Development of one-class classification method for identifying healthy *T. granosa* from those contaminated with uncertain heavy metals by LIBS

Zhonghao Xie[1], Xi'an Feng[1], Xiao Chen[2], Guangzao Huang[3], Xiaojing Chen[3], Limin Li[3*], Wen Shi[3], Chengxi Jiang[3], Shuwen Yu[4]

(1. *School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China*;
2. *Wenzhou Institute of Industry & Science, Wenzhou 325035, Zhejiang, China*;
3. *College of Electrical and Electronic Engineering, Wenzhou University, Wenzhou 325035, Zhejiang, China*;
4. *Information Technology Center, Wenzhou University, Wenzhou 325035, Zhejiang, China*)

Abstract: Laser-induced breakdown spectroscopy (LIBS) can be used for the rapid detection of heavy metal contamination of *Tegillarca granosa* (*T. granosa*), but an appropriate classification model needs to be constructed. In the one-class classification method, only target samples are needed in training process to achieve the recognition of abnormal samples, which is suitable for rapid identification of healthy *T. granosa* from those contaminated with uncertain heavy metals. The construction of a one-class classification model for heavy metal detection in *T. granosa* by LIBS has faced the problem of high-dimension and small samples. To solve this problem, a novel one-class classification method was proposed in this study. Here, the principal component scores and the intensity of the residual spectrum were combined as extracted features. Then, a one-class classifier based on Mahalanobis distance using the extracted features was constructed and its threshold was set by leave-one-out cross-validation. The sensitivity, specificity and accuracy of the proposed method were reached to 1, 0.9333 and 0.9667 respectively, which are superior to the previously reported methods.

Keywords: laser-induced breakdown spectroscopy, Heavy metal contamination, *Tegillarca granosa*, one-class classification
DOI: 10.25165/j.ijabe.20231604.7666

Citation: Xie Z H, Feng X, Chen X, Huang G Z, Chen X J, Li L M, et al. Development of one-class classification method for identifying healthy *T. granosa* from those contaminated with uncertain heavy metals by LIBS. Int J Agric & Biol Eng, 2023; 16(4): 200–205.

1 Introduction

In recent years, heavy industries have been developing rapidly and discharging numerous polluting heavy metals into rivers, oceans, and other water bodies. The pollution, consequently, results in heavy metal contamination of aquatic products and creates serious problems[1]. Heavy metals mainly include Hg, Cd, Pb, Zn, Cu, Co, Sn, etc. Harmful heavy metals not only disturb the natural system of aquatic animals but also lead to the poisoning and death of aquaculture species. In addition, heavy metals accumulate in the human body by eating food contaminated with heavy metals. When heavy metals interact with enzymes, they can cause acute, subacute or chronic poisoning, which seriously affects human health and safety[2]. Therefore, it is very important to improve the detection ability of heavy metal contamination in aquatic products to ensure quality and safety.

Tegillarca granosa (*T. granosa*) has become one of the most important commercial seafood products in eastern Asia due to its delicious taste, high nutritional value and health care effect[3]. However, *T. granosa* is a heavy metal accumulating organism, which is mainly due to the mudflat aquaculture and non-selective filter-feeding habits. *T. granosa* absorbs heavy metals from the water around its gills. Heavy metal ions are then transferred to different parts of *T. granosa* through its blood. *T. granosa* can be used as an indicator organism for the determination of heavy metal contamination in sea areas because *T. granosa* can well reflect the content and species of heavy metals in their surrounding area[4,5].

At present, there are many methods to detect heavy metals. Traditional heavy metal detection methods include biochemical detection and chemical spectroscopic detection such as flame atomic absorption spectroscopy (FAAS)[6], inductively coupled plasma atomic emission spectrometry (ICP-AES)[7] and graphite furnace atomic absorption spectrometry (GFAAS)[8]. Though these methods have high sensitivity and accuracy, they are time-consuming, labor-intensive and require high professional knowledge and ability. Compared with these traditional detection methods, laser-induced breakdown spectroscopy (LIBS) technology does not require time-consuming and labor-intensive pretreatment of samples and involves simple equipment. The principle of LIBS is that the atoms in the sample are excited by high power pulse laser to form a high-temperature plasma spark, and the excited atoms and ions emit characteristic spectral lines in the process of de-excitation[9]. Due to the sample matrix effect, laser energy instability, element spectral line collapse and overlapping, the collected laser-induced breakdown spectral curves may have some deviation. The composition characteristics of *T. granosa* samples

