PEER-REVIEWED RESEARCH
Spectroscopy papers have undergone a double-blind peer review process and are available on our open-access website.

Identification of Different Dairy Products Using Raman Spectroscopy Combined with Fused Lasso Distributionally Robust Logistic Regression

Xiang Xu, Wentao Xiao, Yiyun Cao, and Zhengyong Zhang

Raman spectroscopy has been more widely used recently in the quality detection of dairy products. Because Raman spectroscopy can conduct rapid analyses of small sample sizes at high dimensions, its use in the dairy industry is becoming a hot topic for researchers. To improve the robustness and accuracy of logistic regression identification method, a new Raman spectroscopy identification method was proposed that combines a distributionally robust optimization technique and fused lasso technique with logistic regression. Then, Raman spectroscopy was used to analyze two types of dairy products that were collected for anti-jamming identification testing to verify the effectiveness of the new method. The experimental results show that the proposed method is more robust and has a higher recognition accuracy than the traditional logistic regression.

As one of the important food sources for humans, dairy products are regularly consumed globally. Recently, the dairy industry has seen an increase in the sale of wantonly counterfeit dairy products driven by economic interests, which has seriously endangered the life and health of consumers. As a result, it has become imperative to efficiently identify and detect adulterated substances in dairy products, such as milk powder, to ensure the safety of dairy products safety, which is of vital significance to China's current national conditions. Among the various detection technologies, Raman spectroscopic imaging technology is one of the few technologies that can simultaneously make it possible to rapidly conduct micro-region analysis and large-area scanning, and it has a great application potential in the analysis of adulterated substances of milk powder with its unique advantages of high resolution and high throughput. The Raman spectral analysis technique is based on the Raman scattering effect discovered by Indian scientist C. V. Raman. It analyzes the scattering spectrum different from the incident light frequency to obtain information on molecular vibration and rotation, and it is applied to the study of molecular structure.

Figure 1 shows the Raman spectroscopy of two different substances. We can see that the differences in the Raman spectroscopy of F and I are mainly reflected in the stokes peaks, and the algorithm proposed in this paper is mainly based on the extraction of peak segment characteristics as a means of identifying different material components.

The Raman spectrum has many raw data features, including a small sample size and several distinct peaks, but also a large amount of noise interference. Therefore, it is particularly important to pretreat the spectrum, remove noise interference, and preserve stokes peaks. When analyzing and applying Raman spectroscopy, because of the influence of noise, the small sample size and sequence correlation of features, the existing methods face many challenges, such as noise interference, low efficiency, and difficulty in identification. To solve these problems, a fused lasso logistic regression model based on the distributionally robust optimization, which uses the robust optimization idea to weaken the influence of noise, is proposed. The fused lasso method extracts the key feature terms and ensures the sequencing between the feature to highlight the spectral peaks. Meanwhile, we designed an algorithm for the model, and the iterative process of the algorithm has a simple closed form, which is convenient to be ap-

plied in practice.

Dairy products have long faced challenges, such as adulteration and counterfeiting, drug residues, contaminants, and excessive use of food additives and illegal additives. Compared to the high cost and long cycle of laboratory testing methods, Raman spectroscopy is non-destructive, efficient, and pollution-free, which is why it has been used in several food analysis applications. For example, Raman spectroscopy was used to perform real-time detection of the quality of different types of cooking oil under frying conditions, and it had also been used to classify and certify 70 servings of Spanish peppers (1,2). Ríos-Reina and others distinguished Spanish Protected Designation of Origin wine vinegar for categories with Raman spectroscopy (3). Raman spectroscopy was also used to classify and identify Spanish gasoline (4), and the exhaust soot of diesel engines and gasoline engines under different laser powers (5). Five fat samples using Raman spectroscopy were identified by using the kernel principal components analysis nearest neighbor model (6).

Because the signal is subject to experimental testing conditions and baseline interference, there are interference factors, such as noise, baseline drift, overlapping peaks, cosmic ray energy, and fluorescence that impact the analysis of Raman spectroscopy data. To resolve these problems, researchers have made many attempts to improve Raman measurements in recent years. One study made use of the correlation coefficients of wavelets of different levels to identify noise and complete the denoising of Raman spectra (7). Angeyo and Gari conducted correlation analysis between the spectra of seaweed samples and traditional inductively coupled plasma (ICP) spectra (1). In their research, they used the fully cross-verified partial least squares (PLS) regression method and the nonlinear iterative PLS algorithm. Principal component analysis (PCA) combined with linear discriminant analysis (LDA) was used for multivariate statistical analysis to distinguish the Raman spectra of different blood groups (8). However, most of Raman spectroscopy applications in dairy products are still used for qualitative detection. There are still some interference factors in the actual quantitative analysis process, such as poor reproducibility and signal susceptibility to experimental conditions. Infrared (IR) spectroscopy was combined with stoichiometry to detect the quality parameters in milk powder (9). Fourier transform Raman spectroscopy was explored as a fast and reliable screening method for assessing milk powder quality and identifying doped whey (10,11). Based on Swiss cheese maturation process, the PLS regression and artificial neural networks were utilized to model the relationship between spectral profiles and hardness values (12).

With the development of machine learning technology, a quantitative analysis method was established for polycyclic aromatic hydrocarbon (PAH) surface-enhanced Raman spectroscopy (SERS) with PCA dimensionality reduction and the support vector machine (SVM) algorithm (13). It was proposed that an adaptive genetic algorithm for point-by-point selection mixed terahertz absorption spectral wavelengths, dynamically adjusting the crossover and mutation probability (14). The extremely

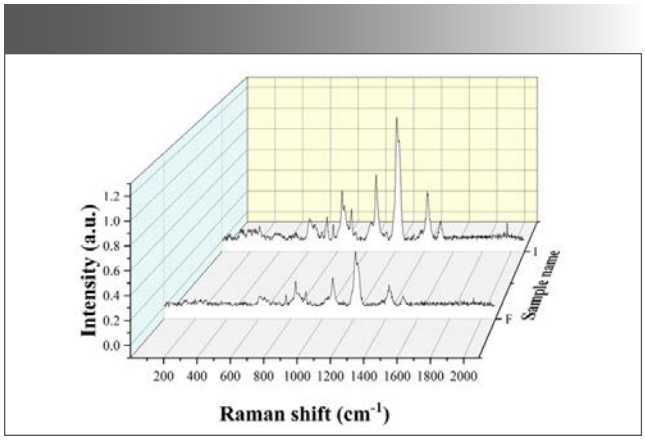

FIGURE 1: Raman spectra of two substances.

randomized trees model was used to accurately match the entire spectral range to their respective minerals (15). A dynamic spectrum matching method was developed, which was based on convolutional conjoined neural networks (16). Sha and others studied the uniformity of rice flours of four different particle sizes using relative standard deviation analysis of Raman spectra and hierarchical cluster analysis (17). A robust correction model was designed for blood glucose spectral monitoring using a support vector machine algorithm (18). Most studies have shown that Raman spectroscopy has broad application for rapid and non-destructive quality testing of dairy products. However, the advanced analysis methods of Raman spectroscopy in dairy products analysis applications still need to be explored.

