Detection of adulteration by hydrolysed leather protein in infant formula based on least squares support vector machine and near-infrared spectroscopy

YISEN LIU – SONGBIN ZHOU – WEI HAN – CHANG LI – KEJIA HUANG – WEIXIN LIU

Summary

Near-infrared (NIR) spectroscopy and least squares support vector machine (LS-SVM) were applied to detect hydrolysed leather protein (HLP) adulteration in infant formula. In the classification approach, principal components (PCs) extracted by principal component analysis (PCA) were used as the inputs of the LS-SVM class model. The correct classification rates of authentic samples and adulterated samples in validation set were 96.7 % and 95.6 %, respectively. In the regression approach, PCs extracted by PCA were also utilized as inputs of LS-SVM to develop PCs-LS-SVM regression models. The optimal PCs-LS-SVM model was obtained with 4 PCs and the root mean squared error of calibration and prediction were 1.02 g·kg^{-1} and 1.46 g·kg^{-1}, respectively. Moreover, sensitive wavelengths (SWs) were selected according to the regression coefficients obtained by partial least squares regression model, and utilized as the input data of LS-SVM to develop SWs-LS-SVM regression models. The optimal SWs-LS-SVM model was obtained with 19 SWs and the root mean squared error of calibration and prediction were 0.64 g·kg^{-1} and 1.18 g·kg^{-1}, respectively. The results demonstrated that NIR spectroscopy combined with LS-SVM could be applied as a fast, simple and non-destructive way to detect HLP adulteration in infant formula.

Keywords

hydrolysed leather protein; infant formula adulteration; least squares support vector machine; near-infrared spectroscopy

Milk and infant formula adulteration have been of high concern for a long time, especially after the global food safety scares caused by melamine adulteration in 2008 [1–3]. In 2011, it was reported that a new type of adulterant, hydrolysed leather protein (HLP), was intentionally added into dairy products by some illegal factories, in order to boost the apparent protein content [1, 4]. HLP is collagen that is extracted from leather scraps. As an adulterant in dairy products, it may cause a great risk for osteoporosis and heavy metal (Cr) poisoning, for the reason that a large amount of potassium dichromate and sodium dichromate are added during the tanning and dyeing process of leather.

Although HLP adulteration has drawn great attention from the Chinese government and consumers, there are only a few scientific reports concentrating on the detection method of HLP in dairy products. Current analytical methods for detecting HLP are mainly chromatography-based methods [1, 5, 6], through detecting hydroxyproline (Hyp) as a marker of the collagen content. These methods are time-consuming, expensive, requiring complicated sample pre-treatment and well-trained technicians. Therefore, there is an urgent and increasing need to develop a quick, simple and economical method for detecting HLP in dairy products.

Near-infrared (NIR) spectroscopy is an analytical technique that has advantages in real-time response, simplicity in testing, relatively low-cost instrument, non-destructive and environmentally friendly analysis. NIR spectroscopy has been widely used in numerous fields, such as nutrient content analysis in food industries [7–13], process monitoring in pharmaceutics [14], composition control in petroleum industries [15, 16] and also in

Yisen Liu, Songbin Zhou, Wei Han, Chang Li, Kejia Huang, Weixin Liu, Guangdong Key Laboratory of Modern Control Technology, Guangdong Public Laboratory of Modern Control and Manufacturing Technology, Guangdong Institute of Intelligent Manufacturing, Xianliezhong Road No. 100, 510070 Gangzhou, China.

Correspondence author:
Songbin Zhou, e-mail: zhousongbin728@126.com, tel.: +8620 87685580

adulteration detection of dairy products [17, 18]. For instance, MAUER et al. detected melamine in infant formula on the basis of near- and mid-infrared (MIR) spectroscopy [19]. BORIN et al. detected starch, whey and saccharose adulteration in milk powder by NIR spectroscopy [20].

So far, quite a few efforts have been made on chemometric algorithms to achieve better detection accuracy of dairy products adulteration. BALABIN and SMIRNOV [21] compared several calibration methods, including partial least squares regression (PLS), polynomial PLS, artificial neural network (ANN), support vector regression (SVR) and least squares support vector machine (LS-SVM) to detect melamine in dairy products. They found that the relationship between NIR/MIR spectrum of milk products and melamine content was non-linear, and the non-linear regression methods (SVR, LS-SVR, ANN) had better predicted accuracy for melamine content. Support vector machine (SVM) is a promising method proposed by VAPNIK [22], which can perform non-linear classification and multivariate function estimation or non-linear regression. LS-SVM, proposed by SUYKENS et al. [23], is an optimized algorithm based on SVM, which applies a linear set of equations instead of quadratic programming used in standard SVM [24]. This can reduce the computational complexity as well as improve the predicted accuracy of the model. The combination of NIR spectroscopy and LS-SVM was used in food brand discrimination [25], nutrient content analysis [26–28] and several other applications [29, 30].

In the classification part, principal components (PCs) extracted by principal component analysis (PCA) were used as the input of LS-SVM to establish the PCs-LS-SVM classification model.

In the regression part, two types of regression model were established and compared, namely, PCs-LS-SVM and sensitive wavelengths (SWs) LS-SVM. In these two methods, PCs extracted by PCA and SWs selected by PLS were used as the input data of the LS-SVM regression model, respectively.