Received date: 2022-11-13 **Accepted date:** 2023-02-04

Biographies: Zhonghao Xie, PhD, research interest: data analysis and modelling, Email: xzh@mail.nwpu.edu.cn; **Xi'an** Feng, Professor, research interest: signal analysis, Email: fengxa@nwpu.edu.cn; **Xiao Chen**, Postgraduate, research interest: spectroscopy and Chemistry, Email: m18801133301@163.com; **Guangzao Huang**, PhD, research interest: machine learning and spectroscopy, Email: guangzh@wzu.edu.cn; **Xiaojing Chen**, Professor, research interest: machine learning and spectroscopy, Email: 00101094@wzu.edu.cn; **Chengxi Jiang**, PhD, research interest: data analysis and modelling, Email: 20190408@wzu.edu.cn; **Shuwen Yu**, Postgraduate, research interest: data analysis and modelling, Email: gsyusw@163.com

***Corresponding author: Limin Li**, PhD, research interest: machine learning and spectroscopy. College of Electrical and Electronic Engineering, Wenzhou University, Wenzhou 325035, Zhejiang, China, Email: lilimin@wzu.edu.cn

are complex， changeable and have a large number of elements. Therefore, the detection of heavy metal contamination in *T. granosa* by LIBS needs to apply chemometrics methods to construct analysis models with thousands of spectral variables.

Currently, there are some published studies on using traditional supervised classification methods for heavy metal detection in *T. granosa* by LIBS. For example, Ji et al.[10] used the algorithms of wavelet and information gain for features selection to construct a random forest classifier and achieved a classification accuracy of 93.3%. Xie et al.[11] combined linear regression classification and threshold feature selection to construct classifiers, and the best classification accuracy was 90.67%. For the actual heavy metal contamination detection in *T. granosa*, the state is uncertain, such as single heavy metal contamination and multiple heavy metal cross-contamination. However, the traditional supervised classification method needs to know all the information of samples to ensure a satisfactory accuracy rate. This leads to great limitations in the actual heavy metal contamination detection of *T. granosa* by LIBS.

In the field of spectral detection of food safety, some studies based on one-class classification methods have been published[12-15]. For the one-class classification problem, only target class samples are used to train the one-class classifier to achieve the purpose of distinguishing abnormal class samples[16-18]. Therefore, this study deals with the heavy metal contamination detection of *T. granosa* by LIBS from the perspective of one-class classification. For the problem studied, there are only a dozen samples used for training the classification model, and the spectral variables have tens of thousands of dimensions, which are high-dimensional and small samples. Therefore, how to improve the generalization of classification methods on high-dimensional and small sample data is the key issue to be solved while using one-class classification methods for the heavy metal detection in *T. granosa* by LIBS.

In this study, a new one-class classification method is proposed, which combines principal component analysis[19] and Mahalanobis distance[20], named PCAMA. Firstly, the principal component scores of the global spectrum and the intensity of the residual spectrum were obtained and combined as the extracted features. Then, Mahalanobis distance was used to evaluate the similarity of samples with the extracted features. The performance of the proposed method was compared with classical one-class classification methods including a data-driven version of soft independent modeling of class analogy (DD_SIMCA)[21], one-class partial least squares (OCPLS)[22-25] and support vector data description (SVDD)[26,27].

2 The proposed one-class classification method (PCAMA)

The proposed one-class classification method includes two parts, feature extraction and classifier construction. As this method is to distinguish abnormal classes only by learning the target class, so, the extracted features should preserve the information of the target class as much as possible. Besides, features extraction inevitably loses feature information and results in reduced specificity for abnormal classes. Therefore, the information of the global feature space should be preserved as much as possible. Based on the above requirements for feature dimension reduction of one-class classification, the following dimension reduction method was adopted.

Let X be the (n×m) centralized matrix that represents a set of values collected for n training samples at m variables. Principal component analysis of X is performed according to the following equation

$$X = TP^t + E \tag{1}$$

where, $T = \{t_{na}\}$ is the $(N \times A)$ scores matrix, $P = \{p_{ma}\}$ is the $(M\times A)$ loadings matrix, $E = \{e_{nm}\}$ is the $(N\times M)$ matrix of residuals and A is the number of principal components (PCs). The first-order norm of e_{im} in E is calculated and spliced to form the $(N\times 1)$ residual intensities matrix $S = \{|e_{im}|\}$. T and S are spliced together to form the $[N\times(A+1)]$ extracted feature matrix $F = [t_{n1},\ldots,t_{na},|e_{nm}|]$. For a centralized test sample x， its extracted feature vector $k = [t,|e|]$ is calculated by the equations

$$t = xP \tag{2}$$

$$\tilde{x} = tP^t \tag{3}$$

$$e = x - \tilde{x} \tag{4}$$

where, t and e represent the feature information of sample x in main space and residual space respectively. Therefore, the feature extraction method proposed in this paper is oriented to the one-class classification problem, while the traditional PCA method is oriented to the unsupervised problem. Note the residual e is important for the proposed method and it reflects the outlyingness of the corresponding sample with respect to main space. For different kinds of heavy metal contaminated samples. Their residuals can be significantly different.