The contributions of this paper are as follows: first, a new Raman spectroscopy recognition model is proposed by using the fused lasso term, distributionally robust optimization and logistic regression. Second, we transform the distributionally robust optimization model that cannot be solved directly into a tractable form through optimization theory. And third, we develop an effective algorithm for solving the transformed model. Its outer loop is the cutting plane method and its inner loop is the Variant Auxiliary Problem Principle (VAPP) method.

Spectrum Recognition Method Based on FLDRO

In standard statistical learning, it is assumed that all samples $(x,y) \in \Xi = \mathbb{X} \times \{0,1\}$ follow the distribution $\mathbb{P}$ independently, where $x \in \mathbb{X} \subseteq \mathbb{R}^{1\times m}$ represents feature vector, and $y \in \{0,1\}$ is the classification label. If the distribution $\mathbb{P}$ is known, the undetermined parameter $b \in \mathbb{R}^m$ of the logistic regression model can be obtained by solving the following random optimization problem:

$$\inf_b \left\{ \mathbb{E}^{\mathbb{P}}[l_b(x,y)] = \int_{\mathbb{X}\times\{0,1\}} l_b(x,y)\mathbb{P}\big(d(x,y)\big) \right\}, \quad [1]$$

where the loss function $l_b(x,y) = -yxb + \log(1 + e^{xb})$.

However, distribution $\mathbb{P}$ cannot be directly observed by training samples in practical application. Therefore, the distribution $\mathbb{P}$ is uncertain, and we can solve equation [1] by using the distributionally robust optimization (DRO) theory to solve the following problem:

$$\inf_{b} \sup_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}^{\mathbb{Q}}[l_b(x,y)], \quad [2]$$

where $\mathcal{P}$ is a fuzzy set of distributions supported by training samples and contains the unknown distribution $\mathbb{P}$. In probability space, we use the Wasserstein metric to characterize a spherical fuzzy set $\mathcal{P}$. And then, we need to introduce the following definition of *Wasserstein distance*: Let $M(\Xi^2)$ denote the set of probability distributions on $\Xi\times\Xi$.

The Wasserstein distance between two distributions $\mathbb{P}$ and $\mathbb{Q}$ supported on Ξ is defined as:

$$W(\mathbb{Q},\mathbb{P}) = \inf_{\Pi\in M(\Xi^2)} \left\{\int_{\Xi^2} d(\xi,\xi')\Pi(d\xi,d\xi') : \int_{\Xi}\Pi(d\xi,\xi') = \mathbb{Q}(\xi'), \int_{\Xi}\Pi(\xi,d\xi') = \mathbb{P}(\xi)\right\}$$

where $\xi = (x,y)$, and $d(\xi,\xi')$ is a metric on Ξ.

In this paper, we used Wasserstein balls $\mathbb{B}_{\delta}(\mathbb{P}) = \{\mathbb{Q}: W(\mathbb{Q},\mathbb{P}) \le \delta\}$ with radius δ and center $\mathbb{P}$ as fuzzy sets. Given the training samples $\{(\hat{x}_i,\hat{y}_i)\}_{i=1}^n$, we can use the empirical distribution $\hat{\mathbb{P}}_n = \frac{1}{n}$ to approximate the unknown distribution $\mathbb{P}$. Therefore, equation [2] in this study can be written as the following form.

$$\text{(DRO)} \qquad \inf_{b} \sup_{\mathbb{Q}\in\mathbb{B}_{\delta}(\hat{\mathbb{P}}_n)} \mathbb{E}^{\mathbb{Q}}[l_b(x,y)]. \quad [3]$$

Problem (DRO) cannot be solved directly, and Theorem 1 shows that it is equivalent to a tractable semi-infinite programming problem.

Theorem 1 (Tractable Reformulation): *The problem (DRO) is equivalent to the following semi-infinite programing problem.*

$$\begin{aligned} &\min_{b,u,v} \frac{1}{n}\sum_{i=1}^{n} v_i + \delta u \\ &s.t. \begin{cases} l_b(\xi) - v_i - u d(\xi,\xi_i) \le 0, \forall \xi \in \Xi, \forall i \in [n], \\ u \ge 0. \end{cases} \end{aligned} \quad [4]$$

where $\xi = (x,y)$, and $d(\xi,\xi_i)$ is a metric on $\Xi = \mathbb{X}\times\{0,1\}$.

Proof: By definition of the Wasserstein ball and the empirical distribution, we have:

$$\begin{aligned} \mathbb{B}_{\delta}(\hat{\mathbb{P}}_n) = \Bigg\{ &\mathbb{Q}: \exists \Pi \; s.t. \sum_{i=1}^{n} \Pi(\xi,\xi_i) = \mathbb{Q}(\xi), \forall \xi \in \Xi, \int_{\Xi} \Pi(d\xi,\xi_i) \\ &= \frac{1}{n}, \forall i, \int_{\Xi} \sum_{i=1}^{n} d(\xi,\xi_i)\Pi(d\xi,\xi_i) \le \delta \Bigg\}. \end{aligned} \quad [5]$$

Then, the inner problem (DRO) is formulated as the following conic linear program:

$$\sup_{\mathbb{Q}\in\mathbb{B}_{\delta}(\hat{\mathbb{P}}_n)} \mathbb{E}^{\mathbb{Q}}[l_b(x,y)] = \sup_{\mathbb{Q}\in\mathbb{B}_{\delta}(\hat{\mathbb{P}}_n)} \int_{\Xi} l_b(\xi)\mathbb{Q}(d\xi) \quad [6]$$

$$= \begin{cases} \sup\limits_{\Pi\in M(\Xi^2)} \int_{\Xi} \sum\limits_{i=1}^{n} l_b(\xi)\Pi(d\xi,\xi_i) \\ s.t. \int_{\Xi} \Pi(d\xi,\xi_i) = \frac{1}{n}, \forall i \in [n], \\ \quad \int_{\Xi} \sum\limits_{i=1}^{n} d(\xi,\xi_i)\,\Pi(d\xi,\xi_i) \le \delta, \\ \quad \Pi \ge 0. \end{cases} \quad [7]$$

Thus, we obtain the following sample-based dual problem of [7].

$$\min_{u,v} \frac{1}{n}\sum_{i=1}^{n} v_i + \delta u \quad [8a]$$

$$s.t. \begin{cases} l_b(\xi) - v_i - u d(\xi,\xi_i) \le 0, \forall \xi \in \Xi, \forall i \in [n], & [8b] \\ u \ge 0. & [8c] \end{cases}$$

Where $\boldsymbol{\xi} = (x,y)$ and $d(\boldsymbol{\xi},\boldsymbol{\xi}_i)$ is a metric on $\Xi = \mathbb{X}\times\{0,1\}$. Therefore, the result of Theorem 1 follows by combing equation [8] with the outer minimization problem (DRO).

