Check for updates

Received: 3 July 2020 | Revised: 10 September 2020 | Accepted: 3 October 2020

DOI: 10.1111/jfs.12866

ORIGINAL ARTICLE

Journal of Food Safety WILEY

Detection for lead pollution level of lettuce leaves based on deep belief network combined with hyperspectral image technology

Jun Sun | Yan Cao | Xin Zhou | Minmin Wu | Yidan Sun | Yinghui Hu

School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, China

Correspondence
Jun Sun, School of Electrical and Information Engineering of Jiangsu University, Zhenjiang 212013, China.
Email: sun2000jun@sina.com

Funding information
Jiangsu Agriculture Science and Technology Innovation Fund, Grant/Award Number: CX (19)3089; Jiangsu University Student Practice Innovation Training Program; National natural science funds projects, Grant/Award Number: 31971788; Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD); Project of Agricultural Equipment Department of Jiangsu University, Grant/Award Number: 4121680001; Synergistic Innovation Center of Jiangsu Modern Agricultural Equipment and Technology, Grant/Award Number: 4091600030

Abstract

Fast detection for heavy metal in vegetables is one of the most important steps to ensure the food safety. A novel method to identify lead pollution levels of lettuce based on hyperspectral image technology was proposed in this study. Firstly, hyperspectral images of lettuce samples cultivated under four lead stress levels (0 mg/L, 50 mg/L, 100 mg/L and 200 mg/L) were collected using hyperspectral image system. Then, a total of 240 spectra were calculated from region of interest (ROI) in the range of 478–978 nm covering 399 bands. Besides, chemical test showed that the excessive level of lead residues content in lettuce leaves were none, slight, moderate and severe. Moreover, conventional models and deep belief network (DBN) were established to determine the best identification model. The discriminant DBN model reached the highest accuracy with the training set of 100% and test set of 96.67%. Finally, t-distribution stochastic neighbor embedding (t-SNE) was successful to visualize the feature values in DBN's last hidden layer. This study indicates that it is viable to detect the lead pollution levels of lettuce leaves based on hyperspectral image technology coupled with DBN discriminant model.

1 | INTRODUCTION

Nowadays, the environment pollution has become a serious social issue, and heavy metal pollution has been paid more public attention (Bhat, Kiran, Arun, & Karim, 2010). The distribution of heavy metal pollution in central and south China is quite wide due to the exploitation of mineral resources. Lead pollution is the most prominent problem (Wei, Baiming, & Lin, 2013). According to survey reports, the lead content in many regional vegetables has exceeded the safety standard (Liu, Guo, & Guo, 2017; Yuping, Xiaohui, & Yan, 2017). Study has shown that the absorption of Pb by plants can damage the chloroplast structure and reduce the activity of chlorophyll enzyme, thus causing the decrease of chlorophyll content in leaves. In addition, Pb can inhibit stomatal opening, damage cell structure and reduce the leaf expansion, resulting in plant growth restriction and yield loss (Yu et al., 2018). Excessive amount of lead in the environment can be absorbed by crops, and finally enters the human body through food chain. Lead will damage the nerves, cause hypertension and anemia, affect intellectual development, and reduce immunity even at a low concentration, especially for children (Norton et al., 2015; Pourrut, Shahid, Dumat, Winterton, & Pinelli, 2011).

Lettuce is a common leafy vegetable widely cultivated in the world (Sun et al., 2017a). However, researches show that lettuce which grows in sludge-treated soil will accumulate lead in its leaves gradually. This characteristic means that lettuce can be used as the indicator crop to identify the pollution level of lead contamination in the environment. Lettuce is one of the main sources of trace metal intake through the diet (Dala-Paula et al., 2018). Therefore, the real-time and fast detection for the lead pollution levels of lettuce is essential to ensure the food safety, supervise the growth environment of vegetables, and reduce the risk of lead poisoning.

Conventional methods to detect heavy metals including atomic fluorescence spectrometry (Wen, Wu, Chen, & Hou, 2009), atomic absorption spectrometry (Jahromi, Bidari, Assadi, Hosseini, &

 © 2020 Wiley Periodicals LLC

Jamali, 2007), ultraviolet visible spectrophotometry (Lonni, Scarminio, Silva, & Ferreira, 2005) have high accuracy and sensitivity, but depend on complex and time-consuming chemical processing. In general, these methods cannot meet the requirement for quick and real-time detection of heavy metals of vegetables in actual agricultural production.

Hyperspectral image technology is a popular nondestructive testing technology in recent years (Khulal, Zhao, Hu, & Chen, 2016; Sunli et al., 2018; Wu, Fu, Tian, Wu, & Sun, 2017), which is an effective method to analyze the change information of spectrum for detection of heavy metals (Zhou et al., 2020). It has been applied in the researches on heavy metal in agricultural products. Zhu, Qu, Liu, and Chen (2014) studied the spectral changes in copper-polluted leaves using seven spectral signatures coupled with the spectral angle method, and the results indicated that leaf structural parameter could be used as an indicator of copper pollution. Gu, Li, Gao, and Wei (2015) used six sensitive parameters and regression models to detect the cadmium content in the leaves of *Brassica rapa* chinesis with the highest R^2 of 0.811. Sun et al. (2019) combined transform and least square support vector machine regression to estimate the cadmium content in tomato leaves with the results of R_{cv}^2= 0.9357, RMSECV = 0.1455 mg/kg and RPD = 3.081. The above researches were based on two steps including variable selection and model establishment to deal with high-dimensional hyperpsectral data. But these models have a few limits in different application backgrounds, such as how to choose the best variable selection algorithm and establish the best model (Zhou et al., 2018).

Deep belief network (DBN), a typical deep learning method (Hinton & Salakhutdinov, 2006), has the strong ability to extract abstract features from the input data layer by layer. It has achieved good results in many regression and classification tasks. Recently, it has been applied in the field of spectroscopy such as predicting the germination rate of rice seed (Lu, Guo, Zhang, & Wang, 2018), classifying mildew disease degrees of apples (Zhaoyong et al., 2017), and identifying green potatoes (Li, Ku, Yan, Xu, & Jin, 2016). Therefore, it is feasible to detect the lead pollution levels of lettuce by hyperspectral image technology combined with deep belief network.

