengthen the nonlinear expression capability of the neural network so that it can fit any nonlinear function. Common activation functions include Rectified Linear Unit (ReLU), Sigmoid, Tanh, and Softmax. In this study, the MLP contained five fully connected layers as described in Table 1. During the MLP training process, learning\_rate (learning rate) was set at 0.0001, epochs (number of iterations) at 800, batch\_size (batch size) at 64, validation\_freq (validation frequency of test set) at 1, and the Adam algorithm optimizer was used.

As a supervision model based on the criterion of structural risk minimization, Support Vector Machine (SVM)<sup>[24]</sup> maps the input spectral information space into the high-dimensional characteristic space via kernel functions and facilitates multiclassification by constructing the optimal classification hyperplane. The common kernel functions include linear kernel function, polynomial kernel

function, radial basis function, and Sigmoid kernel function (Sigmoid Tanh). In this study, the optimal key parameters of the model, kernel (type of kernel function), gamma (kernel function coefficient), and C (penalty coefficient), were determined through the grid search method.

**Table 1 MLP network structure**

Layers	Number of neurons	Activation function
Fully connected layer FC1	512	ReLU
Fully connected layer FC2	512	ReLU
Fully connected layer FC3	256	ReLU
Fully connected layer FC4	128	ReLU
Output layer	5	Softmax

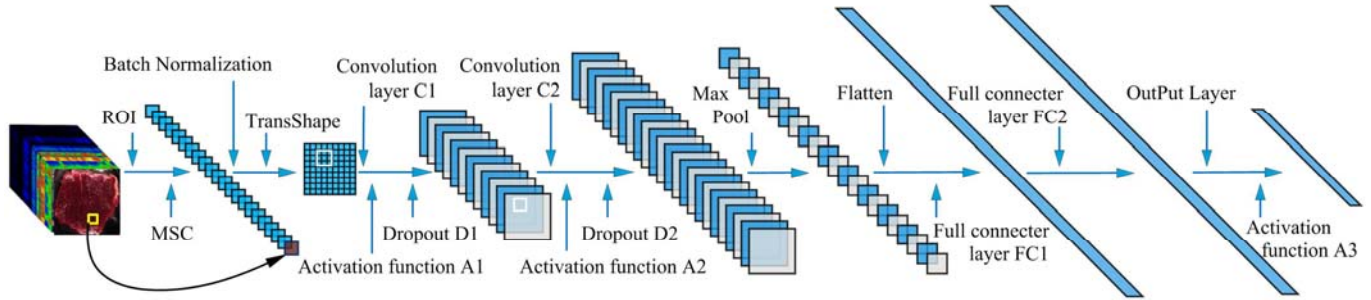
Note: MLP: Multilayer Perceptron; ReLU: Rectified Linear Unit.

As bagging integrated supervised learning method based on layered non-parameters, Random Forest (RF)<sup>[25]</sup> establishes multiple weak learning decision trees and determines the classification or regression prediction result through voting of the decision trees, or the arithmetic mean value. Its main advantages are that standard dataset processing is not needed, model training is fast, and it is not easily stuck in overfitting. The optimal key parameters of the model, max\_depth (maximum depth of decision trees), max\_features (number of maximum features), and n\_estimators (number of decision trees), were determined through the grid search method.

### 2.7.2 Convolutional neural network

CNN<sup>[26]</sup>, an MLP-based deep learning network, has been extensively applied in the field of computer vision. It generally consists of convolution layer, activation layer, dropout layer, pooling layer, and fully connected (FC) layer, where the convolution layer, which has powerful feature extraction capability, transmits the acquired features layer-by-layer so that the receptive field is continuously expanded. The activation layer generates the non-linearity via activation functions, and further improves the training speed and robustness of the neural network. The dropout layer alleviates the overfitting phenomenon by randomly removing some neurons. As a classifier with superior network learning ability, the FC layer can map the features learned by the convolution layer into the sample space. Through the orderly stacking of these layers, the CNN will be suited to the task of classification of graph-like samples by virtue of non-deformation in translation, finite sparse connection, and local weight sharing.

The CNN structure applied in this study is shown in Figure 4. Hyperspectral beef sample images comprised  $2000 \times 1 \times 1 \times 680$  after extraction of ROI, preprocessing and noise clipping. To facilitate the 2D convolution operation, four points on the extremes of the 680 wavebands were abandoned, the remaining data were transformed to  $2000 \text{ all\_batches} \times 26 \text{ width} \times 26 \text{ height} \times 1 \text{ channel}$  and the CNN structural parameters were set (Table 2).



Note: ROI: Regions of interest; MSC: Multiple Scattering Correction.