The Raman spectroscopy data is not only high in dimension, but also has a close correlation between adjacent features, which is known as Raman spectral peaks. Therefore, the fused lasso term $R(b)$ is added to the model to preserve the sparsity of the difference between neighboring features to highlight the importance of spectral peaks. The fused lasso model can automatically and continuously select features while performing parameter estimation, which makes model simplification explanatory:

$$R(b) = \lambda_1\|b\|_1 + \lambda_2 \sum_{i=2}^{m} \|b_i - b_{(i-1)}\|_1, \lambda_1 \ge 0, \lambda_2 \ge 0. \quad [9]$$

Thus, we propose the following fused lasso distributionally robust optimization (FLDRO) model:

$$\text{(FLDRO)} \qquad \inf_{b} \sup_{\mathbb{Q}\in\mathbb{B}_{\delta}(\hat{\mathbb{P}}_n)} \mathbb{E}^{\mathbb{Q}}[l_b(x,y) + R(b)]. \quad [10]$$

The loss function $l_b(x,y) = -yxb + \log(1 + e^{xb})$ of the logistic regression model is known, and the distance function is $d\big((x,y),(\hat{x}_i,\hat{y}_i)\big) = \|x-\hat{x}_i\|^2 + \|y-\hat{y}_i\|^2$ According to Theorem 1, we can convert the problem (FLDRO) into the following tractable reformulation form:

$$\min_{b,u,v} \frac{1}{n}\sum_{i=1}^{n} v_i + \delta u + \lambda_1\|b\|_1 + \lambda_2 \sum_{i=2}^{m} \|b_i - b_{(i-1)}\|_1 \quad [11a]$$

$$s.t. \begin{cases} -xb + log\big(1 + e^{xb}\big) - v_i - u(\|x-\hat{x}_i\|^2 + \|1-\hat{y}_i\|^2) \le 0, \forall x \in \mathbb{X}, \forall i \in [n], & [11b] \\ log\big(1 + e^{xb}\big) - v_i - u(\|x-\hat{x}_i\|^2 + \|\hat{y}_i\|^2) \le 0, \forall x \in \mathbb{X}, \forall i \in [n], & [11c] \\ u \ge 0. & [11d] \end{cases}$$

In the next section, we will introduce an algorithm to solve equation [11].

Algorithm for the Tractable Reformulation FLDRO

In the previous section, we give the tractable form of equation [11], which is a semi-infinite programming problem that can be solved by using the cutting-surface algorithm. The idea of the cutting-surface method is to solve a relaxation problem of the semi-infinite program at each iteration, where the relaxation problem has a finite number of constraints. Constraints are added to the current set of constraints at the solution of the relaxation problem.

Let $\chi = [b,u,v]$ be the decision variables of equation [11], and define the following functions.

$$f(\chi) = \frac{1}{n}\sum_{i=1}^{n} v_i + \delta u + \lambda_1\|b\|_1 + \lambda_2 \sum_{i=2}^{m} \|b_i - b_{(i-1)}\|_1, \quad [12]$$

$$g_i(\chi,\xi) = l_b(\xi) - v_i - u d(\xi,\xi_i), \forall \xi \in \Xi \forall i \in [n]. \quad [13]$$

Thus, the problem [11] can be rewritten as

$$\min_{\chi} f(\chi) \quad [14a]$$

$$\text{s.t.} \begin{cases} g_i(\chi,\xi) \le 0, \forall \xi \in \Xi, \forall i \in [n], & [14b] \\ \chi \in \mathrm{X}. & [14c] \end{cases}$$

Equation [14] is a semi-infinite program. We give the following cutting-surface algorithm to solve this problem:

Algorithm 1 A cutting-surface algorithm to solve [14].

Input: training data $\{(\hat{x}_i, \hat{y}_i)\}$,parameters δ,λ_1,λ_2.

Output: χ^*

1: (Initialization) Set $\Xi^0 \leftarrow \emptyset$, $k \leftarrow 0$.
2: Determine an optimal solution χ^k of the problem
3: $\min_{\chi \in \mathrm{X}} f(\chi)\, \mathrm{s.t.}\, g_i(\chi,\xi) \le 0, \forall \xi \in \Xi^k, \forall i \in [n].$
4: **for** i=1,2,⋯,n **do**
5: Determine an optimal solution ξ^{k+1} of i-th separation problem
6: $\max_{\xi \in \Xi} g_i(\chi^k, \xi)$
7: **if** $g_i(\chi^k, \xi^{k+1}) > 0$
8: $\Xi^k \leftarrow \Xi^k \cup \{\xi^{k+1}\}$
9: **end if**
10: **end for**
11: **if** no cut is added in the current iteration, **then**
12: STOP.
13: **end if**
14: Update $\Xi^{k+1} \leftarrow \Xi^k, k \leftarrow k+1$ and go to Step 2.

The key to Algorithm 1 is to solve the following two optimization problems.

$$\min_{\chi \in \mathrm{X}} f(\chi)\, \mathrm{s.t.}\, g_i(\chi,\xi) \le 0, \forall \xi \in \Xi^k, \forall i \in [n]. \quad [15]$$

$$\max_{\xi \in \Xi} g_i(\chi^k, \xi), \forall i \in [n]. \quad [16]$$

Because equation [16] has the same structure for all $i \in [n]$, according to equation [11], we drop the index i in $g_i(\chi,\xi)$ and use the following notations to simplify our presentation:

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_{2n} \end{pmatrix}, \quad [17]$$

$$g(\chi, X) = \begin{pmatrix} -x_1 b + \log\left(1 + e^{x_1 b}\right) - v_1 - u(\|x_1 - \hat{x}_1\|^2 + \|1 - \hat{y}_1\|^2) \\ \vdots \\ -x_n b + \log\left(1 + e^{x_n b}\right) - v_n - u(\|x_n - \hat{x}_n\|^2 + \|1 - \hat{y}_n\|^2) \\ \log\left(1 + e^{x_{n+1} b}\right) - v_1 - u(\|x_{n+1} - \hat{x}_1\|^2 + \|\hat{y}_1\|^2) \\ \vdots \\ \log\left(1 + e^{x_{2n} b}\right) - v_n - u(\|x_{2n} - \hat{x}_n\|^2 + \|\hat{y}_n\|^2) \end{pmatrix}, \quad [18]$$

then its gradient about X is:

$$\nabla_X g(\chi, X) = \begin{pmatrix} -b^T + \frac{e^{x_1 b}}{1 + e^{x_1 b}} b^T - 2u(x_1 - \hat{x}_1) \\ \vdots \\ -b^T + \frac{e^{x_n b}}{1 + e^{x_n b}} b^T - 2u(x_n - \hat{x}_n) \\ \frac{e^{x_{n+1} b}}{1 + e^{x_{n+1} b}} b^T - 2u(x_{n+1} - \hat{x}_1) \\ \vdots \\ \frac{e^{x_{2n} b}}{1 + e^{x_{2n} b}} b^T - 2u(x_{2n} - \hat{x}_n) \end{pmatrix}. \quad [19]$$

Thus, equation [16] can be solved by the projected gradient descent algorithm, and its iterative process is as follows:

$$X^{k+1} = \prod_{\mathbb{X}} \left(X^k + \varepsilon \nabla_X g(\chi^k, X^k) \right), \quad [20]$$

where $\prod_{\mathbb{X}}(\cdot)$ is the projection function onto $\mathbb{X}$, $\varepsilon > 0$ is the learning rate.

Without loss of generality, given a finite $\mathbb{X}^k$ set, the number of its elements is $|\mathbb{X}^k| = l$. We write equation [15] as follows:

$$\min_{b,u,v,\eta} \frac{1}{n} \sum_{i=1}^{n} v_i + \delta u + \lambda_1 \|b\|_1 + \lambda_2 \|\eta\|_1 \quad [21a]$$

$$\text{s.t.} \begin{cases} -xb + \log\left(1 + e^{xb}\right) - v_i - u(\|x - \hat{x}_i\|^2 + \|1 - \hat{y}_i\|^2) \le 0, \forall x \in \mathbb{X}^k, \forall i \in [n], & [21b] \\ \log\left(1 + e^{xb}\right) - v_i - u(\|x - \hat{x}_i\|^2 + \|\hat{y}_i\|^2) \le 0, \forall x \in \mathbb{X}^k, \forall i \in [n], & [21c] \\ Ab = \eta, & [21d] \\ u \ge 0. & [21e] \end{cases}$$

$$\text{Where } A = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(m-1)\times m}$$

$b \in \mathrm{R}^m$ is the regression coefficient, $\lambda_1 \ge 0$ and $\lambda_2 \ge 0$ are the penalty parameters, $\delta \ge 0$ is Wasserstein balls radius, $\hat{x}_i \in \mathbb{R}^{1\times m}$ is the feature of i-th training sample, and $\hat{y}_i \in \{0,1\}$ is the corresponding label. Through analysis, the objective function of equation [21] is a non-smooth, the constraints [21b] and [21c] are nonlinear, we can use the Variant Auxiliary Problem Principle (VAPP) method proposed by Zhao and Zhu (19) to solve it. For the sake of brevity, we write equation [21] as follows:

$$\min_{\substack{b \in \mathbb{R}^m, u \ge 0 \\ v \in \mathbb{R}^n, \eta \in \mathbb{R}^{m-1}}} f(b,u,v,\eta) \quad [22a]$$

$$\text{s.t.}\, \theta(b,u,v,\eta) \in \{x \in \mathbb{R}^{2nl} | x \le 0\} \times \{0\}^m. \quad [22b]$$

where

$$f(b,u,v,\eta) = \frac{1}{n} \sum_{i=1}^{n} v_i + \delta u + \lambda_1 \|b\|_1 + \lambda_2 \|\eta\|_1, \quad [23]$$

$$\theta(b,u,v,\eta) = \Omega(b) - \begin{pmatrix} \mathrm{vec}(\mathbb{V}) \\ \mathrm{vec}(\mathbb{V}) \\ \eta - Ab \end{pmatrix} + \Phi(u), \quad [24a]$$

$$\Omega(b) = \begin{pmatrix} \Omega_1(b) \\ \Omega_2(b) \\ \mathbb{0}_{m-1} \end{pmatrix} \in \mathbb{R}^{2nl+m-1}, \quad [24b]$$

$$\mathbb{V} = \begin{pmatrix} v_1 & \cdots & v_1 \\ \vdots & \ddots & \vdots \\ v_n & \cdots & v_n \end{pmatrix} \in \mathbb{R}^{n\times l}, \quad [24c]$$

$$\Phi(u) = \begin{pmatrix} \Phi_1(u) \\ \Phi_2(u) \\ \mathbb{0}_{m-1} \end{pmatrix} \in \mathbb{R}^{2nl+m-1}, \quad [24d]$$

$$\mathrm{mat}(\Omega_1(b))_{ij} = -x_j b + \log\left(1 + e^{x_j b}\right), \forall i \in [n], \forall j \in [l], \quad [24e]$$

$$\mathrm{mat}(\Omega_2(b))_{ij} = \log\left(1 + e^{x_j b}\right), \forall i \in [n], \forall j \in [l], \quad [24f]$$

$$\mathrm{mat}(\Phi_1(u))_{ij} = -u\left(\|x_j - \hat{x}_i\|^2 + \|1 - \hat{y}_i\|^2\right), \forall i \in [n], \forall j \in [l], \quad [24g]$$

$$\mathrm{mat}(\Phi_2(u))_{ij} = -u\left(\|x_j - \hat{x}_i\|^2 + \|\hat{y}_i\|^2\right), \forall i \in [n], \forall j \in [l], \quad [24h]$$

The main idea of VAPP is to linearize the non-separable term $\Omega(b)$ to achieve the purpose of parallel computing. Then the iterative process of solving equation [22] by VAPP is obtained.

$$\begin{aligned}(b^{k+1}, u^{k+1}, v^{k+1}, \eta^{k+1}) &= \underset{\substack{b\in\mathbb{R}^m, u\in\mathbb{R}_+^n \\ v\in\mathbb{R}_+^n, \eta\in\mathbb{R}^{m-1}}}{\operatorname{argmin}} f(b,u,v,\eta) \\ &+ \langle \Pi\left(p^k + \gamma\Theta(b^k,u^k,v^k,\eta^k)\right), \nabla\Omega(b^k)b - \begin{pmatrix}\mathrm{vec}(\mathbb{V})\\ \mathrm{vec}(\mathbb{V})\\ \eta - Ab\end{pmatrix} + \Phi(u)\rangle \\ &+ \frac{1}{2\varepsilon}\left(\|b - b^k\|^2 + \|u - u^k\|^2 + \|v - v^k\|^2 + \|\eta - \eta^k\|^2\right), \end{aligned} \quad [25]$$

where $\gamma > 0$, $\varepsilon > 0$.