The specific objectives in this study were listed as follows: (a) to explore and prove that the identification of lead pollution levels of lettuces can be achieved by using DBN combined with near infrared hyperspectral image technology; (b) to compare the results of deep belief network and conventional modeling algorithms and determine the best modeling method; (c) to visualize the original spectral data, feature selection data, feature values of DBN's last hidden layer, and intuitively demonstrate the effectiveness of deep belief network.

2 | MATERIAL AND METHODS

2.1 | Sample cultivation

The experiment in this study was conducted in a greenhouse at laboratory venlo of Modern Agricultural Equipment in Jiangsu University, Zhenjiang of China. The lettuce seed (cold butter lettuce, Shanghai, China) germination was carried out under hydroponics cultivation. According to the risk control standard for soil contamination of agricultural land in China (GB15618-2018) and with reference to one of the national research subjects in the seventh five-year science and technology plan "Study on soil environmental background values in China" (Wei & Chen, 1991), Pb solution with concentration of 0 mg/L, 50 mg/L, 100 mg/L, and 200 mg/L was added into the corresponding nutrient solution to obtain four lettuce samples with different Pb stress gradients. The corresponding mass of $Pb(CH_3COO)_2$ was weighed accurately to prepare the corresponding concentration of Pb solution. Group A, group B, group C and group D were set to correspond to four different Pb stress gradients, respectively. Each group had 60 plants, and group A was the control group. When lettuces put forth 4–5 leaves, they were transplanted into pearlite plot. Lettuces in each group were watered with the corresponding mixed solution of the same volume for 7 days. Afterwards, the non-polluted nutrient solution was used to water the lettuces until they grew to the rosette stage to carry out leaf sampling. One middle lettuce leaf at the same position was picked from each plant. A total of 240 leaves were collected and sealed in plastic bags with corresponding labels and placed in a professional plant fresh-keeping box (Temperature – 5°C). Finally, lettuce samples were subsequently sent to the laboratory for the acquisition of hyperspectral images and chemical test of lead content.

2.2 | Acquisition of hyperspectral image and data extraction

Hyperspectral images of lettuce samples were obtained by hyperspectral image system. The laboratory hyperspectral image system consists of spectrometer (V10E, SPECIM, Finland) with the wavelength range of 400 ~ 1,000 nm and wavelength resolution of 2.8 nm, CCD camera (Zyla4.4, Andor, UK) with the resolution of 2048 × 478, halogen light source (3900E, ISUZU OPTICS, Taiwan), motor stage (MSI300, Taiwan) with movement accuracy of 10 um, a computer and a dark box. The composition diagram of hyperspectral imaging system was shown in Figure 1. Experimental lettuce sample was put flat in the center of the displacement Table (25 × 25 cm) for acquisition of hyperspectral images. The sample exposure time was 17 ms, the move speed of motor stage was 4.1306 mm/s. The acquired hyperspectral images were processed by black and white correction to remove interference information from dark current and light source (Sun, Cong, Mao, Wu, & Yang, 2018).

After image correction, ENVI 4.5 software (Research system, Inc, Boulder, Co) was used to extract hyperspectral data. The whole area of lettuce leaf was chosen as the region of interest (ROI), which was separated from the background by threshold segmentation algorithm. Then, a total of 240 mean hyperspectral data was calculated from the ROI to represent all the test lettuce samples. Due to the limit of hardware, the hyperspectral data was greatly influenced by noise in the first 61 bands and last 18 bands. The specific progress of extracting hyperspectral image can be seen in our formal study (Sun et al., 2019). Hence, the final wavelength range of spectral data was from 478 nm to 978 nm covering 399 bands. The general sample set was divided according to the ratio of 3:1, containing 180 samples in training set and 60 samples in test set.

17454565, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfs.12866 by Universita Del Salento Ser Inf Bibliot "La Stecca", Wiley Online Library on [30/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

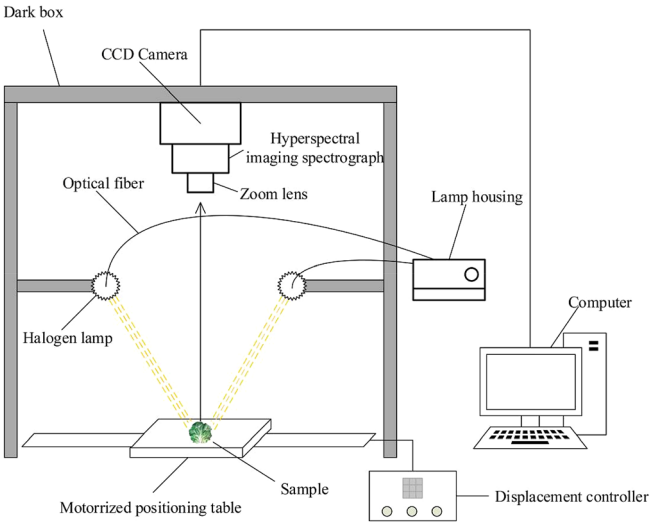


FIGURE 1 Hyperspectral imaging system

2.3 | Determination of Pb content in lettuce leaf

After the hyperspectral image of lettuce leaf was collected, the Pb content in lettuce leaf was determined by graphite furnace atomic absorption spectrometry (GFAAS) method according to the national standard for detecting Pb content in food (GB 5009.12–2010). The formula for calculating Pb content is as follows:

$$c = \frac{(c_1 - c_0) \times v}{m \times 1000}$$

Where c (mg/kg) is the content of Pb in the lettuce sample. c_0 and c_1 (μg/ml) are the content of Pb in blank solution and the sample digestion solution, respectively. v (ml) is the total volume of the digested liquid. m (g) is the mass of the sample. In order to avoid accidental factors, the repeated test on each sample was carried out. The absolute difference among the three independent measurements does not exceed 0.01% of the arithmetic mean of the two measurements. The Pb content of each sample is the average of three measurements.

2.4 | Multivariate data analysis

In this study, conventional classification modeling methods and DBN were used to establish discriminant models. Besides, t-SNE was used to visualize high-dimensional data for visual explanation of feature extraction ability of DBN. The flow chart of processing hyperspectral images of samples was shown in Figure 2.