For the extracted features, Mahalanobis distance is used as the one-class classifier. The Mahalanobis distance has the advantages of not being affected by the dimension and elimination of the interference of the correlation between variables. Moreover, it is an effective indicator for evaluating the similarity between a sample and a data set. The Mahalanobis distance from a sample v to a target matrix D is calculated as

$$g = \sqrt{(v-\mu)\sum^{-1}(v-\mu)^T} \tag{5}$$

where, μ and $\sum$ are the central vector and covariance matric of D respectively; g indicates the similarity between the sample v and the target matrix D, a smaller value represents a higher similarity of the sample to the target class.

For the one-class classification problem, a threshold is needed, which is set according to the type I error α (the probability of misclassifying a sample belonging to the target class as an abnormal class sample). The leave-one-out cross-validation method is used to calculate the extracted features and Mahalanobis distance of training samples. The $(1-\alpha)\times 100$ quantile of Mahalanobis distance of the training is set as the threshold θ. The unknown sample can be classified by comparing its Mahalanobis distance g with the threshold θ according to the following equation.

$$h(g) = \begin{cases} 1, & if\ g \le \theta \\ 0, & if\ g \ge \theta \end{cases} \tag{6}$$

where, 1 and 0 indicate the target class and abnormal class, respectively.

3 Experiment

3.1 Sample preparation

The *T. granosa* samples were provided by the Zhejiang Mariculture Research Institute (Wenzhou, China). High purity chemical reagents ($PbCH_3COO\cdot 3H_2O$, $CdCl_2$, Zn $SO_4\cdot 7H_2O$) were purchased from the Chemical Reagent Co. Ltd., Shanghai, China for heavy metal solutions. The *T. granosa* samples were divided into five groups, labeled as groups I, II, III, IV and V. The samples were preserved in a water tank with seawater (at pH 8.05±0.10, temperature 22.4°C±5.6°C, dissolved oxygen content >6 mg/L and salinity 21‰). Groups Ⅰ, Ⅱ, and Ⅲ were fed in water dissolved with highly concentrated $PbCH_3COO·3H_2O$ (1.833 mg/L), $CdCl_2$ (1.634 mg/L), Zn $SO_4·7H_2O$ (4.424 mg/L), respectively. Group IV was fed in the water evenly mixed with the above three chemicals. Group V was fed in seawater without adding any heavy metal. The *T. granosa* samples from each group were fed for 10 d to accumulate heavy metals. After the incubation period, 30 samples of *T. granosa* were taken from each group and placed in a refrigerator at –4°C for 30 min for execution. Then, they were frozen, dried and ground for spectral analysis.

3.2 Collection of spectral data

The LIBS experimental device system mainly consists of a laser and a spectrometer. The laser used in this experiment was a pulsed Nd: YAG laser (Litron Nano SG 150-10, Litron Lasers, Warwickshire, England) with a wavelength of 1064 nm, pulse duration of 6 ns, energy of 150 mJ and pulse repetition frequency of 5 Hz. The high-energy, short-pulse laser beam emitted by the laser was vertically focused onto the surface of the sample through a convex lens with a diameter of 30 mm and a focal length of 100 mm. A beam splitter was used to split a small portion (10%) of the pulse energy into an energy meter for monitoring. The optical fiber probe was used to collect the emission spectra at different wavelengths. The emission spectra were transmitted to the spectrometer analysis and processed by computer software. To reduce the fluctuations caused by the instability of the laser pulse, the laser was excited 20 times during the collection of spectral signals. Then, the 20 spectral signals were accumulated and averaged. A total of 150 *T. granosa* samples (30 samples from each group) were collected for this study and the obtained LIBS data was a (150×30 267) matrix.