The present study aimed at evaluating the performance of NIR spectroscopy and LS-SVM for HLP adulteration detection in infant formula.

MATERIALS AND METHODS

Samples

A total of 150 samples, including 60 pure infant formula samples and 90 adulterated samples, were analysed.

In order to improve the model robustness, samples in this work contained infant formula of three brands: Wyeth (Dallas, Texas, USA), Mead Johnson (Chicago, Illinois, USA) and Beingmate (Hangzhou, China) and their mixtures with random proportions [21]. All the infant formulae were obtained from a local food safety supervision organization (Guangzhou Quality Supervision and Testing Institute, Guangzhou, China). HLP powder was purchased from three producers: Kaitai (Beijing, China), Cargill (Hamburg, Germany) and AccoBio (Wuxi, China). Hyp powder was purchased from Sigma (St. Louis, Missouri, USA).

Adulteration contents of HLP ranged from 1.0 $g \cdot kg^{-1}$ to 10.0 $g \cdot kg^{-1}$. All of the adulterated samples were well stirred before NIR spectroscopy measurements.

Near-infrared spectroscopy measurement

The NIR spectra were acquired with a handheld NIR analyser DLP NIRscan Nano (Texas Instruments, Dallas, Texas, USA). The spectra ranged from 900 nm to 1 700 nm (11 00 cm^{-1} to 5 880 cm^{-1}) with the scanning resolution of 2.8 nm (18 cm^{-1}). Sixteen diffuse reflectance scans were averaged for each spectrum. All of the measurements were taken at room temperature (24–27 °C).

Data processing

In order to improve the calibration accuracy, the initial and terminal sections of 100 nm of the spectra were deleted because rather high instrument noise could be observed at these regions. Therefore, the spectra ranging from 1 000 nm to 1 600 nm were used for analysis. Prior to building the class and regression models, pre-treatment of first-order gap segment derivative followed by the multiplicative scatter correction and first derivative was applied for denoising.

Before class modelling, PCA was applied to evaluate the possibility of discrimination between pure infant formula and infant formula adulterated with HLP. To improve the model accuracy and reduce the complexity of computing, PCs were extracted by PCA and then used as the input data of the LS-SVM classification model. In the PCA analysis and class modelling, half of the samples were used for calibration, while the others were used for validation.

In the part of quantitative analysis, PCs-LS-SVM and SWs-LS-SVM models were used to predict the content of HLP in adulteration samples. PCs-LS-SVM regression model utilized PCs extracted by PCA as input data, which was the same as in the classification model. On the other hand, SWs with higher absolute values of regres-

sion coefficients in PLS were selected and used as the input data in the SWs-LS-SVM regression model.

The root mean squared error of calibration ($RMSE_C$) was used to evaluate the calibration results, while the root mean squared error of leave-one-out cross validation ($RMSE_{CV}$) was used to select parameters of the models, such as the number of PCs, the number of SWs, the kernel function parameter σ^2 and the regularization parameter γ in LS-SVM model. The root mean squared error of prediction ($RMSE_P$) and the determination coefficient of validation set (R^2) were used to evaluate the prediction capacity of each regression model.

In this study, LS-SVM models performed in MATLAB R2015a (Math Works, Natick, Massachusetts, USA). The free LS-SVM toolbox was applied with MATLAB to derive all of the LS-SVM models. The spectra pre-treatment, PCA and regression coefficients calculation were done in Unscrambler ver. 10.3 (CAMO, Oslo, Norway).

Least squares support vector machine

The LS-SVM algorithm is introduced as follows. LS-SVM maps the input data to a higher dimensional feature space:

$$y(x) = w^T\varphi(x) + b \tag{1}$$

where w is the weight vector, and b is the bias. Then the optimization problem is to minimize a cost function (C):

$$C = \frac{1}{2}w^T w + \frac{1}{2}\gamma\sum_{i=1}^{N} e_i^2 \tag{2}$$

subjecting to the constraints:

$$y_i = w^T\varphi(x) + b + e_i \quad i = 1, 2, \ldots, N \tag{3}$$

In Eq. 2, γ is the regularization parameter and e_i is the random error.

In LS-SVM algorithm, Lagrange function is used to solve this optimization problem:

$$L = \frac{1}{2}\|w\|^2 + \gamma\sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N}\alpha_i\{w^T\varphi(x) + b + e_i - y_i\} \tag{4}$$

where α_i is Lagrange multipliers named support value. In order to obtain the optimum, the partial first derivative of each variable was set to zero:

$$\begin{cases} \dfrac{\partial L}{\partial w} = w - \sum_{i=1}^{N}\alpha_i\varphi(x_i) = 0 \Rightarrow w = \sum_{i=1}^{N}\alpha_i\varphi(x_i) \\ \dfrac{\partial L}{\partial e} = \sum_{i=1}^{N}\gamma e_i - \alpha_i = 0 \Rightarrow \alpha_i = \gamma e_i \end{cases} \tag{5}$$

It can be known from Eq. 5 that:

$$w = \sum_{i=1}^{N}\alpha_i\varphi(x_i) = \sum_{i=1}^{N}\gamma e_i\,\varphi(x_i) \tag{6}$$