2.4.1 | Competitive adaptive reweighted sampling

Competitive adaptive reweighted sampling (CARS) is a kind of linear feature selection methods (Sun et al., 2017b). In the progress of feature selection, a subset of multiple wavelength variables is selected, which contains wavelengths with large regression coefficient values in the PLS model through exponentially decreasing function (EDF) and adaptive reweighted sampling (ARS). According to the cross-validation method, the subset with the smallest cross-validation mean square error (RMSECV) is determined. Therefore, wavelength variables contained in this subset are the combination of optimal wavelengths.

2.4.2 | Support vector machine

Support vector machine (SVM) is a machine learning method based on VC dimension theory and structural risk minimization principle to find the optimal classification hyperplane that meets the classification requirements (Friedrichs & Igel, 2005). The RBF kernel function was

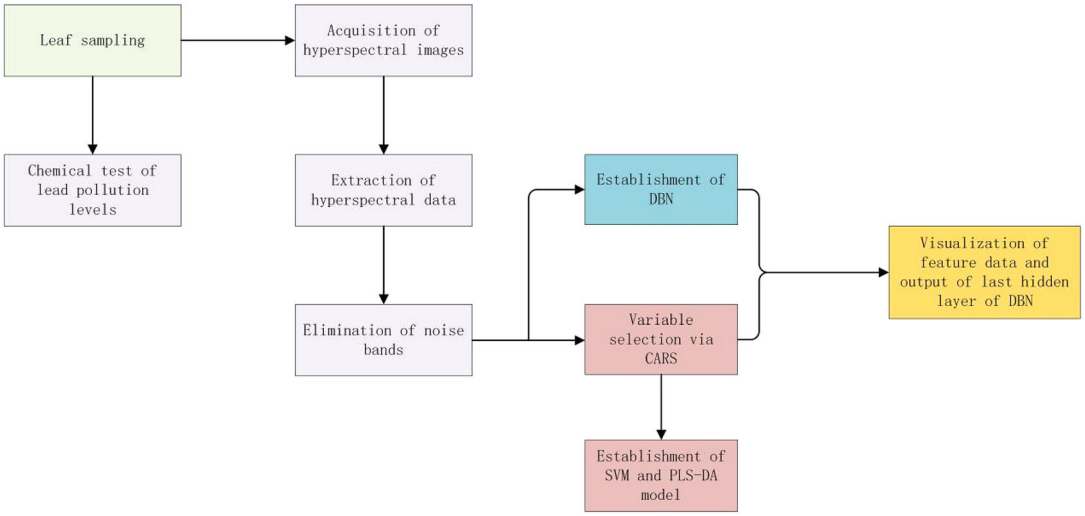


FIGURE 2 Experimental flow chart

17454565, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfs.12866 by Universita Del Salento Ser Inf Bibliot "La Stecca", Wiley Online Library on [30/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

used as a radial basis function in the paper. A grid search algorithm was used to search the best values of penalty parameter (c) and kernel function parameter (g) in the range of $2^{-8} - 2^{8}$.

2.4.3 | Partial least squares discriminant analysis

Partial least squares discriminant analysis (PLS-DA) is considered as a supervised method of maximum discrimination between the samples (Westerhuis et al., 2008). Leave-one-out cross validation is used to establish PLS-DA model. The accuracy of PLS-DA is determined by the relative error between true label and predicted label.

2.4.4 | Deep belief network

Deep belief network (DBN) is a typical deep learning architecture that consists of multiple restricted Boltzman machines (RBMs). RBM is a probability generation model. The structure of RBM was shown in Figure 3.

In this paper, hyperspectral data is the input to train DBN. As shown in Figure 4, the structure of DBN is composed of three RBMs including one input layer, two hidden layers and one output layer. Besides, logistic function is added in the last layer so that hyperspectral data can be mapped to corresponding label. Therefore, DBN model can be used as a discriminant model. The training process of DBN can boil down to unsupervised pretraining and supervised fine-tuning (MinMin et al., 2019).

In pre-training process, RBM is trained from bottom to top by contrastive divergence algorithm (CD) individually. The output of the former RBM is the input of the next RBM. Hence, the entire pre-training progress of DBN is finally finished. Afterwards, in the fine-tuning process, back propagation (BP) is used to fine-tune the whole parameters of DBN through minimizing the error between the expected label and true label from top to bottom. Therefore, the convergence ability and prediction performance of DBN can be improved.

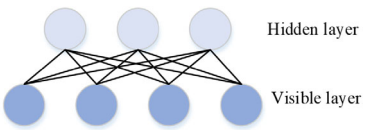


FIGURE 3 Structure of RBM

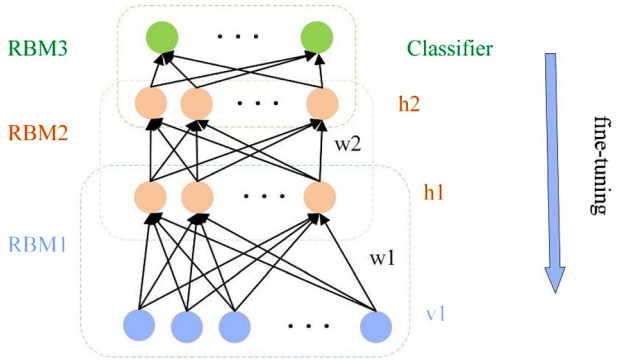


FIGURE 4 Structure of DBN. (a) Raw spectral data. (b) Average spectral data

2.5 | Visualization methods

The t-distribution stochastic neighbor embedding (t-SNE) method converts similarity of data points into probability (Natascha, James, James, & Christopher, 2011). The similarity in the original space is expressed as Gauss joint probability, and the similarity in the embedded space is expressed by t-distribution. It is capable to visualize high-dimensional data by projecting it on a two or three-dimensional map with retaining the local structure of the input data. Therefore, it is used to visualize the abstract feature values from the last hidden layer of DBN.

2.6 | Software tools

The software environment of hyperspectral data processing is MATLAB 2014a (The Math works Inc., Natick, MA), the operating system is win7, the processor is Intel (R) Core (TM) i5-4,460, the CPU is 3.20GHz, and the memory is 4.00 GB.

