



Original Research

Visual detection of microbial community during three bacteria mixed fermentation through hyperspectral imaging technology

Yanxiao Li ¹, Xuetao Hu ², Jiyong Shi ^{2,*}, Baijing Qiu ¹, Jianbo Xiao ^{3,4}

¹School of Agricultural Equipment Engineering, Jiangsu University, China

²School of Food and Biological Engineering, Jiangsu University, China

³International Research Center for Food Nutrition and Safety, Jiangsu University, China

⁴Department of Analytical Chemistry and Food Science, University of Vigo - Ourense Campus, Spain

ARTICLE INFO

Article History

Received 15 October 2021

Accepted 9 November 2021

Keywords

Vinegar fermentation
Microbial community
Hyperspectral imaging technology
Strain colony identification
Strain colony counting

ABSTRACT

Hyperspectral imaging technology with chemometrics was used for identifying and counting each species in microbial community during mixed fermentation. Hyperspectral images of microbial community of *Enterobacter* sp, *Acetobacter pasteurianus*, and *Lactobacillus paracasei* colonies were obtained and the spectra of strain colonies were extracted. Identification models were developed using linear discriminant analysis (LDA) and least-squares support vector machine (LS-SVM) by using 23 variables selected by genetic algorithm. The optimal LS-SVM model with an identification rate of 96.67 % was used to identify colonies and prepare colony distribution maps in color for strains counting. The counting results by hyperspectral imaging technology agreed with that of the manual counting method with an average relative error of 3.70 %. The developed counting method has been successfully used to identify and count the specific strain from the mixed strains simultaneously. The hyperspectral imaging technology has a great potential to monitor changes in the microbial community structure.

© 2021 The Authors. Publishing services by Visagaa Publishing House

This is an open access article distributed under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).

1. INTRODUCTION

Mixed fermentation occurs in many Chinese traditional fermented food, such as vinegar, wine, and soy sauce [1]. It involves more than one microbial strain and many different types of interactions between them exist during fermentation [2]. Vinegar is produced from glutinous rice, wheat bran, and rice hull in a special environment through mixed fermentation of various microorganisms [3, 4]. Numerous studies conducted to analyze the fermentation microflora in vinegar fermentation indicated that *Lactobacillus*, *Acetobacter*, and *Enterobacter* were the dominant bacteria in fermentation stage [5, 6]. *Acetobacter* is involved in the oxidation of ethanol to acetic acid as the main ingredient of vinegar [7]. *Lactobacillus* greatly contributes to the formation of lactic acid, which plays a crucial role in moderating the irritating odor of vinegar [8]. *Enterobacter* can produce a variety of substances in terms of aroma and flavor (such as amylase and lipase). Many vinegar industries always made substandard vinegar products because of unbalanced flora structure causing insufficient or over fermentation. Structure of microbial community can strongly influence the quality and characteristics of the mixed-culture fermentation of vinegar [3, 9]. Therefore, monitoring of the microbial diversity is crucial to know the fermentation state and control vinegar quality.

Traditional culture-dependent method and molecular fingerprinting techniques were widely used to investigate the microbial community in fermentation [10, 11]. Traditional culture-dependent methods require a minimal investment in terms of the equipment and agar cultures, which are relatively low cost [12]. However, this method has a further limitation in counting a specific strain of colonies from mixed strain colonies, which have similar morphologies as well as physiological and biochemical characteristics [13, 14]. Recently, much effort has been focused on development of molecular fingerprinting techniques. Their application in bacterial identification is limited for high costs and the requirement for trained personnel. In addition, it is difficult for the above methods to identify and count the mixed bacteria simultaneously. Hyperspectral imaging technique, as rapid and nondestructive method integrates computer vision and spectral analysis techniques to acquire spectral information of each pixel location in a two-dimensional (2D) objective image, generating a multi-dimensional data (one-dimensional spectral dimension and two-dimensional spatial information) [15, 16]. Each strain colony with unique fingerprints spectrum has been identified successfully by taking full advantage of one-dimensional spectral information [17–19]. Although hyperspectral imaging technology is very useful for strain identification during mixed fermentation, there are still exist a major issue: rapid and automated

*Corresponding author. Email: shi_jiyong@ujs.edu.cn

Peer review under responsibility of the International Association of Dietetic Nutrition and Safety

enumeration of multiple strains. Some research groups studied the visual distribution of target combing the one-dimensional spectral information and two spatial dimensions. The specific strain colonies can be positioned and marked considering the two-dimensional spatial information. Therefore, hyperspectral imaging technology can realize rapid identification and visual counting of strains in mixed fermentation to monitor microbial community.