Combining Eq. 1 with Eq. 6, the following result is obtained:

$$\begin{aligned} y &= \sum_{i=1}^{N}\alpha_i\varphi(x_i)^T\varphi(x) + b = \\ &= \sum_{i=1}^{N}\alpha_i\langle\varphi(x_i)^T, \varphi(x)\rangle + b \end{aligned} \tag{7}$$

The problem is converted into solving a set of linear equations:

$$\begin{bmatrix} K + \dfrac{1}{\gamma} & \vec{1} \\ \vec{1}^T & 0 \end{bmatrix}\begin{bmatrix}\alpha \\ b\end{bmatrix} = \begin{bmatrix} y \\ 0\end{bmatrix} \tag{8}$$

in which

$$\begin{cases} y = [y_1, y_2, \ldots, y_N] \\ \vec{1} = [1, \ldots, 1]^T \\ \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N] \\ K = (x_i, y_i) = \varphi(x_i)^T\varphi(x_j) \end{cases} \tag{9}$$

In Eq. 8, K is the kernel function that follows Mercer's condition. Then the LS-SVM regression model can be written as:

$$y(x) = \sum_{i}^{N}\alpha_i K(x, x_i) + b \tag{10}$$

Commonly used kernel function includes radial basis function (RBF), polynomial functions and linear function. In this paper, RBF was chosen as the kernel function. It can be written as:

$$K(x, x_i) = \exp\left(-\|x_i - x_j\|^2/\sigma^2\right) \tag{11}$$

where σ^2 is the bandwidth of the RBF function.

In LS-SVM, the RBF kernel function parameter σ^2 and the regularization parameter γ should be selected carefully by the users because these two parameters play an important role in obtaining high model accuracy, meanwhile avoiding over-fitting.

RESULTS AND DISCUSSION

Classification modelling

The pre-treated spectra of samples adulterated with HLP are shown in Fig. 1. The pre-treatment method was the first-order gap segment derivative followed by the multiplicative scatter correction, which helped to eliminate the influence of base-

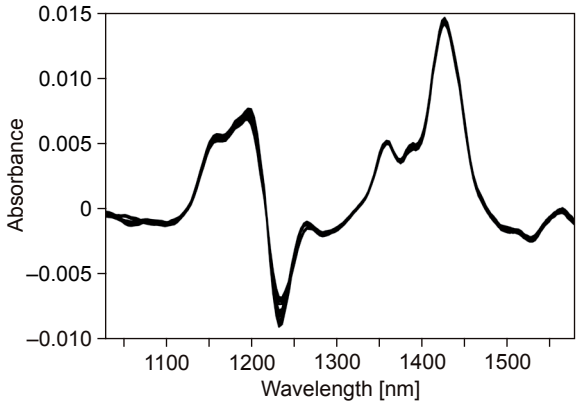

Fig. 1. Near-infrared spectroscopy spectra of samples adulterated with hydrolysed leather protein after pre-treatment of multiplicative scatter correction and first-order gap segment derivative.

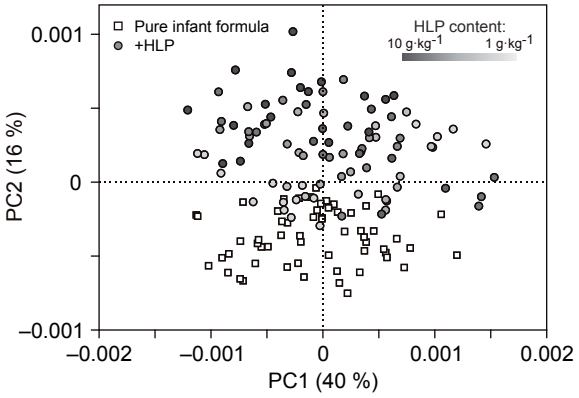

Fig. 2. Principal component analysis score plot of the pure infant formula samples and samples adulterated with hydrolysed leather protein.

HLP – hydrolysed leather protein.

line drift and particle scattering.

In this work, PCA was performed based on the pre-treated NIR spectra of 60 pure infant formula samples and 90 samples adulterated with HLP. The aim of using PCA was to evaluate the potential of discrimination between authentic and adulterated samples, and to extract PCs as the input data of LS-SVM classification model. In the PCA analysis and subsequent class modelling, half of the samples were utilized as calibration sets, while the remaining half of them were assigned into validation sets. The first and second principal component scores are plotted in Fig. 2. PC1 and PC2 explain 56 % of the total variance. As shown in Fig. 2, the pure infant formula samples and samples adulterated with HLP exhibited obvious differences in PC2 and a potential of clustering, although a particle overlap between these two classes could be observed. The contents of HLP in adulteration samples are marked by different shades of gray in Fig. 2. However, as can be seen, the change of HLP contents with the change of PC2 was not-obvious.

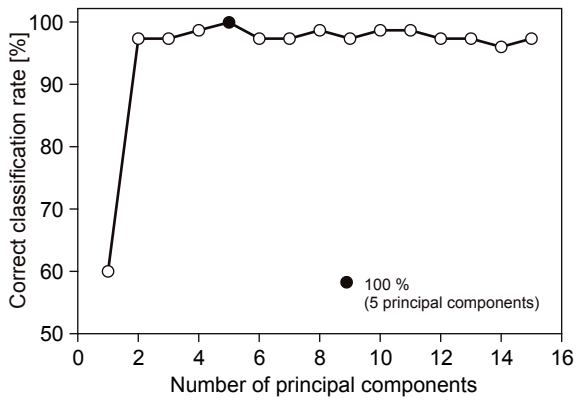

Fig. 3. Effect of the principal component number on the correct classification rates of cross-validation.