Figure 4 Convolutional neural network structure applied in this study

**Table 2 Parameter settings of the convolutional neural network structure**

Layers	Number of DWT layers	
	1-4	5-6
Convolution layer C1	$3 \times 3 \times 32$	$3 \times 3 \times 32$
Activation layer A1	ReLU	ReLU
Dropout layer D1	0.3	0.3
Convolution layer C2	$5 \times 5 \times 64$	$3 \times 3 \times 64$
Regularization L <sub>2</sub> <sub>1</sub>	0.001	0.001
Activation layer A2	ReLU	ReLU
Dropout layer D2	0.5	0.5
Pooling layer MaxPooling	$2 \times 2 \times 2$	$2 \times 2 \times 2$
Fully connected layer FC1	256	256
Fully connected layer FC2	256	256
Regularization L <sub>2</sub> <sub>2</sub>	0.001	0.001
Output layer	5	5
Activation layer A3	Softmax	Softmax

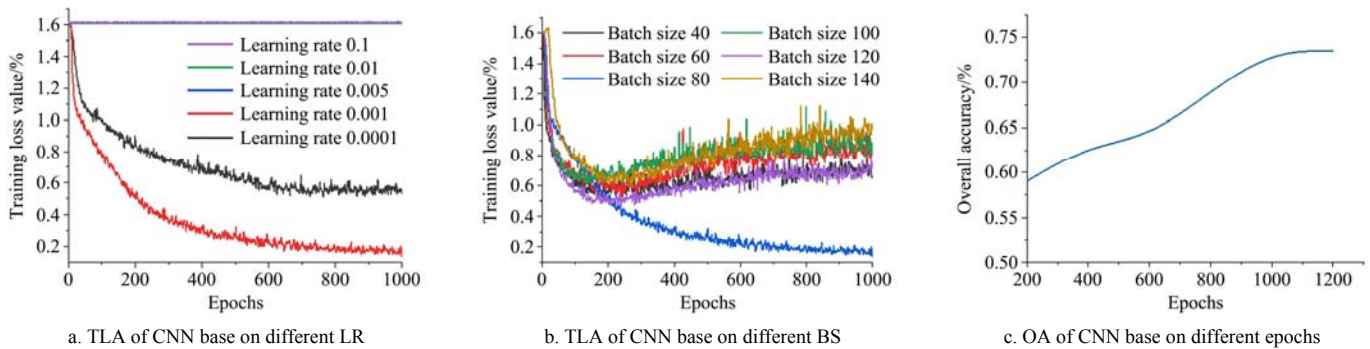
Note: Convolution layer  $n \times n \times m$  represents  $m$  filters with convolution kernel size  $n \times n$ . Pooling layer  $n \times n \times m$  represents pooling kernel size  $n \times n$  and step size  $m$ .

Classification accuracy of the recognition model was directly affected by CNN hyperparameter selection. To determine the key hyperparameters (learning\_rate, batch\_size, and epochs), multi-parameter training was performed multiple times on the hyperspectral data (Figure 5). Training loss remained at 1.6%

when learning\_rate was 0.005, 0.01, and 0.1, indicating that the model could not be converged due to the excessive learning\_rate (Figure 5a). Model convergence rate was high when learning\_rate was 0.001. The decline in training loss value was limited and oscillated significantly when batch\_size was 40, 60, 100, 120, and 240, showing that the model was unstable with weak generalization ability (Figure 5b). When batch\_size was 80, training loss value fell continuously, with minor fluctuations. The overall accuracy (OA) rose with the increasing number of epochs, reaching a peak after 1000 epochs (Figure 5c). Thus, the following hyperparameters were selected: learning\_rate=0.001, batch\_size= 80, epochs=1000, and the Adam algorithm was adopted as the optimizer. Since smaller sample sizes tend to result in model overfitting, the model's generalization ability was strengthened using dropout and L2 regularization techniques.

### 2.8 Model evaluation

Confusion matrix (CM)<sup>[27]</sup>, also called error matrix, expresses the accuracy of a classification algorithm in  $N \times N$  matrix form. Table 3 presents the typical CM form ( $N=2$ ), where TP means true positive, expressing the number of sample 1 correctly classified, FN (false negative) is the number of sample 1 mistakenly classified as sample 2, FP (false positive) is the number of sample 2 mistakenly classified as sample 1, and TN (true negative) is the number of sample 2 correctly classified.



a. TLA of CNN base on different LR  
 b. TLA of CNN base on different BS  
 c. OA of CNN base on different epochs  
 Note: TLA: Training loss value; LR: Learning rate; BS: Batch size; OA: Overall accuracy.

Figure 5 CNN hyperparameter selection process

**Table 3 Confusion matrix**

True value/Predicted value	Sample 1	Sample 2
Sample 1	TP	FN
Sample 2	FP	TN

Overall accuracy (OA)<sup>[28]</sup> is the ratio of the number of correctly classified samples to the total number of samples tested, and indicates the probability that an individual sample will be correctly classified by a test:

$$OA = \frac{TP + TN}{TP + FN + FP + TN} \quad (4)$$

Kappa coefficient is the ratio of classification results to error decrement generated by completely random classification. Used to measure classification accuracy, it effectively solves the problem that OA cannot distinguish the accuracy unevenness of different sample types, the calculation of Kappa coefficient is as follows:

$$Kappa = \frac{OA - P_e}{1 - P_e} \quad (5)$$

where,  $P_e = \frac{(TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN)}{N \times N}$ .