Hyperspectral imaging technique was investigated to establish a microflora detection system for mixed fermentation. Three dominant strains (*Acetobacter pasteurianus*, *Lactobacillus paracasei* and *Enterobacter* sp.) were used in mixed fermentation to simulate the mixed fermentation of vinegar. Their colonies will be recognized and each strain can be counted after plating culture. It is expected to provide a new way for rapid monitoring the microbial community in mixed fermentation.

2. MATERIALS AND METHODS

2.1. Bacterial strains and culture conditions

This study was conducted using three strains, namely *Acetobacter pasteurianus* (China Center of Industrial Culture Collection (CICC) 20064), *Lactobacillus paracasei* (CICC 7000), and *Enterobacter* sp. (CICC 10595) from China Center of Industrial Culture Collection. These strains were classified as *Acetobacter* sp., *Lactobacillus* sp., and *Enterobacter* sp., respectively. The strains were revived to live cells from frozen stocks through culturing in a sterilized Luria-Bertani (LB) broth (1% tryptone, 0.5% yeast extract, and 1% NaCl) for 24 h at 37°C. A series of diluted suspensions (10^3 -, 10^4 -, 10^5 - and 10^6 -fold,) and their mixed suspensions with same volume were prepared with sterile water. The pure and mixed suspensions (0.1 mL) were spread onto LB agar plates (1% tryptone, 0.5% yeast extract, 2% agar, and 1% NaCl) in 9 cm petri dishes and incubated at 37°C for 24 h. Finally, these pure ($2 \times 3 \times 4 = 24$) and mixed ($2 \times 4 = 8$) plates are prepared in duplicate. The experimental procedures were performed in triplicate to evaluate repeatability ($3 \times 24 = 72$ plates for pure plates, $3 \times 8 = 24$ plates for mixed plates).

2.2. Acquisition of hyperspectral image

Data acquisition was conducted using a hyperspectral imaging system developed by the Agricultural Product Processing and Storage Lab at Jiangsu University in China. This system consists of mainly a line-scan spectrograph (ImSpector, VI7E, Spectra Imaging Ltd., Finland) with a spectral range of 431 ~ 962 nm in 618 spectral bands and a resolution of approximately 0.858 nm; a CMOS camera (XEVA-FPA-1.7-320, XenIcs Ltd., Belgium) with a resolution of 1628×618 pixels; a direct-current illuminator (2900, Illumination Technologies Inc., USA); a conveyor belt (Zolix TS200AB, Zolix. Corp., China); data acquisition and preprocessing software (Spectra-Cube, Auto Vision Inc., USA), an enclosure; and a computer. For more details regarding the imaging system, refer to the study of Zou et al [20].

The agar plates were placed on the conveyor for acquisition of hyperspectral images, and a hypercube was created using spatial ($a \times b$ pixels) and spectral information (m wavebands). The

information was represented by a three-dimensional (3D) image with x-axis, y-axis, and L-axis coordinate information. A single hyperspectral image of each agar plate was stored in raw form before being processed, and the intensities of the images varied with their wavelengths. Hyperspectral images can include signal noise such as that caused by illumination and detector sensitivity. For calibration of the reflectance intensity values at each pixel of the hyperspectral images, the reflectance spectrum at every pixel was calculated using a “dark” image (approximately 0% reflectance) and a “white” image (approximately 99.9% reflectance). The details of hyperspectral calibration can be found in our previous study [21].

2.3. Spectral data extraction

After image calibration, rectangular regions of interest (ROIs) were extracted from the hyperspectral images of the agar plates by using ENVI 4.5 (ITT Visual Information Solutions, Boulder, CO, USA) software. A ROI with a size of 15×15 pixels was selected from the center of each colony. The species of the colonies in each image were estimated from the responses of each pixel in the ROI. Simultaneously, the average relative reflectance of each image was calculated using the spectral response of each pixel in the ROI. Then, five spectra from five colonies of the same species were averaged into one spectrum. Eventually, 90 average reflectance spectra were obtained from the three strains. All the spectrum data were randomly divided into calibration and prediction datasets in a ratio of 2:1. The calibration dataset is used for building calibration models for identifying colony regions and the prediction dataset is for testing the robustness of calibration models.