Tab. 1. Correct classification rates of the least squares support vector machine classification model.

	Calibration	Cross-validation	Prediction
Authentic	100 %	100 %	96.7 %
Adulterated	100 %	100 %	95.6 %

In LS-SVM, several crucial parameters need to be carefully selected, including the kernel function, the kernel function parameter σ^2 and the regularization parameter γ, as well as the optimal input features. In this study, the non-linear kernel function RBF was selected, and the parameters of σ^2 and γ were optimized by two-dimensional grid search. The grid search was performed on the basis of leave-one-out cross-validation and evaluated by the correct classification rate in cross-validation.

Regarding input features, LS-SVM classification model was built up with PCs extracted by PCA as input data. The number of PCs (1 to 15) was optimized through testing its influence on correct classification rate of leave-one-out cross-validation. According to the results presented in Fig. 3, the cross validation achieved the best correct classification rate (100 %) with a number of PCs of 5. The first 5 PCs were selected as the inputs to build PCs-LS-SVM classification model, and the results are shown in Tab. 1. As can be seen in the table, for the calibration set and cross validation, the correct classification rates

of authentic group and adulterated group were both 100 %, while, for the validation set, the correct classification rates of the authentic group and the adulterated group were 96.7 % and 95.6 %, respectively.

It can be observed from Fig. 2 and Fig. 3 that PC2 played an important role in detecting the presence of HLP. Loadings of each PC reflect how much the individual variable contributes to that PC. Therefore, loadings of PC2 (shown in Fig. 4) can be a potential indicator for important wavelengths. It can be seen in Fig. 4 that wavelengths between 1 392 nm and 1410 nm, and between 1 426 nm and 1 450 nm had large absolute values of PC2 loadings and should have been seen as of great importance.

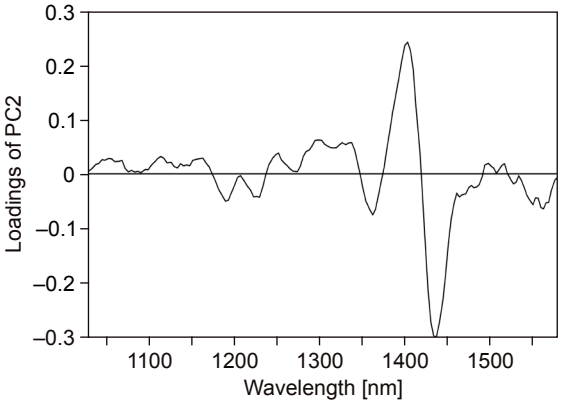

Fig. 4. Loading curve of the second principal component.

HLP is a collagen extracted from leather scraps, and Hyp ($C_5H_9NO_3$) is a unique component of collagen. It is often used as a marker of collagen content in traditional analytical methods. The content of Hyp in HLP is about 9 %. To understand what chemical composition determined the change of the important NIR band (1 392–1 450 nm), the average spectrum of pure infant formula, the average spectrum of pure HLP powder and the spectrum of Hyp powder were measured and shown in Fig. 5A. It can be observed that, in the important region between 1 392 nm and 1 450 nm (marked by grey vertical lines in Fig. 5A), the pure infant formula powder and the HLP powder had similar shape in spectrum, while the spectrum of Hyp powder showed different trend in this region. The difference spectrum of the Hyp and the pure infant formula (marked as Hyp – pure infant formula in Fig. 5B) was calculated after adjusting the spectra to the same baseline. The loadings of PC2 between 1 392 nm and 1 450 nm are also shown as grey bars in Fig. 5B. It can be clearly seen that the difference spectrum of Hyp and the pure infant formula had a quite similar trend with the PC2 loadings in this region. Therefore, we deduced that the spectrum change between 1 392 nm and 1 450 nm in adulterated samples could be mainly attributed to the presence of Hyp, and Hyp is also essential for detecting HLP in infant formula by NIR spectroscopy.

Regression modelling

In the regression modelling approach, 120 samples, including 90 HLP adulterated samples and 30 adulteration-free samples, were involved. Among these samples, 25 pure infant formula samples and 70 HLP adulterated samples were