### 3 Results and discussion

#### 3.1 Spectral analysis

To visualize the spectral differences between the groups of samples, the 400 spectra in each group were averaged (Figure 6).

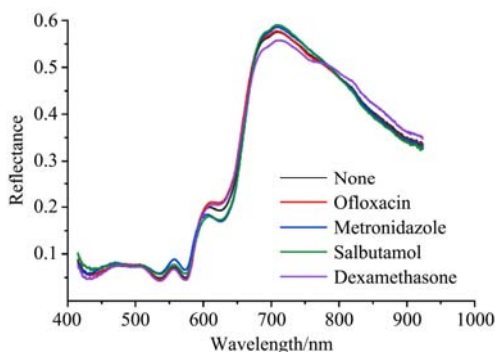


Figure 6 Averaged spectra of beef samples

Figure 6 shows that spectral reflectance was low in the visible light range (415-570 nm) and high in the shortwave NIR range (610-750 nm). As the beef sample surface was red due to the presence of myoglobin and hemoglobin, spectral absorption was high and reflectance was low in the complementary blue-green region. In the NIR region, spectral absorption was associated with energy absorption and the transition of groups between energy

levels in the drug molecules. The reflectance spectrum of metronidazole-treated beef was different from the salbutamol spectrum between 415 and 600 nm but they overlapped in the range 640-700 nm, probably due to the -OH and -CH groups present in both drug compounds. The pattern of the curves for the residue-free, ofloxacin-treated and dexamethasone-treated samples differed in the 600-650 nm and 690-720 nm ranges, probably due to the different effects of ofloxacin and dexamethasone on the conformation of bovine serum albumin, resulting in a large difference in spectral absorption rates. In conclusion, these provided a technical basis for identifying these veterinary drug residues in beef by spectral means.

#### 3.2 Analysis of dimension reduction algorithms

##### 3.2.1 Dimension reduction via DWT

The Haar wavelet was used to compare dimension reduction effects. DWT was applied to the averaged spectra, after which high-frequency noise components were excised and low-frequency components retained. The data dimensions obtained through six layers of DWT dimension reduction are shown in Table 4. Data dimensions progressively decreased by  $2^{-m}$  ( $m$ =number of DWT layers) with each DWT layer.

**Table 4 DWT data dimensions**

DWT layers	Data dimensions
Original data	680
One-layer DWT	340
Two-layer DWT	170
Three-layer DWT	85
Four-layer DWT	43
Five-layer DWT	22
Six-layer DWT	11

Note: DWT: Discrete Wavelet Transform.

Figure 7 illustrates the gradually increasing differences between the original spectra and spectra following successive DWT layers. Figure 7a shows the low-frequency component spectra after one-layer DWT. These transformed spectra were very similar to the untreated spectra in Figure 6 even though only half the original data were utilized. The spectra following six-layer DWT (Figure 7f) used only eleven data dimensions but the relative positional relationships of the spectra at key nodes were still evident.

##### 3.2.2 Dimension reduction via PCA and CARS

PCA was also used for dimension reduction of the hyperspectral data. The cumulative contribution of the first 11, 22, 43, 85, 170, and 340 principal components are displayed in Table 5.