2.4. Development of identification models

The hyperspectral images acquired using the hyperspectral imaging system present 3D data, including numerous spatial and spectral data. The spectral data between sequential wavelengths are closely related and contain redundant features [22]. Employing spectral data directly to develop a calibration model is difficult without using chemometrics. Feature wavelength selection is crucial and has a profound influence on the performance of a model. In this study, linear discriminant analysis (LDA) and a least-squares support vector machine (LS-SVM) were employed to build calibration models for colony identification, and a genetic algorithm (GA)-partial least-squares (PLS) (GA-PLS) algorithm was applied to select feature wavelengths from the spectral data to optimize the LS-SVM and LDA procedures.

2.4.1. Spectral data processing

During the acquisition of the hyperspectral images, spectral information is affected by high-frequency random noise and background information, such as illumination intensity and dark current, when the same sample is tested repeatedly [23]. This leads to instability and unreliability in calibration models. In this study, several spectral pretreatment methods, including standard normal variate (SNV), multiplicative scatter correction, first order derivative, and second order derivative were employed to manage the raw data. Characteristic wavelength selection

In this study, a GA was applied to select the feature wavelengths, and the PLS method was applied to develop regression models from the characteristic wavelengths. The characteristic wavelengths resulting in the lowest value in terms of the root mean square error of cross-validation (RMSECV) or the highest value in regression coefficients were selected [24]. Cross-validation was used during the GA procedure. Parameters were initialized randomly in the first stage and then are updated using evolutionary methods. The number of runs, initial population size, recombination rate, and mutation rate were set to 130, 30, 0.50, and 1% according to previous literature [25], respectively.

2.4.2. Characteristic wavelengths selection

In this study, a GA was applied to select the feature wavelengths, and the PLS method was applied to develop regression models from the characteristic wavelengths. The characteristic wavelengths resulting in the lowest value in terms of the root mean square error of cross-validation (RMSECV) or the highest value in regression coefficients were selected [24]. Cross-validation was used during the GA procedure. Parameters were initialized randomly in the first stage and then are updated using evolutionary methods. The number of runs, initial population size, recombination rate, and mutation rate were set to 130, 30, 0.50, and 1% according to previous literature [25], respectively.

2.4.3. Development of identification models

The performance of the final model was estimated using the correlation coefficient of calibration (R_c) and the root mean square error of cross-validation (RMSECV) and was tested using the correlation coefficient of prediction (R_p) and the root mean square error of prediction (RMSEP) [26]. In general, a good model should have high R_c and R_p values with low RMSECV and RMSEP values.

LDA is a pattern recognition method that reduces the dimensions of a matrix to extract the characteristic variances and condenses the information into latent variances [27]. It is a supervised learning method and has been successfully implemented to identify strains. In LDA, optimal transformation is conducted to minimize within-class distance and maximizes between-class distance simultaneously by detecting the affinity of a set of unknown samples and separating the objects into clusters or groups, thus achieving maximum discrimination. The latent variances in LDA are linear combinations of the raw variances. The first two LDA functions are named LD1 and LD2. A 2D spatial expression of LD1 and LD2 is obtained to reveal the cluster trends of the training dataset [28]. After clustering, the spectra obtained by SNV processing and a GA were applied to establish the LDA classification model.

The LS-SVM algorithm was proposed by Suykens and Vandewalle and is a modified support vector machine. LS-SVMs are powerful for rapid linear and nonlinear multivariate calibration while maintaining the advantages of a support vector machine. LS-SVMs can use a kernel function to map nonlinear data and input variables into a high-dimensional feature space, thus transforming the optimization problem into a problem of satisfying equality constraints [29]. The determination of an appropriate kernel function and optimum kernel parameters are crucial to the results of

an LS-SVM model. A radial basis function (RBF) kernel is the most efficient function for minimizing the computational complexity of a procedure and could be applied to manage the linear and nonlinear relationships between spectra and target attributes [30]. Therefore, an LS-SVM combined with an RBF kernel was used in this study. To obtain high performance, leave-one-out cross-validation was employed to obtain the optimal combination of the regularization parameter (γ) and kernel function parameter (σ^2) in terms of the lowest cost value.