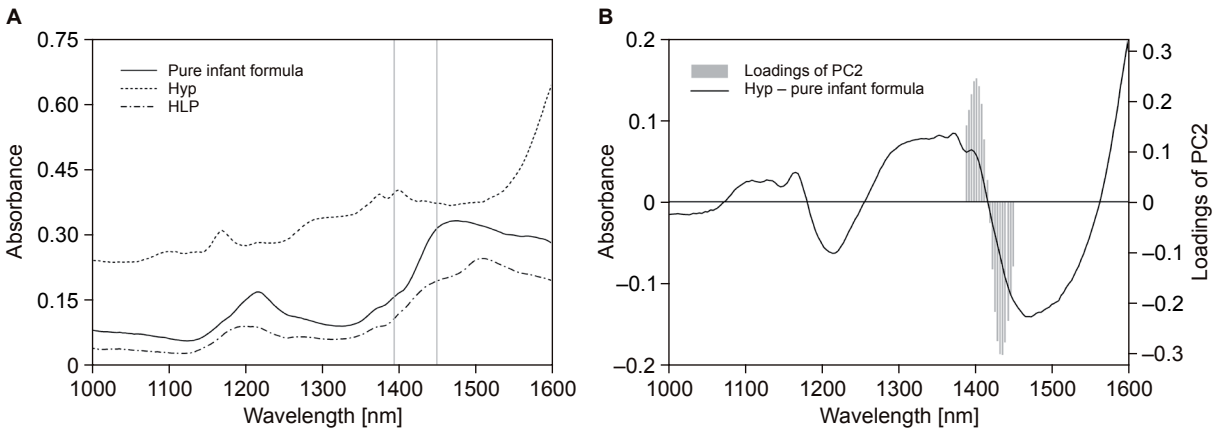

Fig. 5. The impact of hydroxyproline on spectra of infant formula.

A – original spectra of the pure infant formula, hydrolysed leather protein and hydroxyproline, B – difference spectrum of hydroxyproline and of the pure infant formula, and the loadings of the second principal component.
Grey vertical lines and grey bars mark region between 1 392 nm and 1 450 nm.
HLP – hydrolysed leather protein, Hyp – hydroxyproline.

Tab. 2. Description of datasets used in the regression models.

	Number of samples			HLP content [g·kg^{-1}]		
	Total	Authentic	Adulterated	Range	Mean	*SD*
Calibration	95	25	70	0–10.0	4.1	3.5
Validation	25	5	20	0-10.0	4.4	3.4

HLP – hydrolysed leather protein, *SD* – standard deviation.

used for calibration, while 5 pure infant formula samples and 20 samples adulterated with HLP were used for validation. The detailed information about the datasets utilized in regression modelling are presented in Tab. 2**.** In each regression model, RBF was adopted as kernel function, and the parameters of σ^2 and γ were optimized by two-dimensional grid searches. The grid search was performed by searching the appropriate σ^2 and γ with minimized $RMSE_{CV}$. Nevertheless, the detailed process of grid search will not be discussed in this paper.

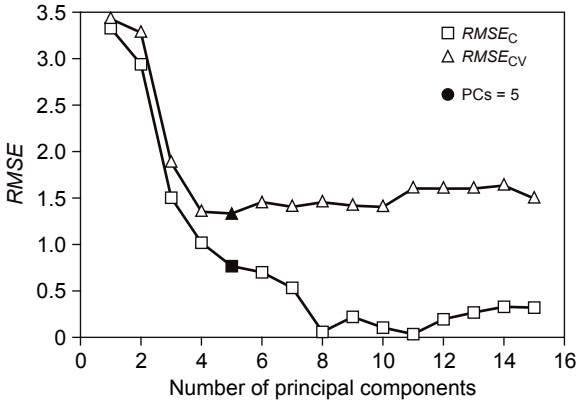

Fig. 6. Influence of the principal components number on root mean squared error of calibration and root mean squared error of cross-validation on the least squares support vector machine model with principal components as input.

$RMSE_C$ – root mean squared error of calibration, $RMSE_{CV}$ – root mean squared error of cross validation, PCs – principal components.

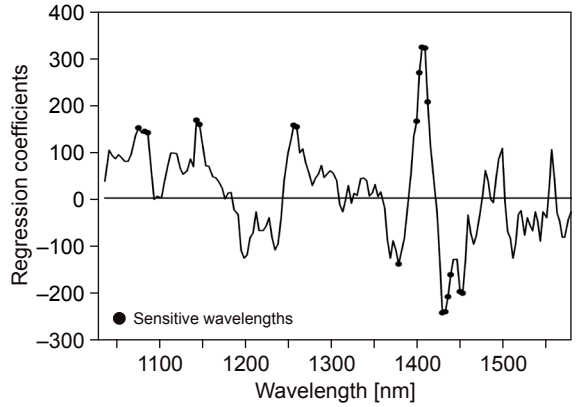

Fig. 7. Regression coefficients of the hydrolysed leather protein content.

Quantification of adulterant contents in infant formula was performed through two types of LS-SVM regression models: PCs-LS-SVM and SWs-LS-SVM. First, PCs extracted by PCA were used as the input data to build PCs-LS-SVM regression model. To keep consistent calibration/validation data partition with the regression modelling, PCA was rebuilt with 95 calibration samples and 25 validation samples. The number of PCs was optimized by evaluating $RMSE_{CV}$ (Fig. 6). As can be seen in Fig. 6, both $RMSE_C$ and $RMSE_{CV}$ decreased dramatically with the increase of PC number from 1 to 4. However, when the PC number was greater than 4, $RMSE_{CV}$ did not show prominent improvement and the difference between $RMSE_C$ and $RMSE_{CV}$ increased obviously, which indicated overfitting. Therefore, the first 4 PCs were selected and the calibration and validation results were listed in Tab. 3. The $RMSE_C$ and $RMSE_P$ values of the PCs-LS-SVM regression model were 1.02 g·kg^{-1} and 1.46 g·kg^{-1}, respectively.