4 Results and discussion

4.1 LIBS analysis of *T. granosa* samples

The LIBS analysis (200-800 nm) of *T. granosa* healthy group and heavy metal contaminated groups is shown in Figure 1. The spectra of the samples were quite complex, having many characteristic spectral lines which represented different elements. According to the data from the Atomic Spectra Database of the American Institute of Standards and Technology, atomic and ion spectral lines of Al, C, Ca, Cd, Pb, Fe, K, Mg, Na, Si, Sr, and Zn were mainly included in the spectral range of 200-800 nm. The prominent spectral lines for Ca Ⅰ (422.7 nm), Ca Ⅱ (393.3 nm, 396.8 nm), Na Ⅰ (588.9 nm, 589.5 nm), Mg Ⅰ (285.2 nm, 517.3 nm, 518.4 nm), Mg Ⅱ (279.5 nm, 280.3 nm) and K Ⅰ (766 nm, 770 nm) were observed. Similarly, less prominent spectral lines for C Ⅰ (247.8 nm), Sr Ⅰ (460.7 nm), Zn Ⅰ (330.3 nm), Si Ⅰ (288.2 nm), Al Ⅰ (394.4 nm, 396.2 nm) and Fe Ⅰ (438.4 nm, 440.5 nm) were noted. Although many characteristic spectral lines were found in LIBS spectra of *T. granosa* samples, it was difficult to distinguish the healthy group and heavy metal contamination groups only by naked eyes. This is because *T. granosa* samples are organisms and their emission lines in LIBS spectra are complex. Most significant lines belong to atomic like Ca I (422.7 nm), Ca II (393.3 nm, 396.8 nm), Na I (588.9 nm, 589.5 nm), Mg I (285.2 nm; 517.3 nm; 518.4 nm). However, affected by the heavy metal, the emission lines of these atomic have strong relevance with the concentration of target heavy metal. One possible reason for the relevance is that biological tissues of the contaminated samples are stressed by heavy metals to generate various metal oxides. In this case, it is possible for chemometric approach to fulfill the classification task. Therefore, it is necessary to do further analysis with chemometrics methods.

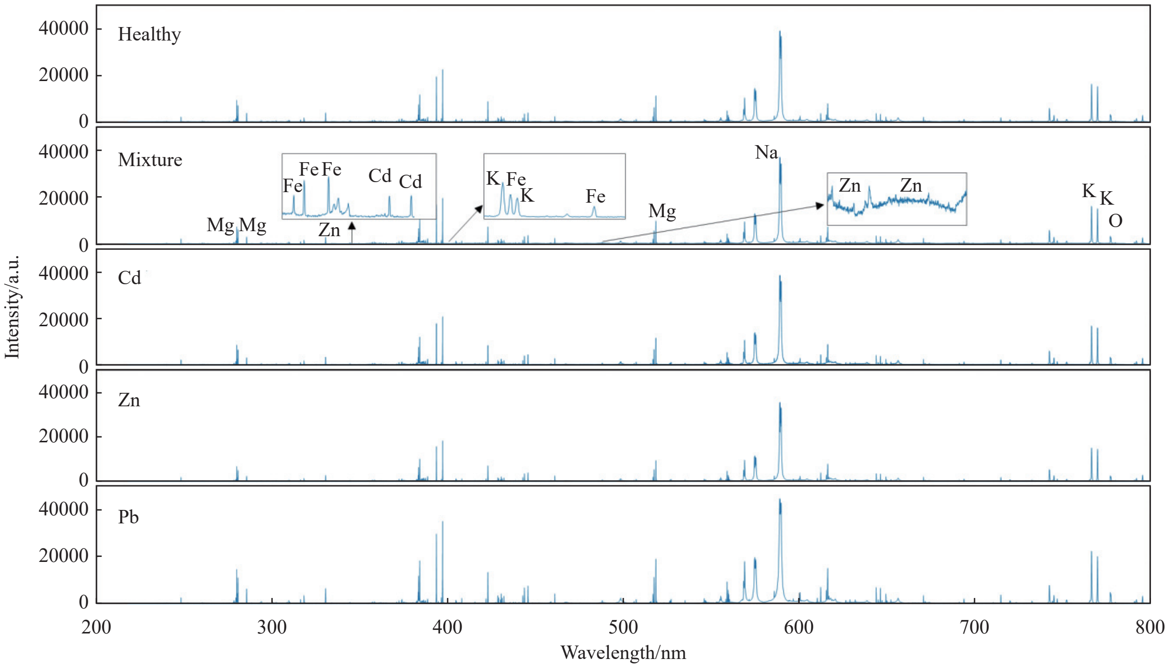

Figure 1 Average LIBS spectra of *T. granosa* samples in each group

4.2 Analysis of one-class classification results

The LIBS of *T. granosa* samples were preprocessed, eliminating spectral variables with intensity less than 0. The spectra were denoised with wavelet filtering, and a 150×29 620 matrix was obtained. The healthy and heavy metal contaminated *T. granosa* samples were regarded as the target class and abnormal class respectively. The LIBS of *T. granosa* was studied by principal component analysis. The projection direction was determined by the target class samples and the first 8 principal components were retained for analysis, which contained 99.57% variation information of target classes. The principal component scores of all *T. granosa* groups are shown in Figure 2. It could be seen that there was significant overlap between different groups in the first 8 principal component scores. Therefore, it was not feasible to directly use the principal component scores as the extracted features when using a one-class classification method for heavy metal detection of *T. granosa* by LIBS. Therefore, we introduced the residual of each sample into the feature set as supplementary information.