$$\nabla\Omega_1(b) = \begin{pmatrix} \mathbb{1}_n\left(\frac{e^{x_1 b}}{1+e^{x_1 b}} - 1\right)x_1 \\ \vdots \\ \mathbb{1}_n\left(\frac{e^{x_l b}}{1+e^{x_l b}} - 1\right)x_l \end{pmatrix} \in \mathbb{R}^{nl\times m}, \quad [26a]$$

$$\nabla\Omega_2(b) = \begin{pmatrix} \mathbb{1}_n\left(\frac{e^{x_1 b}}{1+e^{x_1 b}}\right)x_1 \\ \vdots \\ \mathbb{1}_n\left(\frac{e^{x_l b}}{1+e^{x_l b}}\right)x_l \end{pmatrix} \in \mathbb{R}^{nl\times m}. \quad [26b]$$

Since the above optimization problems are separable, we can write it as follows:

$$\begin{aligned} b^{k+1} &= \underset{b\in\mathbb{R}^m}{\operatorname{argmin}} \lambda_1\|b\|_1 + \langle \Pi_1^k, \nabla\Omega_1(b^k)b\rangle + \langle \Pi_2^k, \nabla\Omega_1(b^k)b\rangle + \langle \Pi_3^k, Ab\rangle \\ &+ \frac{1}{2\varepsilon}\|b - b^k\|^2, \end{aligned} \quad [27a]$$

$$u^{k+1} = \underset{u\geq 0}{\operatorname{argmin}} \delta u + \langle \Pi_1^k, \Phi_1(u)\rangle + \langle \Pi_2^k, \Phi_2(u)\rangle + \frac{1}{2\varepsilon}\|u - u^k\|^2, \quad [27b]$$

$$v^{k+1} = \underset{v\geq 0}{\operatorname{argmin}} \frac{1}{n}\langle \mathbb{1}_n, v\rangle - \langle \Pi_1^k, \mathrm{vec}(\mathbb{V})\rangle - \langle \Pi_2^k, \mathrm{vec}(\mathbb{V})\rangle + \frac{1}{2\varepsilon}\|v - v^k\|^2, \quad [27c]$$

$$\eta^{k+1} = \underset{\eta}{\operatorname{argmin}} \lambda_2\|\eta\|_1 - \langle \Pi_3^k, \eta\rangle + \frac{1}{2\varepsilon}\|\eta - \eta^k\|^2, \quad [27d]$$

where

$$\Pi\left(p^k + \gamma\Theta(b^k,u^k,v^k,\eta^k)\right) = \begin{pmatrix}\Pi_1^k \\ \Pi_2^k \\ \Pi_3^k\end{pmatrix}, \quad \Pi_1^k = \left(p_1^k + \gamma\left(\Omega_1(b^k) - \mathrm{vec}(\mathbb{V}^k) + \Phi_1(u^k)\right)\right)_+ \in \mathbb{R}^{nl}, \quad \Pi_2^k = \left(p_2^k + \gamma\left(\Omega_2(b^k) - \mathrm{vec}(\mathbb{V}^k) + \Phi_2(u^k)\right)\right)_+ \in \mathbb{R}^{nl}, \quad \Pi_3^k = p_3^k + \gamma(Ab^{k+1} - \eta^k) \in \mathbb{R}^{m-1}.$$

And the dual variables iteration process is:

$$p^{k+1} = \Pi\left(p^k + \gamma\Theta(b^{k+1},u^{k+1},v^{k+1},\eta^{k+1})\right), \quad [28]$$

where

$$\text{where } p^k = \begin{pmatrix}p_1^k \\ p_2^k \\ p_3^k\end{pmatrix}$$

then

$$p_1^{k+1} = \left(p_1^k + \gamma\left(\Omega_1(b^{k+1}) - \mathrm{vec}(\mathbb{V}^{k+1}) + \Phi_1(u^{k+1})\right)\right)_+ \in \mathbb{R}^{nl}, \quad [29a]$$

$$p_2^{k+1} = \left(p_2^k + \gamma\left(\Omega_2(b^{k+1}) - \mathrm{vec}(\mathbb{V}^{k+1}) + \Phi_2(u^{k+1})\right)\right)_+ \in \mathbb{R}^{nl}, \quad [29b]$$

$$p_3^{k+1} = p_3^k + \gamma\left(Ab^{k+1} - \eta^{k+1}\right) \in \mathbb{R}^{m-1}. \quad [29c]$$

In order to get the closed form for the above iterative process, we need to introduce the soft-thresholding operator:

$$\min_{b\in R^m} \|b\|_1 + \frac{\beta}{2}\|b - r\|^2, \quad [30]$$

where $\beta > 0$, $r \in R^m$. The solution to the problem can be expressed by using a soft threshold operator:

$$b^* = \mathrm{shrink}(r, 1/\beta) = \mathrm{sign}(r)\cdot\max\{0, |r| - 1/\beta\}, \quad [31]$$

where sign(·) is the sign function. And then, we can write the analytical solution of (b,u,v,η) as follows:

$$b^{k+1} = \mathrm{sign}(r_1^k)\cdot\max\{0, |r_1^k| - \lambda_1\varepsilon\}, \quad [32a]$$

$$r_1^k = b^k - \varepsilon\left(\nabla\Omega_1(b^k)^T\Pi_1^k + \nabla\Omega_2(b^k)^T\Pi_2^k + A^T\Pi_3^k\right), \quad [32b]$$

$$u^{k+1} = \left(u^k - \varepsilon\left(\delta - \sum_{i=1}^{n}\sum_{j=1}^{l}\mathrm{mat}(\Pi_1^k)_{ij}\left(\|x_j - \hat{x}_i\|^2 + \|1 - \hat{y}_i\|^2\right) - \sum_{i=1}^{n}\sum_{j=1}^{l}\mathrm{mat}(\Pi_2^k)_{ij}\left(\|x_j - \hat{x}_i\|^2 + \|\hat{y}_i\|^2\right)\right)\right)_+, \quad [32c]$$

$$v^{k+1} = \left(v^k - \varepsilon\left(\frac{1}{n}\mathbb{1}_n - \sum_{j=1}^{l}\mathrm{mat}(\Pi_1^k)_{\cdot j} - \sum_{j=1}^{l}\mathrm{mat}(\Pi_2^k)_{\cdot j}\right)\right)_+, \quad [32d]$$

$$\eta^{k+1} = \mathrm{sign}(r_2^k)\cdot\max\{0, |r_2^k| - \lambda_2\varepsilon\}, \quad [32e]$$

$$r_2^k = \eta^k + \varepsilon\Pi_3^k. \quad [32f]$$

Materials and Methods

In our experiments, we selected two liquid milky yellow cheese products that are indistinguishable to the naked eye, and their Raman spectroscopy is shown in Figure 2 (denoted as MK and SL respectively). The integration time of Raman spectrometer is 70 s, and 25 samples are collected from each dairy product. A total