For the SWs-LS-SVM model, SWs were used as the input data of LS-SVM. The regression coefficients of wavelengths were calculated by PLS, and adopted for determining the sensitive wavelength. The size of the regression coefficient represents the importance of *X*-variables on predicting *Y*-variable. The regression coefficients calculated by PLS are shown in Fig. 7. Comparing Fig. 4 with Fig. 7, it can be seen that the loadings of PC2 and the regression coefficient curve had very similar shapes. Large positive regression coefficients (at 1392–1409 nm) indicated a strong positive correlation between the absorbance in spectra and the HLP content, while more negative regression coefficients implied a strong negative correlation (at 1426–1453 nm). Therefore, *X*-variables with large absolute values of regression coefficient were selected as SWs, and used as the input data of the LS-SVM regression model.

Tab. 3. Calibration and validation results for the least squares support vector machine regression models.

	Root mean squared error [g·kg^{-1}]			R^2
	Calibration	Cross-validation	Prediction	Prediction
PCs-LS-SVM	1.02	1.35	1.46	0.814
SWs-LS-SVM	0.64	1.29	1.18	0.880

PCs-LS-SVM – least squares support vector machine model with principal components as input, SWs-LS-SVM – least squares support vector machine model with sensitive wavelengths as input, R^2 – determination coefficient.

In this study, the number of SWs was optimized by cross-validation. According to the absolute values of correlation coefficient from large to small, the *X*-variables were realigned, before being input into the LS-SVM model. Different numbers of the most SWs were used to build SWs-LS-SVM regression models and evaluated by the $RMSE_{CV}$. Fig. 8 shows the effects of the SWs number on $RMSE_C$ and $RMSE_{CV}$. As can be seen in Fig. 8, $RMSE_C$ kept decreasing with the increase of SWs number, while the minimum $RMSE_{CV}$ was achieved at a SWs number of 19 (also labelled in Fig. 7). Therefore, the 19 most SWs were selected as the input data.

The regression results and curve of the SWs-LS-SVM model are shown in Tab. 3 and Fig. 9, respectively. The $RMSE_C$ and $RMSE_P$ values of the SWs-LS-SVM regression model were 0.64 g·kg^{-1} and 1.18 g·kg^{-1}, respectively. By comparing the results in the Tab. 3, it can be seen that the SWs-LS-SVM model outperformed the PCs-LS-SVM model.

CONCLUSIONS

In this paper, qualitative and quantitative analyses of adulteration of infant formula with HLP was performed by NIR spectroscopy in combination with LS-SVM. In the classification approach, PCs extracted by PCA were used as the input data of the LS-SVM class model. The correct classification rates of authentic samples and HLP adulterated samples in validation set were 96.7% and 95.6%, respectively. From the PCA score plot and the variation trend of correct classification rate in cross validation, it could be found that PC2 played an important role in discriminating the presence of HLP in infant formula.

In the regression approach, PCs-LS-SVM and SWs-LS-SVM were used to predict the content of HLP with PCs and SWs as input data, respectively. In the PCs-LS-SVM model, the $RMSE_C$ and $RMSE_P$ values were 1.02 g·kg^{-1} and 1.46 g·kg^{-1}, respectively. In the SWs-LS-SVM model, SWs were selected in accordance with the regression coefficients obtained by PLS model. The $RMSE_C$ and $RMSE_P$ values of the SWs-LS-SVM regression model were 0.64 g·kg^{-1} and 1.18 g·kg^{-1}, respectively, which outperformed the PCs-LS-SVM model.

The overall results demonstrated that NIR spectroscopy combined with LS-SVM could pro-

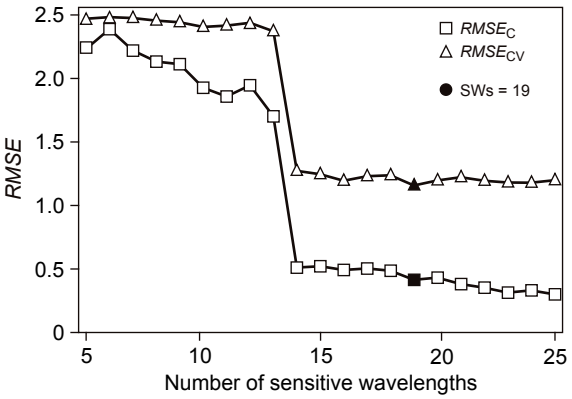

Fig. 8. Influence of the sensitive wavelength number on root mean squared error of calibration and root mean squared error of cross-validation on the least squares support vector machine model with sensitive wavelengths as input.

$RMSE_C$ – root mean squared error of calibration, $RMSE_{CV}$ – root mean squared error of cross validation, SWs – sensitive wavelengths.

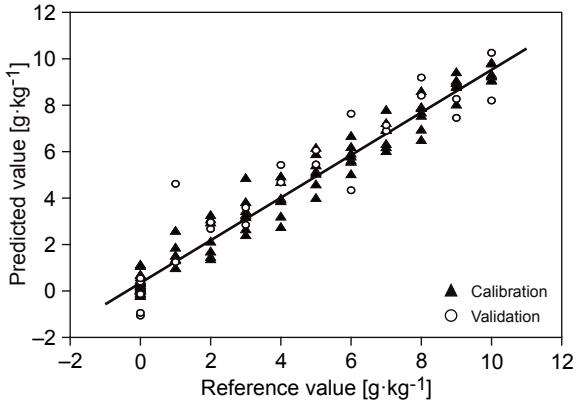

Fig. 9. Regression curves of the least squares support vector machine model with sensitive wavelengths as input data.

vide a simple, rapid, economical and non-destructive detection of adulteration of infant formula with HLP. Based on our results, LS-SVM is a promising technique that can be used for discrimination and quantification of adulteration in dairy products. However, in order to achieve high prediction accuracy and model stability, particular attention has to be always paid to selection of input data and to optimization of parameters.