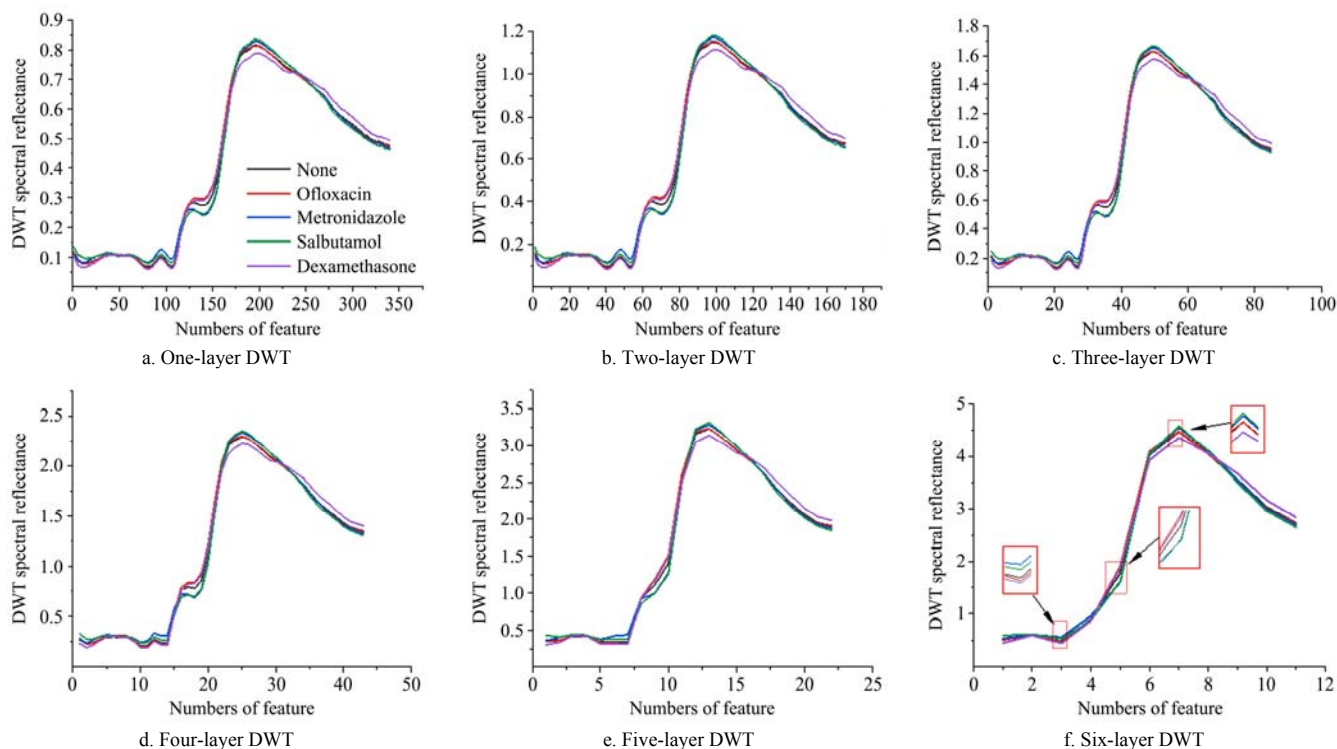


Figure 7 Low-frequency component spectra after DWT based on Haar wavelet basis function

Table 5 Cumulative contribution rates of PCA

Total principal components	Cumulative contribution rate/%
PC11	77.05
PC22	79.78
PC43	83.66
PC85	88.75
PC170	94.24
PC340	98.23

Full-sample dimension reduction was implemented using CARS with sampling frequency, threshold value, number of Monte Carlo sampling times, and number of cross-validation groups being 0.8, 0.8, 100, and 10, respectively. Data were centralized and the screening process is shown in Figure 8. As the number of sampling runs increased, the number of characteristic wavelengths identified reduced until the curve plateaued (Figure 8a). As sampling runs increased, RMSECV declined before increasing sharply, indicating that irrelevant variables were eliminated in the early runs with the lowest error achieved after 35 runs (Figure 8b). Figure 8c illustrates the changing regression coefficients of the characteristic wavelengths, the vertical line indicating the run with the minimum RMSECV. A total of 89 characteristic wavelengths were identified by CARS, accounting for 1.12% of all the wavebands.

3.3 Evaluation of recognition models

3.3.1 Evaluation of DWT-based dimension reduction models using various wavelet bases

The dimension reduction effect varies depending on the wavelet basis function. To achieve optimum classification and recognition performance, Haar wavelet, db3 in Daubechies wavelet family, bior1.5 in Biorthogonal wavelet family, sym5 in Symlets wavelet family, and rbio1.3 in ReverseBior wavelet family were evaluated for the wavelet transform of data from 1 to 6 layers, before recognition models were constructed. The recognition and classification modeling process is illustrated in Figure 9. The model training set comprised 75% of the total beef samples while 25% acted as the prediction set.

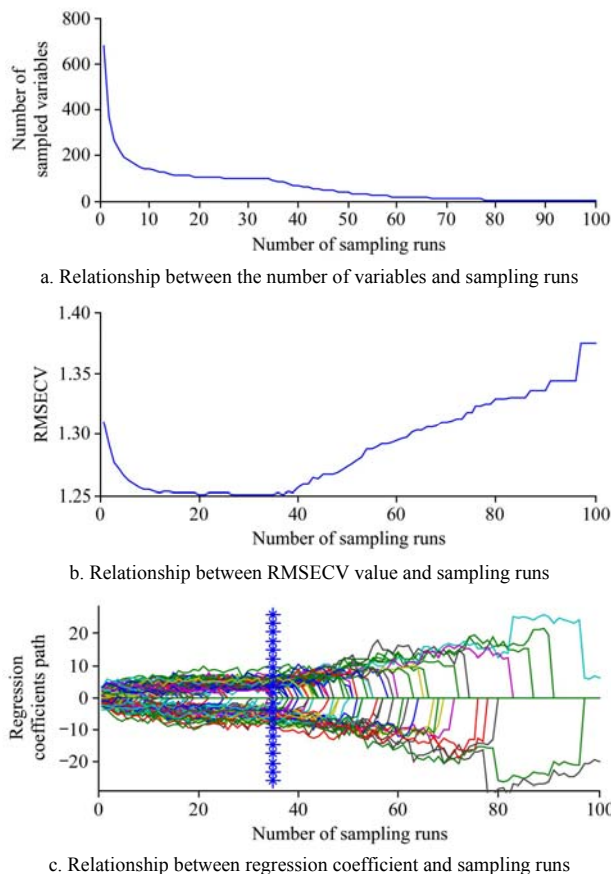
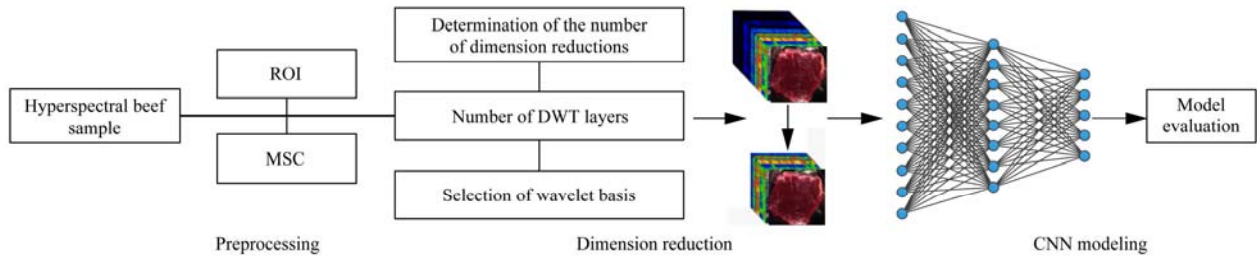