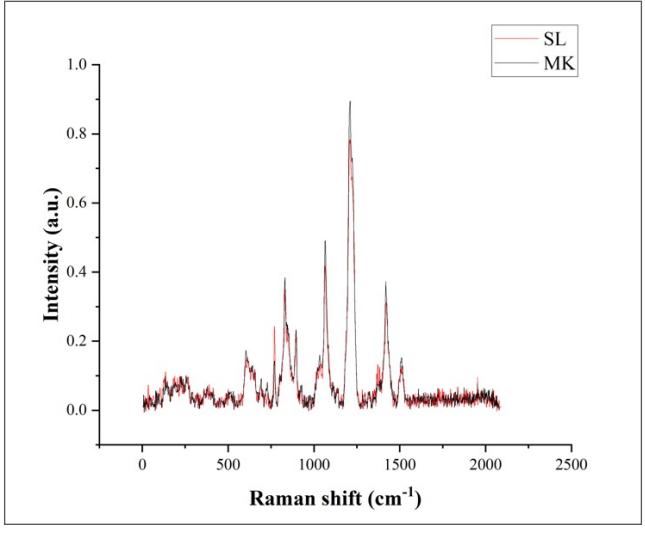

FIGURE 2: Raman spectra of two dairy products.

of n = 50 original Raman spectroscopy samples are collected, and the number of features of every sample is m = 2090. Of the samples, 80 percent were randomly selected as training samples, and the rest are testing samples.

In addition, standard logistic regression (LR), L1 regularized logistic regression (LR-L1) and L2 regularized logistic regression (LR-L2) are taken as the control group. Set FLDRO parameters $\delta = 2, \lambda_1 = 2, \lambda_2 = 2,\ \mathbb{X} = [\hat{x}_{min}, \hat{x}_{max}]$ in equation [20]. The parameter C is equal to 1 in LR-L1 and LR-L2.

$$(\mathrm{LR}) \qquad \min_b l_b(x, y) \qquad [33]$$

$$(\mathrm{LR-L1}) \qquad \min_b \|b\|_1 + C l_b(x, y) \qquad [34]$$

$$(\mathrm{LR-L2}) \qquad \min_b \frac{1}{2}\|b\|^2 + C l_b(x, y) \qquad [35]$$

Figure 3 shows the sparsity of regression coefficient b obtained by the four methods and corresponding testing accuracy. It can be seen from Figure 3 that the recognition accuracy by FLDRO, LR, and LR-L2 is significantly better than that of LR-L1. In terms of sparsity, the FLDRO and LR-L1 are higher than the other two methods. To visually demonstrate the feature selection results by every method without loss of generality, we plot the absolute normalized value of b in one figure with the testing data (as shown in Figure 4). As can be seen from Figure 4, LR and LR-L2 not only take all spectral peaks as key features, but they also include a large number of noise features, which will reduce the generalization ability of the model. Although LR-L1 has no noise feature, only one spectral peak is selected as the key feature, which will lead to underfitting problem. FLDRO takes the spectral peaks marked by the box in the figure as the key feature to distinguish two kinds of dairy products. The experimental results also show that FLDRO has obvious advantages both in regression coefficient sparsity and recognition accuracy.

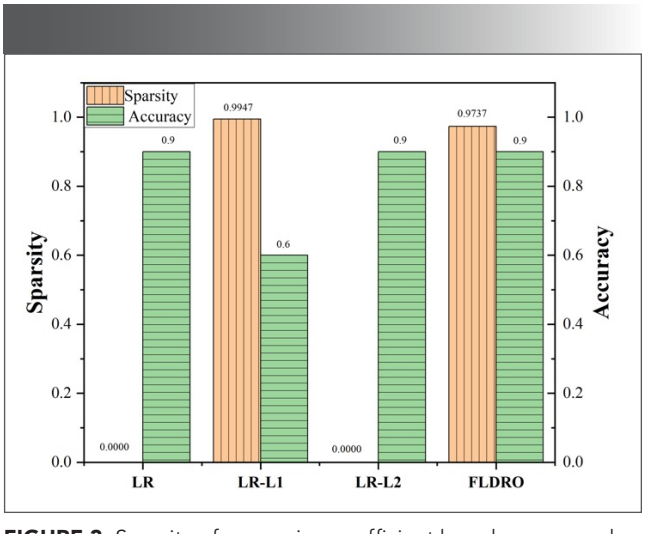

FIGURE 3: Sparsity of regression coefficient b and corresponding testing accuracy.

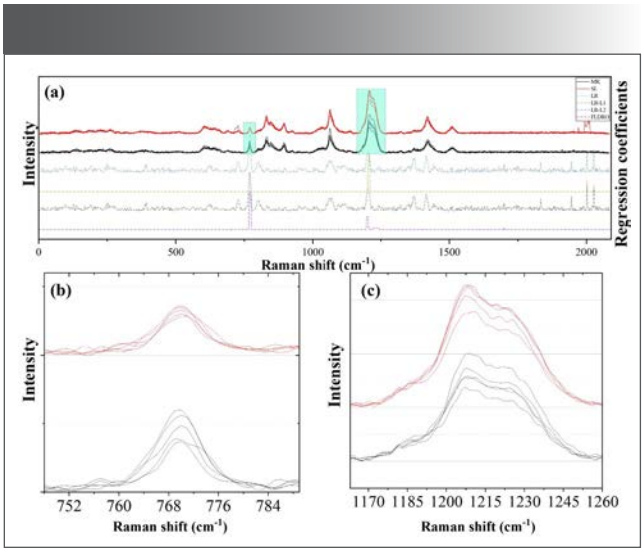

FIGURE 4: (a) Testing samples and regression coefficients obtained by four methods. (b) and (c) Local magnification of selected spectral peaks by FLDRO.