3 | RESULTS AND DISCUSSION

The measurement results of Pb content were shown in Table 1. According to to national food safety standard of limits of contaminants in food (GB2762-2017), the lead-contaminated lettuce was divided into four categories in this paper, that is nonpolluted, slight polluted, moderate polluted, severe polluted. The average values of lead residues in lettuce leaves of four groups (0, 50, 100, 200 mg/L) were 0 mg/kg (nonpolluted), 0.264 mg/kg (slight), 0.549 mg/kg (moderate) and 1.128 mg/kg (severe), respectively. It can be seen that the higher the stress gradient of Pb, the more Pb content in lettuce leaf, which indicated that the Pb contamination degree of lettuce leaf was positively correlated with the Pb stress gradient.

TABLE 1 The measurement results of Pb content

C /mg·L^{-1}	Min/mg·kg^{-1}	Max/mg·kg^{-1}	Mean/mg·kg^{-1}
0	0	0	0
50	0.168	0.352	0.264
100	0.423	0.670	0.549
200	1.097	1.245	1.128

Note: C is the concentration of Pb solution. Min, Max and Mean represent minimum value, maximum value and mean value of Pb content in the lettuce samples, respectively.

17454565, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfs.12866 by Universita Del Salento Ser Inf Bibliot "La Stecca", Wiley Online Library on [30/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.1 | Analysis of hyperspectral data

The raw spectral data of different lead pollution degrees of all lettuce leaves were presented in Figure 5(a). The average spectrum corresponding to each pollution degree was shown in Figure 5(b). It has been found that when crops absorb excessive amount of heavy metal, the cell structure and biochemical components such as chlorophyll will be affected (Liu, Liu, Chen, Xu, & Ding, 2010; Yingqiong et al., 2003), and thus will cause the variation in crop spectrum (Elliott et al., 2012; Zhao, Wang, Ouyang, & Chen, 2011). The reflection peak observed in 550 nm is formed by the strong reflection of chlorophyll a and chlorophyll b, and the absorption valleys observed in 480 nm and 670 nm are formed by the strong absorption of the pigment (Zhou et al., 2020).

As shown in Figure 5(b), when lettuce was under the stress of 200 mg/L, the cell structure of lettuce leave was destroyed, and the content of pigments decreased, leading to the weakness of the reflectivity of lettuce leaves on infrared radiation. Therefore, the spectral reflectance in visible bands (478–700 nm) increased but the reflectance infrared bands (720–965 nm) decreased. Also, there existed blue shift phenomena for the position of red edge (680–750 nm) (Hoque & Hutzler, 1992). Besides, the spectral curves of slight and moderate polluted lettuces showed a small difference. Therefore, it is necessary to explore an effective method to extract the weak feature in hyperspectral data and realize the identification of lettuces of different lead-polluted degrees.

3.2 | Spectral feature wavelength selection and modeling analysis

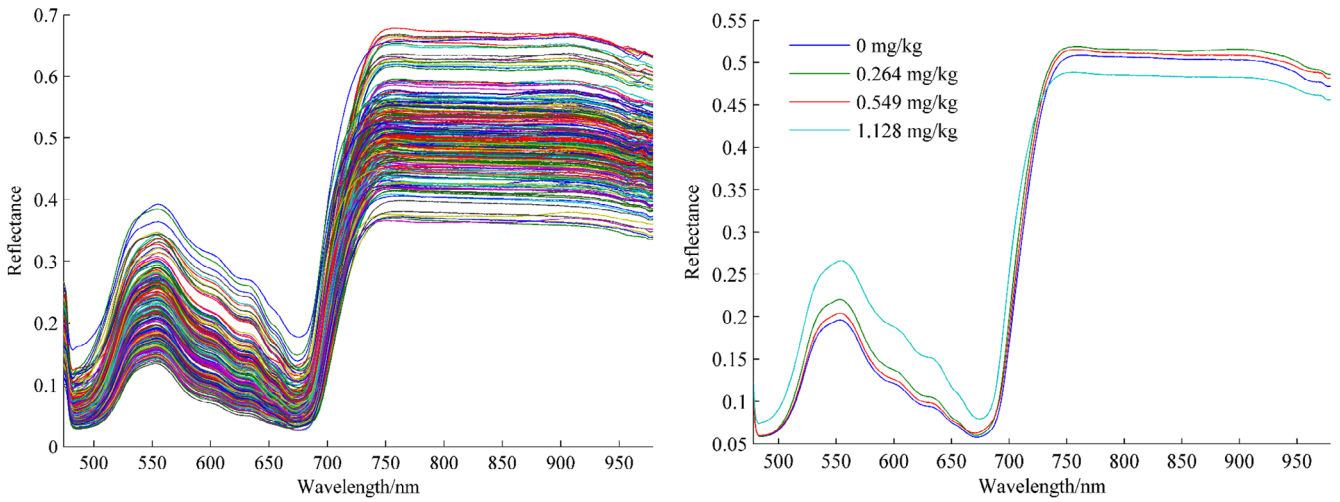
3.2.1 | Variable selection of CARS

The optimal potential wavelength variables were determined according to Monte Carlo cross-validation. The number of MC sampling runs was set to 50. The number of samples is 240, and the number of spectral variables is 399. The multi-dimensional spectrum matrix is displayed as x (240 × 399); the corresponding true value matrix is the degree of heavy metal lead stress, and the variable was set as y (240 × 1). The classification labels of the degree of lead stress were 1,2,3,4, corresponding to four kinds of lettuce samples with different lead stress gradients. The screening results of CARS were shown in Figure 6.

It can be seen from Figure 6(a), the number of wavelengths decreased sharply and then gradually flattened. Therefore, the selection process of CARS including two steps—"rough selection" and "fine selection". Figure 6(b) showed the change trend of root mean square error of five fold cross validation (RMSEV). During the former 17 sampling runs, RMSEV value presented a decreasing trend, which indicated that some independent information related to lead pollution degree were gradually removed. However, RMSEV value increased after 17 sampling runs for the reason that a small amount of important spectral information was eliminated. As shown in Figure 6(c), the regression coefficients of the respective wavelengths changed with the increasing of sampling runs. Besides, RMSEV was smallest at the 17th sampling run. Therefore, a total of 71 wavelengths obtained at the 17th sampling run were determined as the key wavelengths, which were marked in Figure 6(d).