Figure 8 Screening process for characteristic wavelengths using competitive adaptive reweighted sampling

The OA of the four algorithms is displayed in Figure 10a-10d. The OA curves for CNN and MLP were smoother than for RF and SVM, showing the stronger learning and prediction capabilities of the CNN and MLP algorithms. OA fluctuated considerably as the number of DWT layers increased, with Haar wavelet-based dimension reduction tending to give the highest OA. Excluding the db3

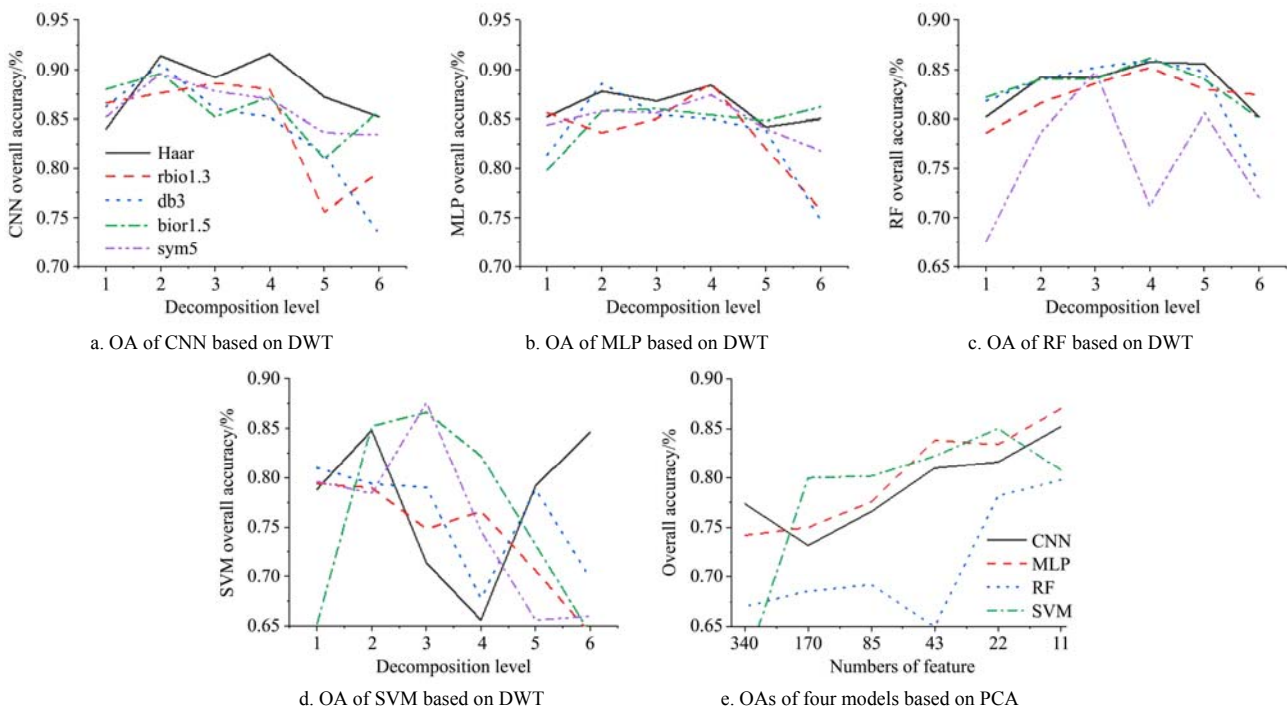
wavelet basis, the OA of CNN-based modeling after dimension reduction fluctuated within 81.4%-91.6%, the highest OA being reached after the 4th layer transform using Haar (Figure 10a). The MLP algorithm reached its highest OA (88.6%) after the 2nd

layer transform using db3 (Figure 10b). The RF algorithm reached its optimal OA (86.2%) after the 4th layer transform using bior1.5 (Figure 10c). The SVM algorithm reached its optimal OA (84.6%) after the 6th layer transform using Haar (Figure 10d).