Acknowledgements

This work was supported by Science and Technology Planning Project of Guangdong Province, China (fund No. 2015B090901025, 2015A010103008, 2016B090918061) and Science and Technology Planning Project of Guangzhou, China (fund No. 201607010093, 201508020071).

REFERENCES

1. Dong, Y. L. – Yan, N. – Li, X. – Zhou, X. M. – Zhou, L. – Zhang, H. J. – Chen, X. G.: Rapid and sensitive determination of hydroxyproline in dairy products using micellar electrokinetic chromatography with laser-induced fluorescence detection. Journal of Chromatography A, *1233*, 2012, pp. 156–160. DOI: 10.1016/j.chroma.2012.02.030.
2. Lim, J. – Kim, G. – Moa, C. – Kim, M. S. – Chao, K. – Qin, J. – Fu, X. – Baek, I. – Cho, B. K.: Detection of melamine in milk powders using near-infrared hyperspectral imaging combined with regression coefficient of partial least square regression model. Talanta, *151*, 2016, pp. 183–191. DOI: 10.1016/j.talanta.2016.01.035.
3. Hanuš, O. – Tomáška, M. – Hofericová, M.– Vyletělová-Klimešová, M. – Klapáčová, L. – Jedelská, R. – Kološta, M.: Relationship between freezing point and raw ewes' milk components as a possible tool for estimation of milk adulteration with added water. Journal of Food and Nutrition Research, *54*, 2015, pp. 281–288. ISSN: 1336-8672 (print), 1338-4260 (online). <http://www.vup.sk/download.php?bulID=1669>
4. Liu, J. X. – Wang, L. L. – Liua, J. – Wang, J. P.: Development of an indirect competitive immunoassay for determination of L-hydroxyproline in milk. Food and Agricultural Immunology, *25*, 2014, pp. 243–255. DOI: 10.1080/09540105.2013.768965.
5. Conventz, A. – Musiol, A. – Brodowsky, C. – Müller-Lux, A. – Dewes, P. – Kraus, T. – Schettgen, T.: Simultaneous determination of 3-nitrotyrosine- tyrosine- hydroxyproline and proline in exhaled breath condensate by hydrophilic interaction liquid chromatography/electrospray ionization tandem mass spectrometry. Journal of Chromatography B, 860, 2007, pp. 78–85. DOI: 10.1016/j.jchromb.2007.10.031.
6. Kataoka, H. – Nabeshima, N. – Nagao, K. – Makita, M.: Selective and sensitive determination of urinary total proline and hydroxyproline by gas chromatography with flame photometric detection. Clinica Chimica Acta, *214*, 1993, pp. 13–20. DOI: 10.1016/0009-8981(93)90298-I.
7. Clément, A. – Dorais, M. – Vernon, M.: Nondestructive measurement of fresh tomato lycopene content and other physicochemical characteristics using visible-NIR spectroscopy. Journal of Agricultural and Food Chemistry, *56,* 2008, pp. 9813–9818. DOI: 10.1021/jf801299r.
8. Nordon, A. – Littlejohn, D. – Dann, A. S. – Jeffkins, P. A. – Richardson, M. D. – Stimpson, S. L.: In situ monitoring of the seed stage of a fermentation process using non-invasive NIR spectrometry. Analyst, *133*, 2008, pp. 660–666. DOI: 10.1039/b719318a.
9. Pissard, A. – Pierna, J. F. – Baeten, V. – Sinnaeve, G. – Lognay, G. – Mouteau, A. – Dupont, P. – Rondia, A. – Lateur, M.: Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. Journal of the Science of Food and Agriculture, *93*, 2013, pp. 238–244. DOI: 10.1002/jsfa.5779.
10. Travers, S. – Bertelsena, M. G. – Kucheryavskiy, S. V.: Predicting apple (cv. Elshof) postharvest dry matter and soluble solids content with near infrared spectroscopy. Journal of the Science of Food and Agriculture, *94*, 2014, pp. 955–962. DOI: 10.1002/jsfa.6343.
11. Meiyan, Y. – Jing, L. – Shaoping, N. – Jielun, H. – Qiang, Y. – Mingyong, X. – Hua, X. – Zeyuan, D. – Weiwan, Z.: Rapid determination of docosahexaenoic acid in powdered oil by near-infrared spectroscopy. Food Science and Technology International, *16*, 2010, pp. 187–93. DOI: 10.1177/1082013209353379.
12. Cevoli, C. – Gori, A. – Nocetti, M. – Cuibus, L. – Caboni, M. F. – Fabbri, A.: FT-NIR and FT-MIR spectroscopy to discriminate competitors, non compliance and compliance grated Parmigiano Reggiano cheese. Food Research International, *52*, 2013, pp. 214–220. DOI: 10.1016/j.foodres.2013.03.016.
13. Graham, S. F. – Haughey, S. A. – Ervin, R. M. – Cancouët, E. – Bell, S. – Elliott, C. T., The application of near-infrared (NIR) and Raman spectroscopy to detect adulteration of oil used in animal feed production. Food Chemistry, *132*, 2012, pp. 1614–1619. DOI: 10.1016/j.foodchem.2011.11.136.
14. Blanco, M. – Coello, J. – Iturriaga, H. – Maspoch, S. – Pezuela, C.: Near-infrared spectroscopy in the pharmaceutical industry. Critical review. Analyst, *123*, 1998, pp. 135–150. DOI: 10.1039/a802531b.
15. Balabin, R. M. – Safieva, R. Z.: Capabilities of near infrared spectroscopy for the determination of petroleum macromolecule content in aromatic solutions. Journal of Near Infrared Spectroscopy, *15*, 2007, pp. 343–349. DOI: 10.1255/jnirs.749.
16. Felizardo, P. – Baptista, P. – Menezes, J. C. – Correia, M. J. N.: Multivariate near infrared spectroscopy models for predicting methanol and water content in biodiesel. Analytica Chimica Acta, *595*, 2007, pp. 107–113. DOI: 10.1016/j.aca.2007.02.050.
17. Capuano, E. – Boerrigter-Eenling, R. – Koot, A. – Ruth, S. M.: Targeted and untargeted