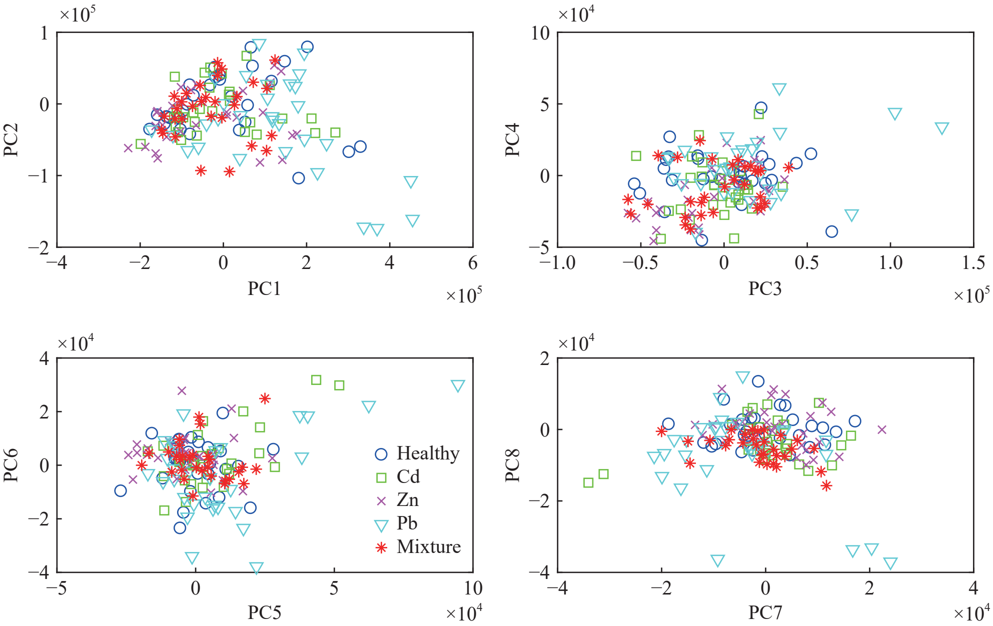

Figure 2 Principal component scores of LIBS spectra of *T. granosa* samples

The effect of one-class classification methods on heavy metal contamination detection of *T. granosa* by LIBS was analyzed as below. According to the Kennard-Stone algorithm[28], two-third of the target class samples were selected as the training set and the remaining one-third of the target class samples and all abnormal class samples were taken as the test set. The one-class classification methods of DD_SIMCA, and SVDD were used to evaluate the performance of the proposed method. Sensitivity (*Sn*), specificity (*Sp*) and accuracy (*Acc*) were used as evaluation indexes, which were defined as

$$Sn = \frac{TP}{TP+FN} \tag{7}$$

$$Sp = \frac{TN}{TN+FP} \tag{8}$$

$$Acc = \frac{Sn+Sp}{2} \tag{9}$$

where, *TP*, *TN*, *FP*, and *FN* were true positive, true negative, false positive, and false negative, respectively.

PCAMA, DD_SIMCA and OCPLS dealt with high-dimensional features by extracting potential variables, while SVDD dealt with this problem by the kernel method. PCAMA and DD_SIMCA extracted potential variables in similar ways, both of which utilized the information of principal component and residual spectrum. OCPLS performed feature extraction by projecting the sample to the center of the training set, which only retained the information of the main space and discards the information of the residual space. Hence, OCPLS was significantly different from the feature reduction strategies of PCAMA and DD_SIMCA.

The experimental results of the four one-class classification methods for heavy metal contamination detection of *T. granosa* by LIBS are listed in Table 1. PCAMA had the best balance in sensitivity and specificity, which were stable at about 1 and 0.9 respectively. Thus, PCAMA maintained a high accuracy rate above 0.9 and achieved the highest accuracy of 0.9667. For DD_SIMCA, the increase in the number of extracted latent variables led to a significant decline in sensitivity, though the specificity was increased. In the optimal number of extracted latent variables, DD_SIMCA achieved the highest accuracy of 0.8417, which was far lower than the highest accuracy of PCAMA. The sensitivity of OCPLS was maintained at the level of 0.9, but its specificity was at a very low level, resulting in an accuracy of only about 0.5. This was because, for the LIBS of *T. granosa* samples, there was little difference between target class and abnormal class, which led to a serious overlap between them in the main space. The discrimination indicators of OCPLS were based on the main space, which severely restricted the identification effect of *T. granosa* contaminated by heavy metals. For the PCAMA and DD_SIMCA, feature extraction was performed on both the main space and the residual space, which was the main reason why DD_SIMCA and PCAMA were better than OCPLS. For the SVDD, the sensitivity and specificity were changed dramatically with the change of the Gaussian kernel parameter, but the balance between the two was not good. When the Gaussian kernel parameter was set to 10^3, the highest accuracy of 0.85 was achieved by the SVDD, which was equivalent to the best result of DD_SIMCA. For SVDD, the kernel method was applied to construct the one-class classifier on high-dimensional data, which was easy to overfit and susceptible to interference from redundant information and resulted in poor separability of the samples (target class and abnormal class). Therefore, using one-class classification methods for heavy metal contamination detection of *T. granosa* by LIBS, feature reduction was a necessary means to improve the generalization of one-class classifiers. PCAMA and DD_SIMCA can retain the feature information of the global feature space to a greater extent than OCPLS to achieve better specificity for abnormal samples.