2.5. Automated count of microbial colonies

The mixed strain colonies were identified by the calibration models and counted with assisted of orientation and marking. Suspensions of three kinds of bacteria were mixed in any proportion and then the mixture was diluted (10^5 -fold) with sterilized water. Finally, 0.1 mL of mixture was smeared onto LB agar plates (1% tryptone, 0.5% yeast extract, 2% agar, and 1% NaCl) in 9 cm petri dishes and incubated at 37°C for 24 h. To count microbial colonies in mixed colony plates, hyperspectral images of colony plates were acquired and spectral data of each pixel were extracted from hyperspectral images at the selected feature wavelengths. Then, the spectral data of each pixel were substituted into the optimal calibration model. The category of each pixel was identified and different strain colonies were visualized with different colors. The microbial colonies on agar plates can be counted according to the colors. The key steps for analyzing hyperspectral imaging data are illustrated in **Figure 1**.

2.6. Software

Spectral Cube (ImSpector, Auto Vision Inc., USA) was used for acquiring and storing the hyperspectral images of the agar plates. ENVI 4.5 (ITT Visual Information Solutions, Boulder, CO, USA) was used to extract the spectral reflectance from the hyperspectral images. Matlab V7.0 (Math Works, USA) was used on a computer with Windows 7 (Microsoft) to conduct image and data analysis.

3. RESULTS AND DISCUSSION

3.1. Appearance of the colonies of different strains

Colonies of *A. pasteurianus*, *L. paracasei*, and *Enterobacter* sp. growing on LB agar after incubation for 24 h at 37 °C is shown in **Figure 2**. For the colonies of *A. pasteurianus* (**Figure 2 a**), the grown colonies are not homogeneous, and the center and edges differ within colonies of the same strain. It is the same as the colonies of *L. paracasei* (**Figure 2 b**), and *Enterobacter* sp. (**Figure 2 c**). By comparing the three stain colonies, each strain colony is roughly circular but the shape and size differences between each strain colony are not great and clear. Although the pseudo-color maps are different, it is difficult to distinguish them by observing their colors. It is because some parts of them have same colors. It was not easy to use the colony shape, size, and color features to segment the specific colonies in the imaging process. Thus, identification and enumeration of specific species from a mixed strains culture cannot be implemented by the manual counting method and automated