detection of skim milk powder adulteration by near-infrared spectroscopy. Food Analytical Methods, *8*, 2015, pp. 2125–2134. DOI: 10.1007/s12161-015-0100-3.
18. Santos, P. M. – Pereira-Filho, E. R. – Rodriguez-Saona, L. E.: Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. Food Chemistry, *138*, 2013, pp. 19–24. DOI: 10.1016/j.foodchem.2012.10.024.
19. Mauer, L. J, – Chernyshova, A. A. – Hiatt, A. – Deering, A. – Davis, R.: Melamine detection in infant formula powder using near- and mid-infrared spectroscopy. Journal of Agricultural and Food Chemistry, *57*, 2009, pp. 3974–3980. DOI: 10.1021/jf9024163.
20. Borin, A. – Ferrao, M. F. – Mello, C. – Maretto, D. A. – Poppi, R. J.: Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. Analytica Chimica Acta, *579*, 2006, pp. 25–32. DOI: 10.1016/j.aca.2006.07.008.
21. Balabin, R. M. – Smirnov, S. V.: Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: A quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. Talanta, *85*, 2011, pp. 562–568. DOI: 10.1016/j.talanta.2011.04.026.
22. Vapnik, V. N.: Statistical learning theory. New York : John Wiley and Sons, 1998. ISBN: 978-0471030034.
23. Suykens, J. A. K. – Van Gestel, T. – De Brabanter, J. – De Moor, B. – Vandewalle J.: Least-squares support vector machines. Singapore : World Scientific Publishing, 2002. ISBN: 978-981-238-151-4.
24. Li, J. Z. – Liu, H. X. – Yao, X. J. – Liu, M. C. – Hu, Z. D. – Fan, B. T.: Structure-activity relationship study of oxindole-based inhibitors of cyclin-dependent kinases based on least-squares support vector machines. Analytica Chimica Acta, *581*, 2007, pp. 333–342. DOI: 10.1016/j.aca.2006.08.031.
25. Liu, F. – He, Y.: Classification of brands of instant noodles using Vis/NIR spectroscopy and chemometrics. Food Research International**,** *41*, 2008, pp. 562–567. DOI: 10.1016/j.foodres.2008.03.011.
26. Chauchard, F. – Cogdill, R. – Roussel, S. – Roger, J. M. – Bellon-Maurel, V.: Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. Chemometrics and Intelligent Laboratory Systems, *71*, 2004, pp. 141–150. DOI: 10.1016/j.chemolab.2004.01.003.
27. Liu, F.– He, Y. – Wang, L.: Comparison of calibrations for the determination of soluble solids content and PH of rice vinegars using visble and short-wave near infrared spectroscopy. Analytica Chimica Acta, *610*, 2008, pp. 196–204. DOI: 10.1016/j.aca.2008.01.039.
28. Wu, D. – He, Y. – Feng, S. – Sun, D.-W.: Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM. Journal of Food Engineering, 84, 2007, pp. 124–131. DOI: 10.1016/j.jfoodeng.2007.04.031.
29. Mora, C. R. – Schimleck, L. R.: Kernel regression methods for the prediction of wood properties of Pinus taeda using near infrared spectroscopy. Wood Science and Technology, *44*, 2010, pp. 561–578. DOI: 10.1007/s00226-009-0299-5.
30. Li, S. – Shi, Z. – Chen, S. – Ji, W. – Zhou, L – Yu, W. – Webster, R.: In situ measurements of organic carbon in soil profiles using vis-NIR spectroscopy on the Qinghai-Tibet plateau. Environmental Science and Technology, *49*, 2015, pp. 4980–4987. DOI: 10.1021/es504272x.

Received 29 December 2016; 1st revised 13 June 2017; 2nd revised 12 July 2017; accepted 31 July 2017; published online 16 September 2017.