Note: ROI: Regions of interest; MSC: Multiple Scattering Correction.

Figure 9 Deep learning classification flowchart based on hyperspectral DWT



Note: OA: Overall accuracy; RF: Random Forest; SVM: Support Vector Machine; PCA: Principal component analysis.

Figure 10 Modeling based on DWT and PCA

### 3.3.2 Evaluation of PCA-based dimension reduction models

To verify the reliability of the proposed DWT dimension reduction algorithm, the first 340, 170, 85, 43, 22, and 11 principal components (corresponding to the data dimensions after one to six layers of DWT) after PCA-based dimension reduction were selected for modeling. The OA of the four algorithms is shown in Figure 10e. As the data dimensions decreased, the OA of the algorithms generally increased, indicating that interfering factors were gradually excluded and the prediction accuracy was significantly improved. CNN and MLP yielded higher OA than RF and SVM, and reached the highest OA (85.2% and 87.0%) when the first 11 principal components were extracted.

### 3.3.3 Transverse evaluation of dimension reduction and modeling algorithms

The OA and Kappa coefficients of the three data dimension reduction algorithms and four modeling algorithms are listed in Table 6. Classification models without a dimension reduction process had a low OA and a long prediction time. The CARS algorithm yielded the lowest classification OA. The OA and Kappa coefficient of the PCA algorithm were higher than CARS. After DWT-based dimension reduction, the OA of the

classification models using CNN, MLP, SVM, and RF were 91.6%, 88.6%, 87.6%, and 86.2%, respectively, and the Kappa coefficients were 0.89, 0.86, 0.85, and 0.83, respectively.

The OA values of the various prediction algorithms are shown in Table 7. The OA for ofloxacin prediction was the highest at 98.0%, which may be the conformation of bovine albumin was changed by the drug, leading to the increase in feature discrimination of spectral reflection peak-valley shift. The OA for metronidazole and salbutamol prediction was generally low (83.0% and 91.0%, respectively), possibly because these drugs contain the same hydroxyl group, limiting their spectral discrimination. Samples containing metronidazole, ofloxacin, and dexamethasone, and those without drug residues all reached their highest OA under the DWT algorithm, indicating that during data dimension reduction the DWT algorithm not only retains the original spectrum shape but can also restore the relative spatial positions of the spectra and improve classification and recognition accuracy. Highest OA for metronidazole, dexamethasone, and residue-free samples was under the DWT (Haar-4)-CNN combination, demonstrating that the DWT-CNN algorithm has favorable dimension reduction, feature extraction, and classification and recognition capabilities.

**Table 6 Overall accuracy and Kappa coefficients of the various algorithms**

Model	Dimension reduction algorithm											
	None			CARS			PCA			DWT		
	OA/%	Kappa	Time/ms	OA/%	Kappa	Time/ms	OA/%	Kappa	Time/ms	OA/%	Kappa	Time/ms
RF	80.4	0.75	157.70	72.0	0.65	48.87	79.8	0.75	247.35	86.2	0.83	140.60
SVM	78.8	0.73	507.67	72.8	0.66	15.96	85.0	0.81	23.96	87.6	0.85	57.85
MLP	83.6	0.79	71.59	78.2	0.72	75.77	87.0	0.84	95.75	88.6	0.86	70.67
CNN	73.6	0.67	180.81	76.6	0.70	364.24	85.2	0.81	99.74	<b>91.6</b>	<b>0.89</b>	87.16

Note: RF: Random Forest; SVM: Support Vector Machine. RF reaches optimal values under PCA (11) and DWT (bior1.5-4); SVM under PCA (22) and DWT (sym5-3); MLP under PCA (11) and DWT (db3-2); and CNN under PCA (22) and DWT (Haar-4). PCA ( $n$ ) denotes the modeling of the first  $n$  principal components, and DWT (W- $g$ ) represents the modeling of the dimension reduction algorithm based on the discrete transform of  $g$ -th layer W wavelet basis. All algorithms were independently run on a Dell T3431 tower-type graphic workstation (Intel® Core i7-9700 CPU 3.00 GHz, RAM 2.60 GHz 16 GB, GPU GeForce GTX 1660 Super).