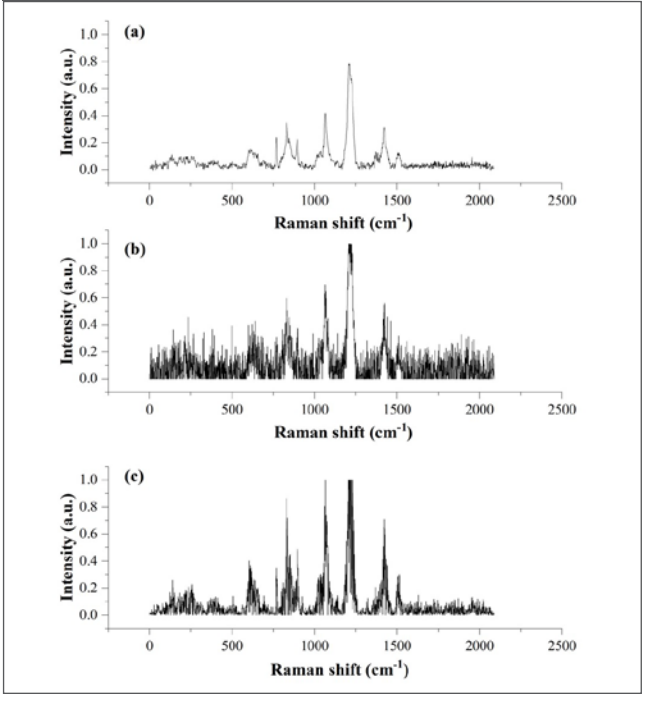

FIGURE 5: Raw data and noisy data. (a) Raw data. (b) Gaussian noise with mean 0.01 and variance 0.01. (c) multiplicative noise with variance 0.04.

Next, we test the robustness of the four methods by adding noise to the original samples. As we can see from Figure 5, although the raw samples retain obvious features after adding noise, a large number of interference factors appear. To test the anti-noise performance of each model, we add Gaussian noise with a mean of 0.01 and a variance of 0.01, and multiplicative noise with a variance of 0.04 to the original test set.

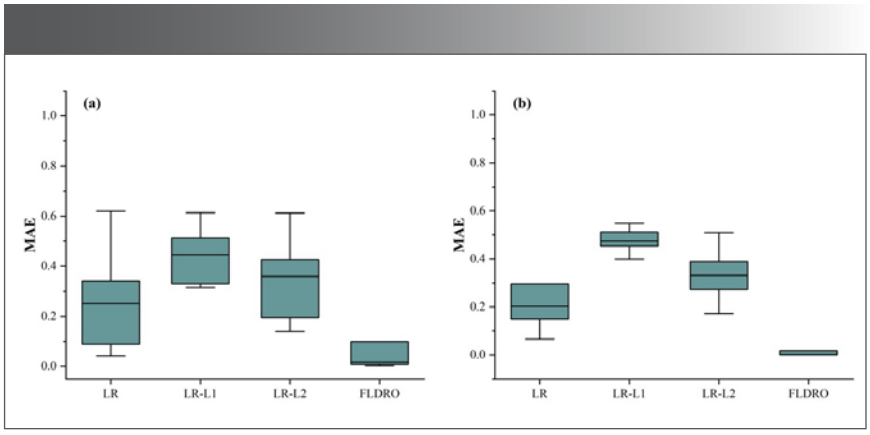

FIGURE 6: Sensitivity for different methods with noise added. (a) MAE boxplot with Gaussian noise setting mean and variance 0.01. (b) MAE boxplot with multiplicative noise setting variance 0.04.

The box plot of mean absolute error (MAE) by the four methods is shown in Figure 6.

$$MAE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left| \log\left(1 + e^{\langle x_{test}^i, b^* \rangle}\right) - y_{test}^i \right| \quad [36]$$

where the x^i_{test} represents the i-th test Raman spectroscopy, y^i_{test} is the corresponding label and n_{test} represents the number of testing samples. The results of the robustness test in Figure 6 show that the MAE box size of FLDRO is the smallest in both noise scenarios, and that the median of the box is also significantly lower than other models, which means that the robustness and accuracy of distributionally robust optimization are significantly better than other models.

Furthermore, we test the recognition accuracy of the four methods under different Gaussian noise intensities. As shown in Figure 7, when the mean value is fixed, and the variance is increasing, the recognition accuracy of all methods is decreasing. However, compared with other methods, FLDRO has the highest accuracy and the smallest influence of variance on accuracy variation. In addition, the accuracy of the three logistic regressions in the figure is close to 50% under the condition of large noise intensity. This is because the sparsity of b by LR and LR-L2 is zero, in other words, all features are selected, which results in serious distortion of its recognition results. Besides, only one spectral peak is selected by LR-L1, which makes it difficult to make a correct identification.

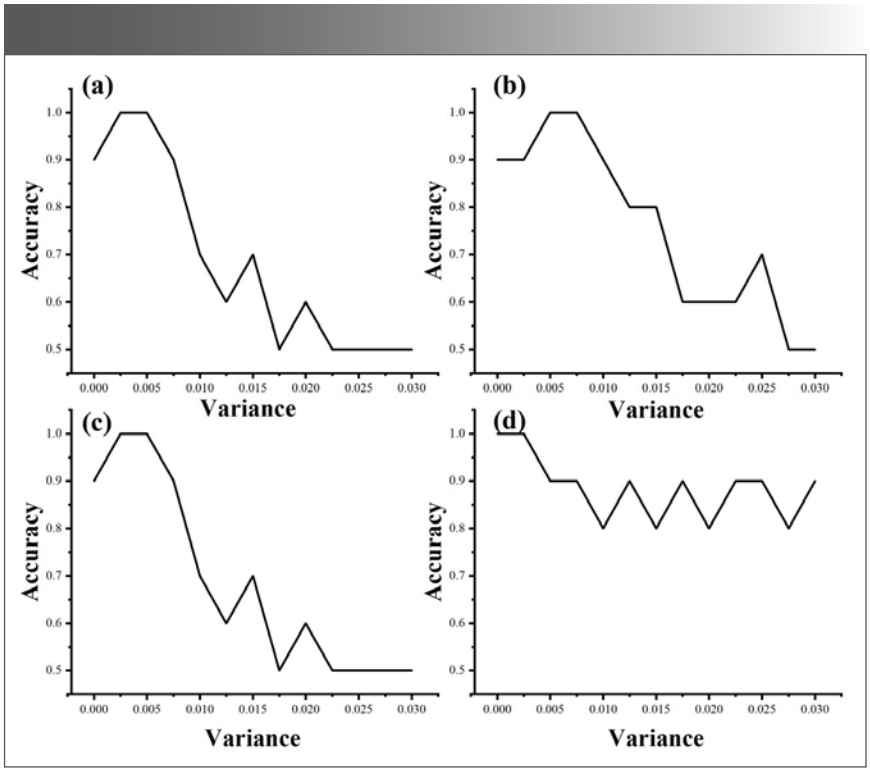

FIGURE 7: Trend of recognition accuracy by different methods with Gaussian noise variance varying and fixed mean value 0.01. (a) LR method. (b) LR-L1 method. (c) LR-L2 method. (d) FLDRO method.