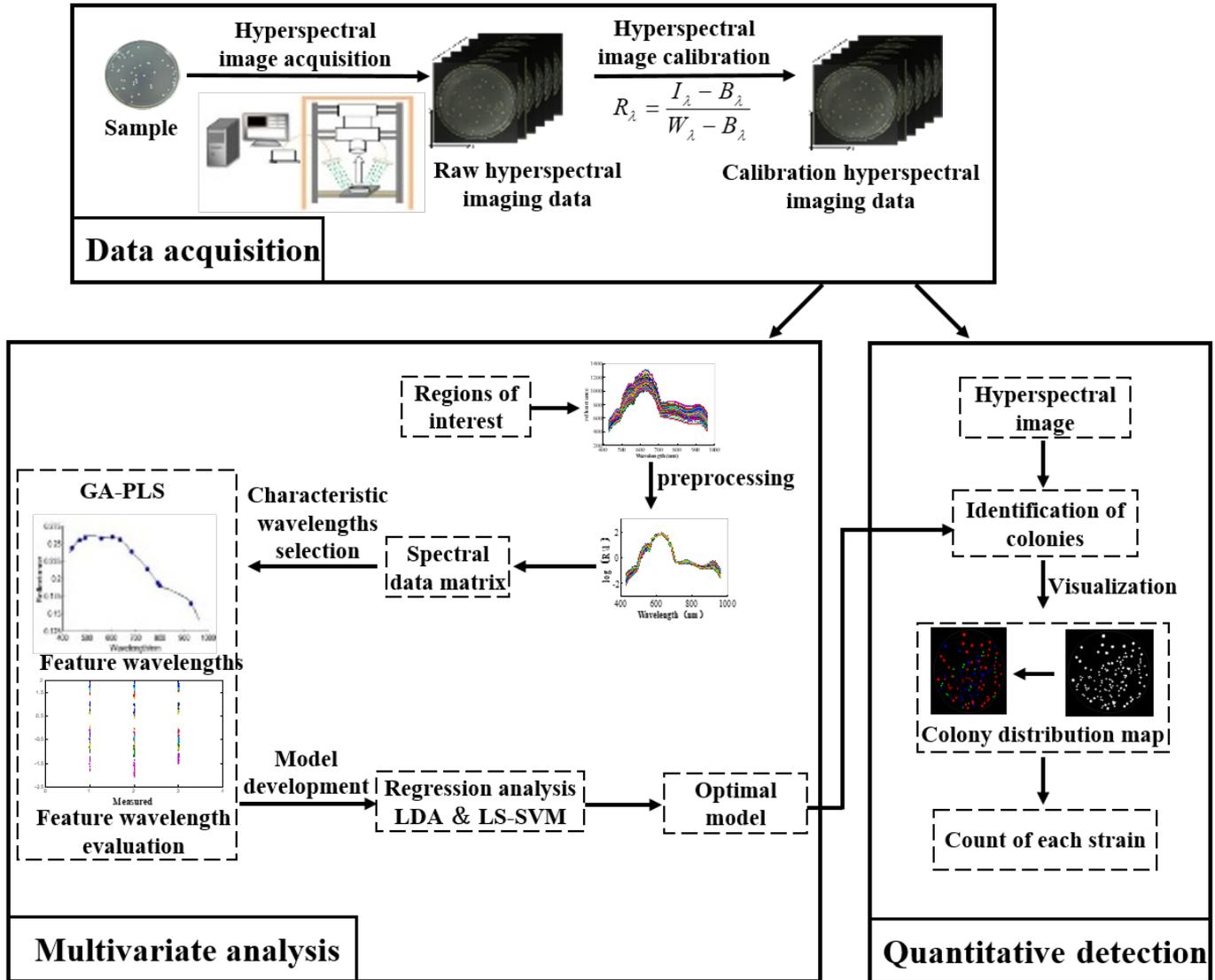


Figure 1 | Key steps for analyzing hyperspectral imaging data and counting strain colonies

counting method with the aid of digital camera images.

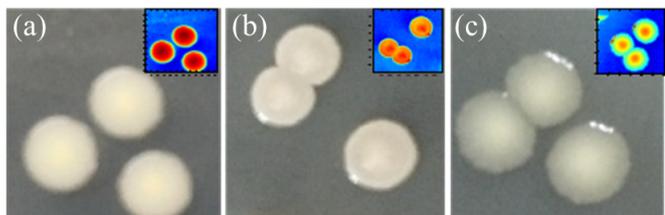


Figure 2 | Digital images of colonies of *A. pasteurianus* (a), *L. paracasei* (b), and *Enterobacter* sp. (c), the insets display the pseudo-color maps of each strain colonies

3.2. Spectra of strain colonies

The spectral curves of the three strain colonies on the agar plates are extracted from the hyperspectral images and presented in Figure 3. Figure 3 a and b reveal the raw reflection spectra and mean reflectance spectra. These spectra of colonies varied with the wavelengths and displayed reflectance peaks (Figure 3 a). The spectral curves at wavelengths of 550 ~ 700 nm exhibited high reflectance, and an obvious reflectance peak was observed at approximately 670 nm. Moreover, a weak reflectance peak was observed at approximately 980 nm and could be attributed to the second O-H stretching of carbohydrates and water [31]. Reflectance spectra are sensitive to changes in the cellular components and structure of the bacterial metabolic product that are related to the response of the X-H bond (such as O-H, C-H and N-H). Each mean spectrum was acquired from 30 spectra of the corresponding strain colonies. As seen from the mean spectra, the spectral curves of strain colonies showed similar trends in wavelengths of 550 ~ 700 nm. The reflection signal responses of each strain colonies exhibited

conspicuous discrepancies in peak strength and peak shape. Despite of these difference, it is still difficult to identify the target strain by using naked eyes. Therefore, the spectral features combined with chemometrics would be used for identifying strains.

The spectra of strain colonies were preprocessed by different pretreatment methods and the spectra after SNV processing were used for further analysis (Figure 3 c). By comparison with several pretreatment methods, SNV processing achieved the highest performance for building identification models. The SNV processing converted each spectrum to a zero-mean intensity or absorbance value with a standard deviation of one [32]. SNV was used to eliminate slope variation and to correct scatter effects in terms of light scattering and changes in light path length [33].