Table 1 One-class classification results on LIBS spectra of *Tegillarca granosa* samples

Methods	Parameters[a]	*Sn*	*Sp*	*Acc*
PCAMA	1	1.0000	0.9333	0.9667
	2	1.0000	0.9333	0.9667
	3	1.0000	0.9167	0.9583
	4	1.0000	0.9167	0.9583
	5	1.0000	0.8667	0.9333
DD_SIMCA	1	1.0000	0.0667	0.5333
	3	1.0000	0.2417	0.6208
	5	1.0000	0.4333	0.7167
	7	0.9000	0.7833	0.8417
	9	0.8000	0.7750	0.7875
	11	0.7000	0.8167	0.7583
OCPLS	3	0.9000	0.0583	0.4792
	4	0.9000	0.0417	0.4708
	5	0.9000	0.0333	0.5167
SVDD	10^3	0.7000	1.0000	0.8500
	10^4	0.1000	1.0000	0.5500
	10^5	1.0000	0.6000	0.8000
	10^6	1.0000	0.0667	0.5333

[a] Number of extracted latent variables for PCAMA, DD_SIMCA, and OCPLS or Gaussian kernel parameter settings for SVDD.

The optimal models of PCAMA and DD_SIMCA were selected for further analysis and the discrimination results of the samples are listed in Table 2. It can be seen that the PCAMA misjudged the samples contaminated by Cd only, while DD_SIMCA misjudged the samples of each group. Especially, DD_SIMCA suffers a low accuracy rate on the group I (Pb), II (Cd), and III (Zn). The similarity evaluation of these samples given by PCAMA is shown in Figure 3. Because the Mahalanobis distances for each group can vary greatly. To make it clear, we place three subplots in Figure 3. The top subplot shows all the groups except for group Mixture, whose Mahalanobis distances are generally the highest. The bottom subplot on the left shows the Mahalanobis distances for group Healthy while the right subplot is for group Zn. The PCAMA was able to distinguish the distribution differences between different groups. In particular, a big difference was observed between group V (Healthy) and group IV (Mixture), which was cross-contaminated by a variety of heavy metals and resulted in the most serious abnormality. Except for group IV, the other groups are relatively similar to each other. For this reason, DD_SIMCA has trouble distinguishing groups I, II, III from group V and consequently has a high misclassification rate. Group II is closest to group V, which explains why 8 samples from group II are misclassified as group V. On the other hand, group IV is far different from group V. Hence, even DD_SIMCA can easily tell them from each other. Also, from Figure 3, it can be seen that PCAMA not only effectively distinguished normal and abnormal samples but also reasonably evaluated the abnormal degree of samples. For example, group IV has the largest variance of the Mahalanobis distance. Hence, this group is most outlying with respect to group V.

Table 2 Optimal discrimination results of PCAMA and DD_SIMCA

Actual class	Predicted class with PCAMA (3 latent variables)		Predicted class with DD_SIMCA (7 latent variables)	
	Healthy	Outlier	Healthy	Outlier
Healthy	10	0	9	1
Cd	8	22	10	20
Zn	0	30	11	19
Pb	0	30	11	19
Mix	0	30	4	26