3.2.2 | Establishment of conventional classification models

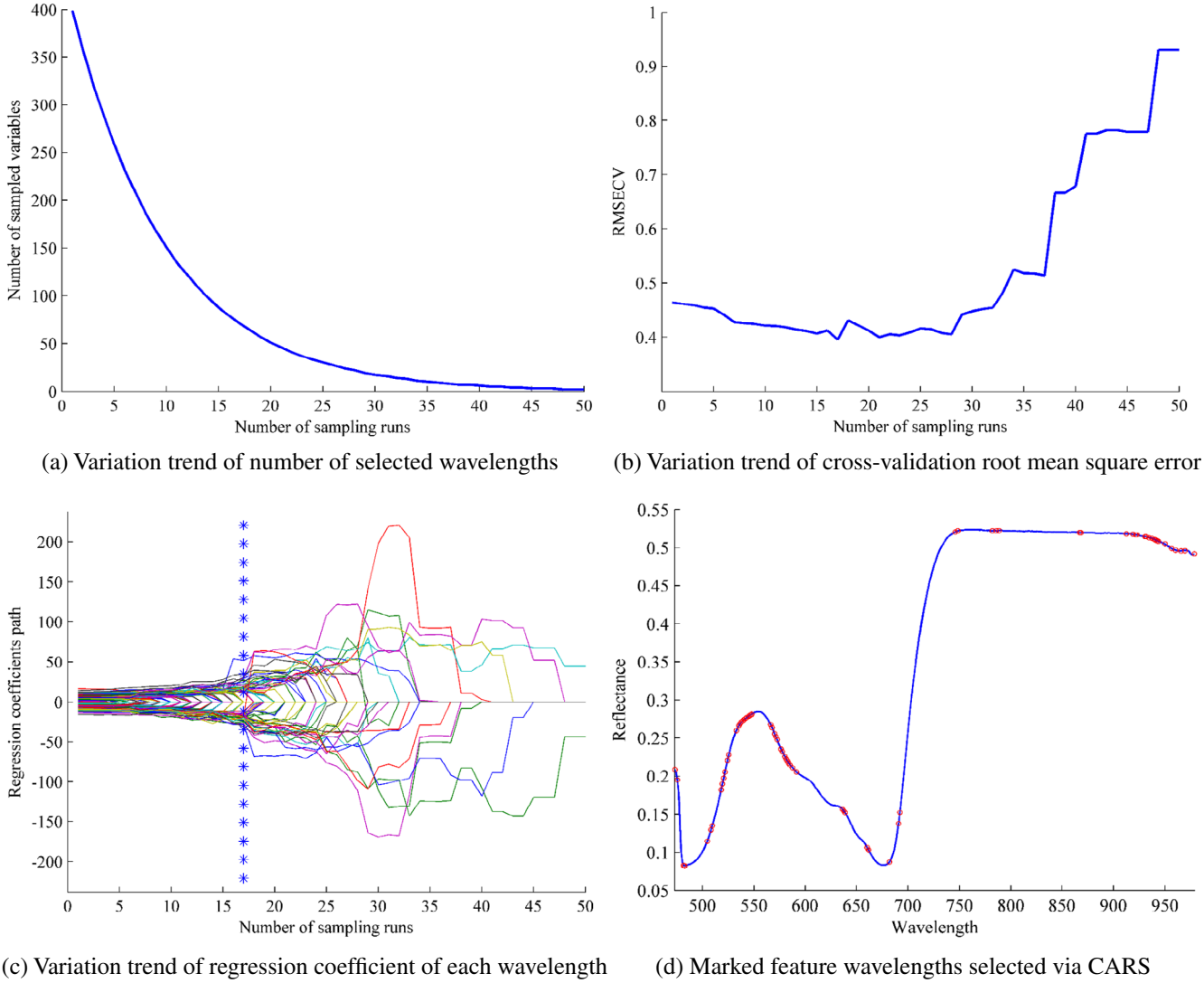
The full band spectral data and feature data selected by CARS were used as the characteristic parameters for detecting the lead pollution level in the lettuce leaves. The category labels corresponding to the four pollution levels of nonpolluted, slight polluted, moderate polluted, severe polluted were set to 1, 2, 3, and 4, and then the traditional



(a) Raw spectral data (b) Average spectral data

FIGURE 5 Raw spectral data and Average spectral data. (a) Variation trend of number of selected wavelengths. (b) Variation trend of cross-validation root mean square error. (c) Variation trend of regression coefficient of each wavelength. (d) Marked feature wavelengths selected *via* CARS

17454565, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfs.12866 by Universita Del Salento Ser Inf Bibliot "La Stecca", Wiley Online Library on [30/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) Variation trend of number of selected wavelengths

(b) Variation trend of cross-validation root mean square error

(c) Variation trend of regression coefficient of each wavelength

(d) Marked feature wavelengths selected via CARS

FIGURE 6 Progress of CARS wavelength selection. (a) Visualization map of raw spectral data. (b) Visualization map of feature data *via* CARS. (c) Visualization map of output from DBN's last hidden layer

discriminant model (PLS-DA and SVM) is established. It could be seen from Table 2 that the performance of PLS-DA and SVM had been improved after feature selection through CARS algorithm, and CARS-SVM model reached the highest accuracy with the training set of 95.56% and test set of 91.67%. However, the classification accuracy (<95%) could not meet the requirement in practical application. Therefore, it was necessary to explore a more sufficient model to improve the classification results of lead pollution degree of lettuce leaves.

3.2.3 | Identification results of DBN

The key parameters of DBN were set as follows. The number of nodes in input layer was 399, which was determined by the dimension of input spectral data. Considering that lead pollution degree were four, the number of nodes in output layer was set to 4. The relationship between the output and lead pollution degree was shown in Table 3. The number of iterations of RBM was set to 30, the learning rate of DBN was 0.01 and the iteration number of DBN was 1,000.

The number of hidden nodes had great influence on the training time of models and classification accuracy, but it always depended on artificial experience. If the number of hidden nodes was too small, it would cause under-fitting problems, that is to say, the classification results were insufficient. If the number of hidden nodes was too large, DBN was easy to fall into local optimum and the operation load would be increased. Therefore, in order to determine the optimal number of hidden nodes, the minimum number of hidden nodes was 10, and increased by 10 until the number of hidden nodes reached 150. The identification results of DBN with different hidden nodes were shown in Table 4.