**Table 7 Overall accuracy of the various prediction algorithms**

Dimension reduction algorithm	Model	Veterinary drug classification/%				
		Metronidazole	Ofloxacin	Salbutamol	Dexamethasone	None
None	RF	62.0	91.0	80.0	93.0	76.0
CARS		49.0	83.0	77.0	83.0	68.0
PCA		67.0	83.0	80.0	87.0	81.0
DWT		70.0	92.0	89.0	89.0	91.0
None	SVM	70.0	92.0	76.0	79.0	77.0
CARS		57.0	80.0	75.0	77.0	75.0
PCA		71.0	95.0	<b>91.0</b>	87.0	81.0
DWT		81.0	93.0	84.0	91.0	89.0
None	MLP	63.0	95.0	81.0	89.0	90.0
CARS		62.0	80.0	87.0	82.0	80.0
PCA		79.0	97.0	88.0	84.0	87.0
DWT		77.0	<b>98.0</b>	90.0	93.0	85.0
None	CNN	55.0	79.0	71.0	92.0	71.0
CARS		68.0	89.0	72.0	78.0	76.0
PCA		75.0	91.0	87.0	86.0	88.0
DWT		<b>83.0</b>	96.0	90.0	<b>96.0</b>	<b>93.0</b>

Note: CARS: Competitive adaptive reweighted sampling; PCA: Principal component analysis; DWT: Discrete Wavelet Transform; RF: Random Forest; SVM: Support Vector Machine; MLP: Multilayer Perceptron; CNN: Convolutional Neural Network.

## 4 Conclusions

In this study, a hyperspectral DWT algorithm was combined with CNN to create a deep learning classification algorithm. Classification and recognition were performed on hyperspectral datasets from five groups of beef samples containing veterinary drug residues.

1) Hyperspectral data subjected to dimension reduction showed improved OA and Kappa coefficients for an independent prediction sample set. The time spent in classification was also shortened considerably, demonstrating that dimension reduction algorithms can effectively eliminate redundant data, reduce data dimensions, alleviate the Hughes phenomenon, and improve the accuracy of modeling algorithms.

2) The DWT algorithm based on the Haar wavelet basis function is capable of filtering out high-frequency interference information through multilayer low-loss filters and achieving data dimension reduction. Furthermore, compared with PCA and CARS, it not only effectively retains the original spectrum shape but also restores the relative spatial positions of the spectra and improves model classification and recognition accuracy.

3) CNN has strong feature extraction and learning capabilities. The OA of the model based on DWT and CNN was 91.6%, which exceeded models based on MLP, SVM, and RF by 3.0%, 4.0%, and 2.4%, respectively. The advantages of DWT and deep learning have been demonstrated in this study which proposes a new method

for the fast, non-destructive detection and recognition of veterinary drug residues in beef.

## Acknowledgements

The authors acknowledge all anonymous reviewers for their valuable and detailed comments. This work was financially supported by China Central Government to Support the Reform and Development Fund of Heilongjiang Local Universities (Grant No. 2020GSP15).

## [References]

- [1] Liu Y F, Shen Q. Commonality and difference analysis of the effects of poultry and livestock products on human health. *Journal of Shaanxi Normal University(Natural Science Edition)*, 2020; 48(1): 114–122. (in Chinese)
- [2] Hou M L. A brief analysis of the classification and detection techniques of veterinary drug residues in animal derived food. *Modern Food*, 2020; 3: 130–131, 134. (in Chinese)
- [3] Lozano A, Hernando M D, Uclés S, Hakme E, Fernandez-Alba A R. Identification and measurement of veterinary drug residues in beehive products. *Food Chemistry*, 2019; 274: 61–70. doi: 10.1016/j.foodchem.2018.08.055.
- [4] Aman I M, Ahmed H F, Mostafa N Y, Kitada Y, Kar G. Detection of tetracycline veterinary drug residues in Egyptian poultry meat by high performance liquid chromatography. *J Vet Med Allied Sci*, 2017; 1(1): 52–58.
- [5] Masiá A, Suarez-Varela M M, Llopis-Gonzalez A, Pico Y. Determination of pesticides and veterinary drug residues in food by liquid chromatography-mass spectrometry: A review. *Analytica Chimica Acta*,