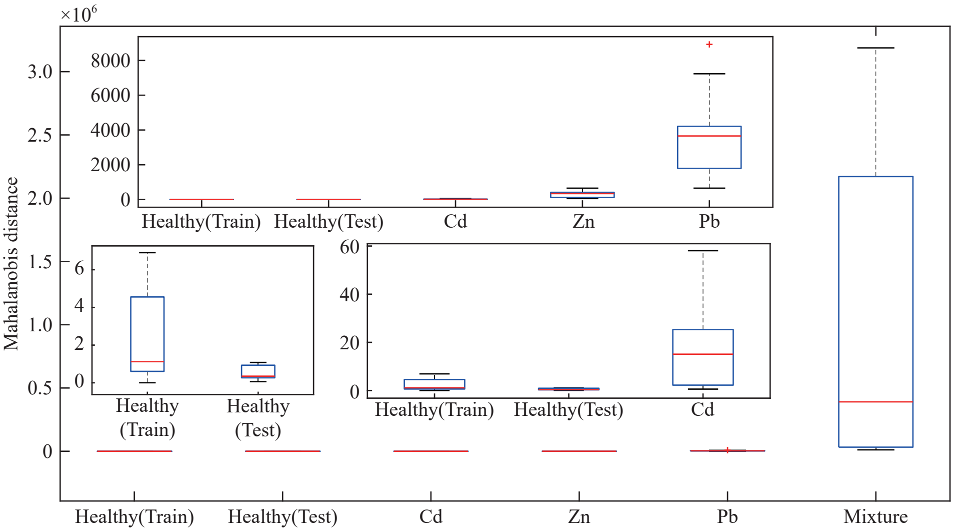

Figure 3 Box plot of similarity evaluation results of PCAMA (3 latent variables) on *T. granosa* samples by LIBS

The feature reduction methods of PCMA and DD_SIMCA are similar, but the difference in the way the classifier was constructed. PCMA directly merged the features of the main space and the residual space into new features to construct the one-class classifier based on Mahalanobis distance. To avoid overfitting, the leave-one-out cross-validation method was adopted to calculate extracted features and Mahalanobis distance of training set, and the threshold of the target class was estimated by quantile. However, DD_SIMCA calculated the orthogonal score and residual score of the training set directly and assumed that the two obey the chi-square distribution, which was prone to large deviations in the case of small samples. According to the experimental results in Tables 1 and 2, PCAMA had better robustness than DD_SIMCA in the number of extracted latent variables. Parameter optimization of the one-class classifier was difficult due to the lack of abnormal class samples. Therefore, for the PCAMA, the requirement for parameter optimization was weakened to achieve better generalization than DD_SIMCA. Besides, the modeling complexity of DD_SIMCA was significantly higher than that of PCAMA. The extracted three latent variables were taken to construct a model as an example. On the researcher's computer (i7-8750H CPU @ 2.20 GHz, 16.0 GB RAM), the PCAMA was constructed in 1.23 seconds, while the construction of DD_SIMCA took 166.84 seconds and consumed 135.24 times as long as PCAMA.

5 Conclusions

A novel one-class classification method was proposed based on PCA and Mahalanobis distance and applied to detect heavy metal contamination in *T. granosa* by LIBS. In this method, the principal component score and the intensity of the residual spectrum were retained and combined as the extracted features. For the extracted features, Mahalanobis distance was used as the evaluation index of sample similarity. During the threshold setting, the leaving-one-out cross-validation was used to calculate the extracted features and Mahalanobis distance of the training set to ensure the generalization. For the detection of heavy metal in *T. granosa* by LIBS, the detection accuracy of the proposed method was significantly better than the reported one-class classification methods. With a small number of extracted latent variables, a good balance between sensitivity and specificity was achieved. In summary, the proposed method is suitable for heavy metals detection in *T. granosa* by using LIBS and has reference significance for one-class classification problems of other high-dimensional and small sample data.

Acknowledgements

This study was financially supported by the Zhejiang Natural Science Foundation of China (Grant No. LY21C200001, LY20F030019) and National Natural Science Foundation of China (Grant No. 62105245, 62071386), Wenzhou Major Scientific and Technological Innovation Projects of China (Grant No. ZG2021029, ZY2021027) and the Wenzhou Science and Technology Bureau General Project (Grant No. S2020011).

[References]

[1] Zhang M, Sun X, Xu J L. Heavy metal pollution in the East China Sea: A review. Marine Pollution Bulletin 2020; 159: 111473. doi: 10.1016/j.marpolbul.2020.111473.

[2] Briffa J, Sinagra E, Blundell R. Heavy metal pollution in the environment and their toxicological effects on humans. Heliyon 2020; 6(9): e04691. doi: 10.1016/j.heliyon.2020.e04691.

[3] Riza S, Gevisioner G, Suprijanto J, Widowati I, Putra I, Effendi I. Farming and food safety analysis of blood cockles (*Anadara granosa*) from Rokan Hilir, Riau, Indonesia. Aquaculture, Aquarium, Conservation & Legislation 2021; 14: 814–812.

[4] Chen X J, Liu K, Cai J B, Zhu D H, Chen H L. Identification of heavy metal-contaminated *Tegillarca granosa* using infrared spectroscopy. Analytical Methods, 2015; 7(5): 2172–2181.

[5] Yuan L M, Chen X J, Lai Y J, Chen X, Shi Y J, Zhu D H, et al. A Novel strategy of clustering informative variables for quantitative analysis of potential toxics element in Tegillarca granosa using laser-induced breakdown spectroscopy. Food Analytical Methods, 2018; 11(5): 1405–1416.