The performance of DBN was gradually improved with increase of the number of hidden nodes. After DBN reached the highest accuracy, the classification accuracy of DBN remained 95%. What's more, when the number of hidden nodes was 110, DBN had the best performance with the training set accuracy of 100% and test set accuracy of 96.67%.

17454565, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfs.12866 by Universita Del Salento Ser Inf Bibliot "La Stecca", Wiley Online Library on [30/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Classification results of traditional models

Feature selection methods	Models	Parameters	Accuracy%	
			Training set	Test set
FS	PLS-DA	10	86.67	83.33
	SVM	(256,1)	88.89	83.33
CARS	PLS-DA	10	93.89	90.00
	SVM	(256,1)	95.56	91.67

Note: PLS-DA model's parameter refers to the optimal number of LVs; SVM model's parameters mean different penalty parameters(c) and kernel function parameters (g), which are shown as (c, g).

TABLE 3 Relationship between output of DBN and lead pollution degree

Label	Output of DBN	Lead pollution level
1	0001	None
2	0010	Slight
3	0100	Moderate
4	1,000	Severe

TABLE 4 Identification results of DBN with different hidden nodes

Network structure of DBN	Accuracy%	
	Training set	Test set
399–10–10-4	81.11	78.33
399–20–20-4	87.22	83.33
399–30–30-4	95.56	90.00
399–40–40-4	95.56	90.00
399–50–50-4	95.56	90.00
399–60–60-4	96.67	91.67
399–70–70-4	98.89	93.33
399–80–80-4	98.89	93.33
399–90–90-4	100.00	95.00
399–100–100-4	100.00	95.00
399–110–110-4	**100.00**	**96.67**
399–120–120-4	100.00	95.00
399–130–130-4	100.00	95.00
399–140–140-4	100.00	95.00
399–150–150-4	100.00	95.00

Note: The bold values in Table 4 indicate that the DBN identification result was the best when the number of hidden nodes was 110, and the test set had the highest accuracy.

3.2.4 | Comparison of conventional models and DBN

The best performance of conventional models and DBN were compared to determine the best discriminant model. The final results of CARS and DBN were shown in Table 5.

TABLE 5 Comparison of traditional model and DBN

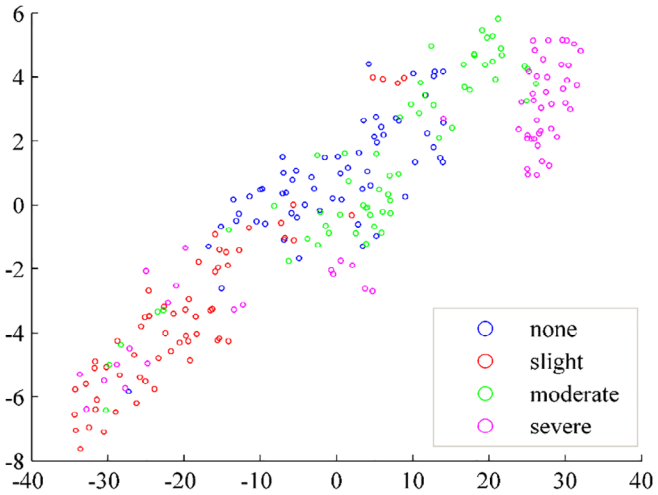
Models	Training time/s	Accuracy %	
		Training set	Test set
CARS-SVM	29.455	95.56	91.67
DBN	23.552	100.00	96.67

Obviously, DBN had better performance than CARS-SVM, which could be explained by the following reasons. Firstly, hyperspectral data was complex and nonlinear, so linear variable selection method (CARS) may ignore weak useful spectral information. SVM belonged to shallow neural network and had limits on dealing with multi-classification problems. However, it was more effective to extract deep features in hyperspectral data by using DBN. Deep belief network usually involved a deep architecture, which could extract more abstract features layer by layer, resulting a better performance than conventional models such as CARS-SVM. Besides, the training of variable selection method was time-consuming, so the training time of CARS-SVM was 29.455 s, which was more than that of DBN with 23.552 s. Therefore, DBN is more suitable to deal with hypersectral data than traditional models due to higher classification accuracy and less training time.

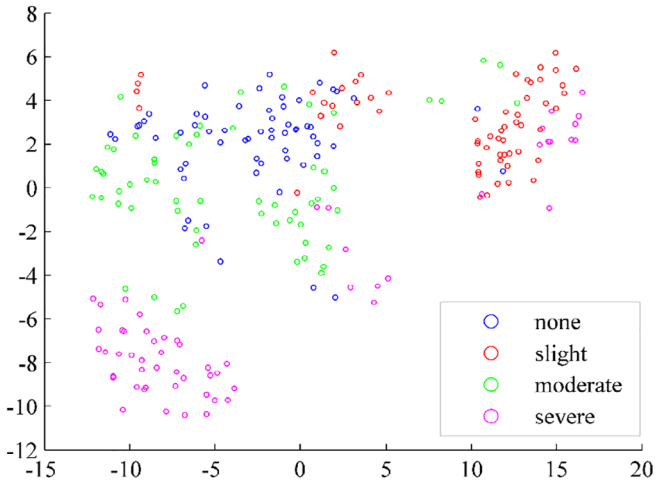
3.3 | Visualizing data

To visually illustrate the effectiveness of DBN for identification of lead pollution degrees based on hyperspectral image technology, t-SNE was used to visualize spectral data, which had the ability to map high-dimensional data into a low-dimensional nonlinear mainfold by aggregating similar data points. As shown in Figure 7, the raw spectral data, feature selection data *via* CARS and feature values from DBN's last hidden layer were visualized on a two-dimensional plane, and different lead pollution degrees were represented by different colors. The visualized data points of original spectral data were mixed and overlapped. Though CARS algorithm was used to select variable wavelengths from the original spectral data, the data points were still difficult to be separated. On the contrast, visual points of feature values collected from the last hidden layer of DBN were clearly separated. There existed obvious boundaries among the visualization data points from t-SNE, which could clearly explain the effectiveness of DBN in dealing with hyperspectral data for detection of lead pollution levels.