Conclusions

To verify the advantages of FLDRO in the recognition accuracy and robustness of Raman spectral identification, we adopt FLDRO, LR, LR-L1, and LR-L2 methods to perform recognition experiments on spectral data of two kinds of dairy products. First, based on the regression coefficients obtained from training samples by different methods, the recognition accuracy of FLDRO is 90%, and the obtained b sparsity is 0.9737, which is the best among all methods. Second, it can be seen from Figure 4 that FLDRO provides a significant regression coefficient value at the spectral peak of Raman spectroscopy, which indicates that FLDRO has a strong ability to identify spectral peaks. Moreover, Figure 6 shows that MAE boxes size of FLDRO are significantly smaller than those of the other three models under the interference of two different noises, indicating that FLDRO model has a strong anti-interference ability. Finally, in the sensitivity test experiments of Gaussian noise, FLDRO outperforms the other three methods in terms of both recognition ability and robustness.

Xiang Xu, Wentao Xiao, Yiyun Cao, and **Zhengyong Zhang** are with the School of Management Science and Engineering at Nanjing University of Finance and Economics, in Nanjing, China. Direct correspondence to: xuxiang2_2009@foxmail.com •

References

(1) Angeyo, H. K.; Gari, S. Direct rapid quality assurance analysis of complex matrix materials: A chemometrics enabled energy dispersive X-ray fluorescence and scattering spectrometry application. *Appl. Radiat. Isot.* **2022,** 110274. DOI: 10.1016/j.apradiso.2022.110274

(2) Campmajó, G.; Saurina, J.; Núñez, O.; et al. Differential mobility spectrometry coupled to mass spectrometry (DMS–MS) for the classification of Spanish PDO paprika. *Food Chem.* **2022**, *390*:, 133141. DOI: 10.1016/j.foodchem.2022.133141

(3) Ríos-Reina, R.; Elcoroaristizabal, S.; Ocaña-González, J. A.; et al. Characterization and authentication of Spanish PDO wine vinegars using multidimensional fluorescence and chemometrics. *Food Chem.* **2017,** *230*: 108–116. DOI: 10.1016/j.foodchem.2017.02.118

(4) Ardila, J. A.; Soares, F. L. F.; dos Santos Farias, M. A.; et al. Characterization of gasoline by Raman spectroscopy with chemometric analysis. *Anal. Lett.* **2017**, *50* (7), 1126–1138. DOI: 10.1080/00032719.2016.1210616

(5) Ge, H.; Ye, Z.; He, R. Raman spectroscopy of diesel and gasoline engine-out soot using different laser power. *J. Environ. Sci.* **2019**, *79*, 74–80. DOI: 10.1016/j.jes.2018.11.001

(6) Wang, H.; Song, C.; Liu J.; et al. Authenticity identification and adulteration analysis of milk powder based on Raman spectroscopy-pattern recognition method. *Spectrosc. Spectr. Anal.* **2017,** *37* (1), 124–128.

(7) Ehrentreich, F.; Sümmchen, L. Spike removal and denoising of Raman spectra by wavelet transform methods. *Anal. Chem.* **2001**, *73* (17), 4364–4373. DOI: 10.1021/ac0013756

(8) Lin, D.; Zheng, Z.; Wang, Q.; et al. Label-free optical sensor based on red blood cells laser tweezers Raman spectroscopy analysis for ABO blood typing. *Opt. Express* **2016**, *24* (21), 24750–24759. DOI: 10.1364/OE.24.024750

(9) Coitinho, T. B.; Cassoli, L. D.; Cerqueira, P. H. R.; et al. Adulteration identification in raw milk using Fourier transform infrared spectroscopy. *J. Food Sci. Technol.* **2017**, *54* (8), 2394–2402. DOI: 10.1007/s13197-017-2680-y

(10) Almeida, M. R.; Oliveira, K. D. S.; Stephani, R.; et al. Fourier-transform Raman analysis of milk powder: a potential method for rapid quality screening. *J. Raman Spectrosc.* **2011,** *42* (7), 1548–1552. DOI: 10.1002/jrs.2893

(11) Mabood, F.; Jabeen, F.; Hussain, J.; et al. FT-NIRS coupled with chemometric methods as a rapid alternative tool for the detection & quantification of cow milk adulteration in camel milk samples. *Vib. Spectrosc.* **2017,** *92*, 245–250. DOI: 10.1016/j.vibspec.2017.07.004

(12) Vásquez, N.; Magán, C.; Oblitas, J.; et al. Comparison between artificial neural network and partial least squares regression models for hardness modeling during the ripening process of Swiss-type cheese using spectral profiles. *J. Food Eng.* **2018**, *219,* 8–15. DOI: 10.1016/j.jfoodeng.2017.09.008

(13) Chen, Y.; Yan, X.; Zhang, X.; et al. Surface-Enhanced Raman Spectroscopy Quantitative Analysis of Polyey-clic Aromatic Hydrocarbon Based on Support Vector Machine Algorithm. *Chin. J. Lasers* **2019,** *46* (3), 1–8. DOI: 10.3788/CJL201946.0311005

(14) Li, Z.; Guan, A.; Ge, H.; et al. Wavelength selection of amino acid THz absorption spectra for quantitative analysis by a self-adaptive genetic algorithm and comparison with mwPLS. *Microchem J.* **2017,** *132*,185–189. DOI: 10.1016/j.microc.2017.02.002

(15) Sevetlidis, V.; Pavlidis, G. Effective Raman spectra identification with tree-based methods. *J. Cult. Herit.* **2019,** 37, 121–128. DOI: 10.1016/j.culher.2018.10.016

(16) Liu, J.; Gibson, S. J.; Mills, J.; et al. Dynamic spectrum matching with one-shot learning. *Chemometrics Intell. Lab. Syst.* **2019**, *184*, 175–181. DOI: 10.1016/j.chemolab.2018.12.005

(17) Sha, M.; Gui, D.; Zhang, Z.; et al. Evaluation of sample pretreatment method for geographic authentication of rice using Raman spectroscopy. *J. Food Meas. Charact.* **2019,** *13*, 1705–1712. DOI: 10.1007/s11694-019-00087-7

(18) Barman, I.; Kong, C. R.; Dingari, N. C.; et al. Development of robust calibration models using support vector machines for spectroscopic monitoring of blood glucose. *Anal. Chem.* **2010**, *82* (23), 9719–9726. DOI: 10.1021/ac101754n

(19) Zhao, L.; Zhu, D. L. On iteration complexity of a first-order primal-dual method for nonlinear convex cone programming. *J. Oper. Res. Soc. China* **2022**, *10,* (1), 53–87. DOI: 10.1007/s40305-021-00344-x