[6] Ipeaiyeda A, Ayoade A. Flame atomic absorption spectrometric determination of heavy metals in aqueous solution and surface water preceded by co-precipitation procedure with copper (II) 8-hydroxyquinoline. Applied Water Science 2017; 7: 4449–4459.

[7] Suo L Z, Dong X Y, Gao X, Xu J F, Huang Z, Ye J, et al. Silica-coated magnetic graphene oxide nanocomposite based magnetic solid phase extraction of trace amounts of heavy metals in water samples prior to determination by inductively coupled plasma mass spectrometry. Microchemical Journal, 2019; 149: 104039.

[8] Vinhal J O, Cassella R J. Novel extraction induced by microemulsion breaking for Cu, Ni, Pb and V determination in ethanol-containing gasoline by graphite furnace atomic absorption spectrometry. Spectrochimica Acta Part B: Atomic Spectroscopy, 2019; 151: 33–40.

[9] Fu Y T, Gu W L, Hou Z Y, Muhammed S A, Li T Q, Wang Y, et al. Mechanism of signal uncertainty generation for laser-induced breakdown spectroscopy. Frontiers of Physics, 2021; 16(2): 1–10.

[10] Ji G L, Ye P C, Shi Y J, Yuan L M, Chen X J, Yuan M S, et al. Laser-induced breakdown spectroscopy for rapid discrimination of heavy-metal-contaminated seafood Tegillarca granosa. Sensors, 2017; 17(11): 2655.

[11] Xie Z H, Meng L W, Chen X J, Chen X, Yuan L M, Shi W, et al . Identification of heavy metal-contaminated Tegillarca granosa using laser-induced breakdown spectroscopy and linear regression for classification. Plasma Science and Technology, 2020; 22(8): 085503.

[12] Chen H, Tan C, Li H J. Untargeted identification of adulterated Sanqi powder by near-infrared spectroscopy and one-class model. Journal of Food Composition and Analysis, 2020; 88: 103450.

[13] Santana F B, Neto W B, Popp.R J. Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. Food chemistry, 2019; 293: 323–332.

[14] Esteki M, Simal-Gandara J, Shahsavari Z, Zandbaaf S, Dashtaki E, Vander Heyden Y. A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. Food Control, 2018; 93: 165–182.

[15] Oliveri P. Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues–A tutorial. Analytica chimica acta, 2017; 982: 9–19.

[16] Huang G Z, Yang Z J, Chen X J, Ji G L. An innovative one-class least squares support vector machine model based on continuous cognition. Knowledge-Based Systems, 2017; 123: 217–228.

[17] Tax D M J. One-class classification. Applied Sciences, 2001; 212p.

[18] Oza P, Patel V M. One-class convolutional neural network. IEEE Signal Processing Letters, 2018; 26(2): 277–281.

[19] Wetzel S J. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. Physical Review E, 2017; 96(2): 022140.

[20] Ghorbani H. Mahalanobis distance and its application for detecting multivariate outliers. Facta Universitatis-Series Mathematics and Informatics, 2019; 34: 583–595.

[21] Zontov Y, Rodionova O Y, Pomerantsev A, Kucheryavskiy S. DD-SIMCA–a Matlab GUI tool for data driven SIMCA approach. Chemometrics and Intelligent Laboratory Systems, 2017; 167: 23–28.

[22] Xu L, Yan S M, Cai C B, Yu X P. One-class partial least squares (OCPLS) classifier. Chemometrics & Intelligent Laboratory Systems, 2013; 126: 1–5.

[23] Xu L, Goodarzi M, Shi W, Cai C B, Jiang J H. A Matlab toolbox for class modeling using one-class partial least squares (OCPLS) classifiers. Chemometrics and Intelligent Laboratory Systems, 2014; 139: 58–63.

[24] Xie Z H, Feng X A, Chen X J. Partial least trimmed squares regression. Chemometrics and Intelligent Laboratory Systems, 2022; 221: 104486.

[25] Xie Z H, Feng X A, Chen X J. Subsampling for partial least-squares regression via an influence function. Knowledge-Based Systems, 2022; 245: 108661.

[26] Tax D. DDtools, the Data Description Toolbox for Matlab, 2014. Software available at http://prlab tudelft nl/david-tax/dd_tools html, 2015

[27] Tao X M, Chen W, Li X K, Zhang X H, Li Y T, et al. The ensemble of density-sensitive SVDD classifier based on maximum soft margin for imbalanced datasets. Knowledge-Based Systems, 2021; 219: 106897.

[28] Chen W H, Chen H Z, Feng Q X, Mo L N, Hong S Y. A hybrid optimization method for sample partitioning in near-infrared analysis. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021; 248: 119182.