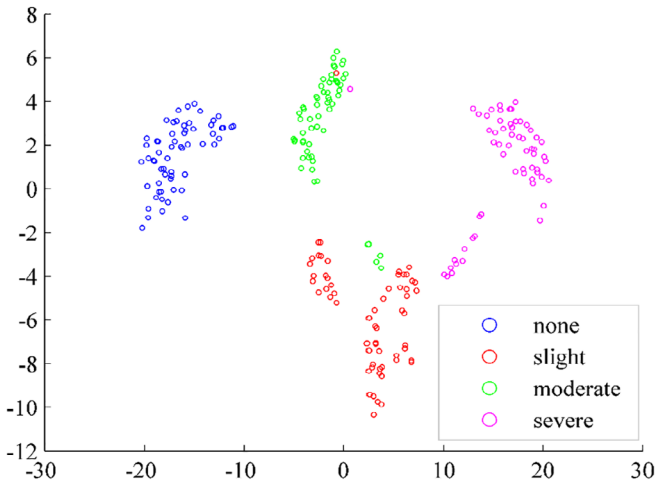
17454565, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfs.12866 by Universita Del Salento Ser Inf Bibliot "La Stecca", Wiley Online Library on [30/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) Visualization map of raw spectral data



(b) Visualization map of feature data via CARS



(c) Visualization map of output from DBN's last hidden layer

FIGURE 7 Visualization maps of different data based on t-SNE

Although hyperspectral imaging technology combined with deep belief network and t-SNE is successfully used to detect lead pollution level in lettuce leaves, the current research is at proof-of-concept stage. Because the experimental environment is highly controlled, the lettuces cultivated in this experiment are different from the lettuces cultivated in the natural environment, and the number of samples in this article is small, so the practicality and applicability still need to be improved. Therefore, it is also necessary to strengthen the combination of indoor simulation and field measurement, and modify the model through large-scale field observation data to improve the accuracy and practicability of the model in future research.

4 | CONCLUSION

Hyperspectral image system was used to obtain hyperspetral images of lettuce leaves, then ENVI software was adopted to extract hyperspetral data of all samples. SVM and PLS-DA discriminant models were established based on full spectral data and spectral data selected by CARS to identify lead pollution degree of lettuce leaves. CARS-SVM had the highest accuracy with training set of 95.56% and test set of 91.67%. Moreover, compared with conventional model (CARS-SVM), DBN model had more excellent learning and prediction ability. The highest accuracy of training set and test set reached 100% and 96.67%, respectively. Also, the training time of DBN was less than that of CARS-SVM, indicating that DBN was practical for quick detection of lead pollution degrees of lettuce leaves. Besides, t-SNE algorithm was used to demonstrate the feature selection ability of CARS and feature extraction ability of DBN. It turned out that DBN had stronger feature extraction ability, which was more suitable for processing high-dimensional nonlinear spectral data.

It could be summarized that DBN discriminant model was better than conventional models in classification ability and less training time in this study. Therefore, it is effective to realize quick and nondestructive detection for lead pollution degrees of lettuce leaves based on DBN and hyperspectral image technology, hoping to provide a reference for subsequent study on heavy metal stress in vegetables.

ORCID

Jun Sun https://orcid.org/0000-0003-3019-6086
Yan Cao https://orcid.org/0000-0003-4638-7717
Minmin Wu https://orcid.org/0000-0003-2814-0577

REFERENCES

Bhat, R., Kiran, K., Arun, A. B., & Karim, A. A. (2010). Determination of mineral composition and heavy metal content of some nutraceutically valued plant products. *Food Analytical Methods*, *3*(3), 181–187.

Dala-Paula, B. M., Custódio, F. B., Knupp, E. A. N., Palmieri, H. E. L., Silva, J. B. B., & Glória, M. B. A. (2018). Cadmium, copper and lead levels in different cultivars of lettuce and soil from urban agriculture. *Environment Pollution*, *242*, 383–389.

Elliott, A. D., Gao, L., Ustione, A., Bedard, N., Kester, R., Piston, D. W., & Tkaczyk, T. S. (2012). Real-time hyperspectral fluorescence imaging of pancreatic-cell dynamics with the image mapping spectrometer. *Journal of Cell Science*, *125*(20), 4833–4840.

Friedrichs, F., & Igel, C. (2005). Evolutionary tuning of multiple svm parameters. *Neurocomputing*, *64*, 107–117.

Gu, Y., Li, S., Gao, W., & Wei, H. (2015). Hyperspectral estimation of the cadmium content in leaves of *Brassica rapa* chinesis based on the spectral parameters. *Acta Ecologica Sinica*, *35*(13), 4445–4453.

17454565, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfs.12866 by Universitá Del Salento Ser Inf Bibliot "La Stecca", Wiley Online Library on [30/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*, 504–507.

Hoque, E., & Hutzler, P. J. S. (1992). Spectral blue-shift of red edge minitors damage class of beech trees. *Remote Sensing of Environment*, *39*(1), 81–84.

Jahromi, E. Z., Bidari, A., Assadi, Y., Hosseini, M. R. M., & Jamali, M. R. (2007). Dispersive liquid–liquid microextraction combined with graphite furnace atomic absorption spectrometry: Ultra trace determination of cadmium in water samples. *Analytica Chimica Acta*, *585*(2), 305–311.

Khulal, U., Zhao, J., Hu, W., & Chen, Q. (2016). Nondestructive quantifying total volatile basic nitrogen (tvb-n) content in chicken using hyperspectral imaging (hsi) technique combined with different data dimension reduction algorithms. *Food Chemistry*, *197*(APR.15PT.B), 1191–1199.

Li, X., Ku, J., Yan, Y., Xu, M., & Jin, R. (2016). Detection method of green potato based on hyperspectral imaging. *Transactions of the Chinese Society for Agricultural. Machinery*, *47*, 228–234.

Liu, D., Liu, X., Chen, Z., Xu, H., & Ding, X. (2010). Bioaccumulation of lead and the effects of lead on catalase activity, glutathione levels, and chlorophyll content in the leaves of wheat. *Communications in Soil Science and Plant Analysis*, *41*(8), 935–944.

Liu, D. J., Guo, X. J., & Guo, W. H. (2017). Investigation of heavy metal contamination of vegetable base in Qinghai province from 2012 to 2014. *Journal of Medical Pest Control*, *33*(07), 779–781.

Lonni, A. A. S. G., Scarminio, I. S., Silva, L. M. C., & Ferreira, D. T. (2005). Numerical taxonomy characterization of baccharis genus species by ultraviolet-visible spectrophotometry. *Analytical Sciences*, *21*(3), 235–239.