- 2016; 936: 40–61. doi: 10.1016/j.aca.2016.07.023.
- [6] Li N, Zhang Y T, Liu L, Shao H, Li H, Li J, Guo Y. Simultaneous determination of 4 kinds of 29 banned and restricted veterinary drugs in animal-derived foods by ultra performance liquid chromatography-tandem mass spectrometry and modified QUECHERS for sample preparation. *Chinese Journal of Chromatography*, 2014; 32(12): 1313–1319. (in Chinese)
- [7] Ji H X, Xia C X, Xu J J, Wu X X, Qiao L, Zhang C. A highly sensitive immunoassay of pesticide and veterinary drug residues in food by tandem conjugation of bi-functional mesoporous silica nanospheres. *Analyst*, 2020; 145: 2226–2232. doi: 10.1039/d1an00929j.
- [8] Zhao X L, Xie S Y, Chen Y, Wan K. Application of immunoanalyt analysis technology in the detection of agricultural and veterinary drug residues in agricultural products. *China Inspection Body & Laboratory*, 2020; 28(3): 18–20. (in Chinese)
- [9] Yu W B, Li Y, Xie B, Ma M F, Chen C C, Li C L, et al. An aggregation-induced emission-based indirect competitive immunoassay for fluorescence “turn-on” detection of drug residues in foodstuffs. *Frontiers in Chemistry*, 2019; 7: 228. doi: 10.3389/fchem.2019.00228.
- [10] Liu C L, Wang C, Guo Y. Detection of veterinary drug residues in animal derived food. *Agriculture and Technology*, 2019; 39(15): 42–43. (in Chinese)
- [11] Tao J J, Liu M H, Yuan H C, Zhao J Y. Surface-enhanced raman spectroscopy for rapid detection of testosterone propionate residues in duck meat. *Chinese Journal of Analysis Laboratory*, 2019; 38(2): 152–156. (in Chinese)
- [12] Guo L. Preparation and application of surface molecular imprinted polymer of sulfadimethoxine. Master dissertation. Shenyang Agricultural University, 2019; 54p. (in Chinese)
- [13] Sun J, Cong S, Mao H, Wu X, Yang N. Quantitative detection of mixed pesticide residue of lettuce leaves based on hyperspectral technique. *Journal of Food Process Engineering*, 2018; 41(2): e12654. doi: 10.1111/jfpe.12654.
- [14] Ji H, Ren Z, Rao Z. Identification of pesticide residue types in spinach leaves based on hyperspectral imaging. *Chinese Journal of Luminescence*, 2018; 39(12): 1778–1784. (in Chinese)
- [15] Gao Z M, Shao Y Y, Xuan G T, Wang Y X, Liu Y, Han X. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artificial Intelligence in Agriculture*, 2020; 4(1): 31–38. doi: 10.1016/j.aiia.2020.04.003.
- [16] Sifuzzaman M, Islam M R, Ali M Z. Application of wavelet transform and its advantages compared to Fourier transform. *Journal of Hsical Sciences*, 2009; 13: 121–134.
- [17] Xu T Y, Guo Z H, Yu F H, Xu B, Feng S. Genetic algorithm combined with extreme learning machine to diagnose nitrogen deficiency in rice in cold region. *Transactions of the CSAE*, 2020; 36(2): 209–218. (in Chinese)
- [18] Gao Z M, Zhao Y R, Khot L R, Hoheisel G, Zhang Q. Optical sensing for early spring freeze related blueberry bud damage detection: Hyperspectral imaging for salient spectral wavelengths identification. *Computers and Electronics in Agriculture*, 2019; 167: 105025. doi: 10.1016/j.compag.2019.105025.
- [19] Yuan Z R, Wei L F, Zhang Y X, Yu M, Yan X. Hyperspectral inversion and analysis of heavy metal arsenic content in farmland soil based on optimizing CARS combined with PSO-SVM algorithm. *Spectroscopy and Spectral Analysis*, 2020; 40(2): 567–573. doi: 10.3964/j.issn.1000-0593(2020)02-0567-07. (in Chinese)
- [20] Peng H G, Jin Y, Zhan Y G, Chen Y, Feng X, Qian F, et al. Quantitative determination of hydrolytic nitrogen content in soil by near infrared spectroscopy combined with competitive adaptive reweighted sampling variable selection algorithm. *Journal of Instrumental Analysis*, 2020; 39(10): 1305–1310. (in Chinese)
- [21] Okayama T, Shintaku Y, Katsuura E. New conformal map for the Sinc approximation for exponentially decaying functions over the semi-infinite interval. *Journal of Computational and Applied Mathematics*, 2020; 373: 112358. doi: 10.1016/j.cam.2019.112358.
- [22] Wang G F, Zhang Y X, Xu W L, Zhou W, Wu H, Xu Z, et al. Estimation of phytoplankton pigment concentration in South China sea from hyperspectral absorption data. *Acta Optica Sinica*, 2021; 41(6): 14–28. (in Chinese)
- [23] Chen M, Hu X Q, Lu W K. Application of artificial neural network in agricultural disease prediction. *Modern Agricultural Science and Technology*, 2020; 11(21): 136–140. (in Chinese)
- [24] Deng L F, Zhang F, Qi Y X, Yuan J. Identification of sodium ion spectral characteristics of halophytes based on parameter optimized SVM method. *Spectroscopy and Spectral Analysis*, 2020; 40(1): 247–254. (in Chinese)
- [25] Li G, Gao X, Xiao N, Xiao Y. Estimation soil organic matter contents with hyperspectra based on SCARS and RF algorithms. *Chinese Journal of Luminescence*, 2019; 40: 1030–1039. (in Chinese)
- [26] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012; 25: 1097–1105.
- [27] Wu H M, Xu J Q, Liu H L, Liu C, Zhang D, Chen K. Identification of water injection meat based on hyperspectral technique and spectrum characteristics. *Transactions of the Chinese Society for Agricultural Machinery*, 2019; 50(11): 367–372. (in Chinese)
- [28] Shi Y, Ma D H, Lu J, Li J, Shi J. Hyperspectral image classification based on manifold spectral dimensionality reduction and deep learning method. *Transactions of the CSAE*, 2020; 36(6): 151–160. (in Chinese)