Lu, W., Guo, Y. M., Zhang, X. Y., & Wang, X. Y. (2018). Rice germination rate detection based on fluorescent spectrometry and deep belief network. *Spectroscopy and Spectral Analysis*, *38*, 1303–1312.

MinMin, W., Jun, S., Bing, L., Xiao, G., Xin, Z., & Mengli, Z. (2019). Application of deep brief network in transmission spectroscopy detection of pesticide residues in lettuce leaves. *Journal of Food Process Engineering*, *42*(3), e13005.

Natascha, B., James, S., James, B., & Christopher, W. (2011). An intuitive graphical visualization technique for the interrogation of transcriptome data. *Nuclceic Acids Research*, *39*(17), 7380–7389.

Norton, G. J., Deacon, C. M., Mestrot, A., Feldmann, J., Jenkins, P., Baskaran, C., & Andrew, A. M. (2015). Cadmium and lead in vegetable and fruit produce selected from specific regional areas of the Uk. *Science of the Total Environment*, *533*, 520–527.

Pourrut, B., Shahid, M., Dumat, C., Winterton, P., & Pinelli, E. (2011). Lead uptake, toxicity, and detoxification in plants. *Reviews of Environmental Contamination and Toxicology*, *213*, 113–136.

Sun, J., Cong, S., Mao, H., Wu, X., & Yang, N. (2018). Quantitative detection of mixed pesticide residue of lettuce leaves based on hyperspectral technique. *Journal of Food Process Engineering*, *41*(2), e12654.

Sun, J., Cong, S., Mao, H., Wu, X., Zhang, X., & Wang, P. (2017a). Cars-abc-svr model for predicting leaf moisture of leaf-used lettuce based on hyperspectral. *Transactions of the Chinese Society of. Agricultural Engineering*, *33*(5), 178–184.

Sun, J., Zhou, X., Mao, H., Wu, X., Zhang, X., & Li, Q. (2017b). Discrimination of pesticide residues in lettuce based on chemical molecular structure coupled with wavelet transform and near infrared hyperspectra. *Journal of Food Process Engineering*, *40*(4), e12509.

Sun, J., Zhou, X., Wu, X., Lu, B., Dai, C., & Shen, J. (2019). Research and analysis of cadmium residue in tomato leaves based on WT-LSSVR and Vis-NIR hyperspectral imaging. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *212*, 215–221.

Sunli, C., Jun, S., Hanping, M., Xiaohong, W., Pei, W., & Xiaodong, Z. (2018). Non-destructive detection for mold colonies in rice based on hyperspectra and gwo-svr. *Journal of the Science of Food & Agriculture*, *98*, 29–35.

Wei, F. S., & Chen, J. S. (1991). Studies on environmental background values in China. *Environmental Science*, *012*(004), 12–19.

Wei, S., Baiming, C., & Lin, L. (2013). Soil heavy metal pollution of cultivated land in China. *Research of Soil and Water Conservation*, *20*(3), 293–298.

Wen, X., Wu, P., Chen, L., & Hou, X. (2009). Determination of cadmium in rice and water by tungsten coil electrothermal vaporization-atomic fluorescence spectrometry and tungsten coil electrothermal atomic absorption spectrometry after cloud point extraction. *Analytica Chimica Acta*, *650*(1), 33–38.

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J. V., … Dorsten, F. A. (2008). Assessment of PLSDA cross validation. *Metabolomics*, *4*(1), 81–89.

Wu, X., Fu, H., Tian, X., Wu, B., & Sun, J. (2017). Prediction of pork storage time using fourier transform near infrared spectroscopy and adaboost-ulda. *Journal of Food Process Engineering*, *40*, e12566.

Yingqiong, D., Jianghua, H., Junjian, C., Xiuguo, W., Xiuqin, Y., Shaoyi, W., & Wenbao, H. (2003). Effects of heavy metals of pb, cd and cr on the growth of vegetables andtheir uptake. *Acta Horticulturae Sinica*, *30*(1), 51–55.

Yu, K., Geel, M. V., Ceulemans, T., Geerts, W., Ramos, M. M., Serafim, C., … Somers, B. (2018). Vegetation reflectance spectroscopy for biomonitoring of heavy metal pollution in urban soils. *Environmental Pollution*, *243*(Part B), 1912–1922.

Yuping, W., Xiaohui, Z., & Yan, X. U. (2017). Analysis of lead contamination in vegetables in Yunnan province from 2012 to 2016. *Journal of Food Safety & Quality*, *8*(10), 3721–3726.

Zhao, J. W., Wang, K. L., Ouyang, Q., & Chen, Q. S. (2011). Measurement of chlorophyll content and distribution in tea plant's leaf using hyperspectral imaging technique. *Spectroscopy and Spectral Analysis*, *31*(2), 512–515.

Zhaoyong, Z., Dongjian, H. E., Haihui, Z., Yu, L., Dong, S. U., & Ketao, C. (2017). Non-destructive detection of moldy core in apple fruit based on deep belief network. *Food Science*, *38*, 297–304.

Zhou, X., Sun, J., Mao, H., Wu, X., Zhang, X., & Yang, N. (2018). Visualization research of moisture content in leaf lettuce leaves based on wt-plsr and hyperspectral imaging technology. *Journal of Food Process Engineering*, *41*, e12647.

Zhou, X., Sun, J., Tian, Y., Lu, B., Hang, Y. Y., & Chen, Q. S. (2020). Hyperspectral technique combined with deep learning algorithm for detection of compound heavy metals in lettuce. *Food Chemistry*, *321*, 126503.

Zhu, Y., Qu, Y., Liu, S., & Chen, S. (2014). Spectral response of wheat and lettuce to copper pollution. *Journal of Remote Sensing*, *18*(2), 335–352.

How to cite this article: Sun J, Cao Y, Zhou X, Wu M, Sun Y, Hu Y. Detection for lead pollution level of lettuce leaves based on deep belief network combined with hyperspectral image technology. *J Food Saf*. 2021;41:e12866. https://doi.org/10.1111/jfs.12866

17454565, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfs.12866 by Universita Del Salento Ser Inf Bibliot "La Stecca", Wiley Online Library on [30/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License