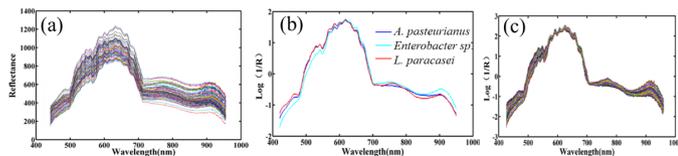


Figure 3 Meanreflectance spectra, (b) raw reflectance spectra, and (c) spectra after SNV preprocessing for different bacterial colonies

3.3. Characteristic wavelengths selection

In this study, GA combined with PLS method was used to process spectral data for selecting characteristic variables. The resulting regression coefficients and RMSECV are used to evaluate the target response of the dataset [34]. Twenty-three wavelengths (583, 580, 582, 587, 434, 585, 638, 617, 697, 774, 861, 433, 435, 486, 518, 558, 561, 584, 591, 639, 692, 753, and 908 nm) were selected as characteristic wavelengths. The RMSECV and RMSEP values of the PLS models were 0.114 and 0.156, respectively. The identification rate of prediction dataset achieved 98.2%, demonstrating the identification model established by characteristic wavelengths had high capability to identify different strain colonies. Moreover, the obtained calibration original from the new matrix with less variance selected by GA was more simple and robust than that from full spectral data. The results suggest that GA is a feasible method to select feature wavelengths for establishment of the optimal calibration model. Then, the new data matrix will be applied to establish an identification model for further analysis in this study. This process was performed thrice to avoid stochastic influence.

3.4. Development of identification models

3.4.1. Results of LDA model

A LDA model was established on the basis of the 23 characteristic wavelengths with identification rates of 95.00% and 90.00% for the calibration and prediction sets, respectively. LDA is a dimension reduction method, and the plane with the smallest dimension is determined from a dataset such that the objects related to the plane are mapped from a higher dimensional space into a lower dimensional space [35]. Figure 4 displays a clear cluster trend of the three species along the top-two score plot of the LDA (LD1 and LD2). LD1 and LD2 explain 80.03% and 8.47% of the variance,

respectively. The accumulated variance contribution rate is 88.50%. Three strains could be well separated by using the LD1 and LD2 planes. *Enterobacter* sp. gathered in an independent area implying that it could be entirely distinguished from other two strain colonies. *L. paracasei* and *A. pasteurianus* had mutual overlaps in a small part of cluster area. The score cluster plot demonstrated that the three strains could be distinguished by the spectral features processed by LDA. After clustering, the top ten scores (LD1, LD2, LD3, LD4, LD5, LD6, LD7, LD8, LD9, and LD10) were used to establish an identification model (LDA model). The recognition rates (Rc and Rp) are 95% and 90% in the calibration set and prediction set, respectively (Table S1Appendix A). *Enterobacter* sp. colonies were correctly identified with 100% of recognition rate *A. pasteurianus* and *L. paracasei* were mutually identified as each other, with more than 80% and 90% of recognition rate, respectively.

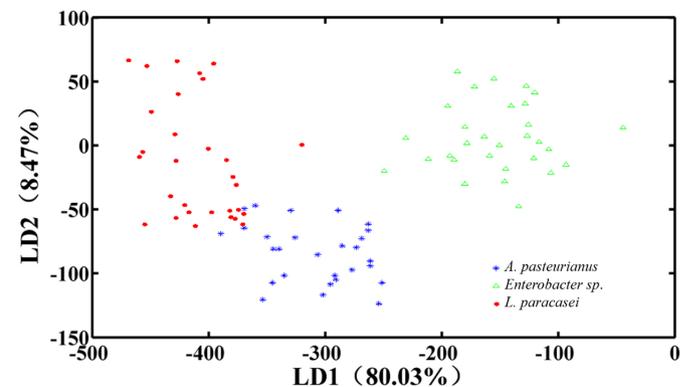


Figure 4 Score cluster plot with LD1 and LD2 values for three strain colonies

3.4.2. Results of LS-SVM model

In this study, LS-SVM was used to develop LS-SVM model based on the 23 characteristic wavelengths. Figure S1Appendix A presents a contour chart in which different cost values are displayed using different colors. Here, the best combination of two parameters (γ and σ^2) was searched during the two stages. First, a large stage in a 10×10 grid (represented as “.”) was implemented. Subsequently, a smaller stage in a 10×10 grid (represented as “x”) was applied to obtain the optimal combination of the two parameters. In both stages, the lowest cost value was determined by using the contour chart. The best parameters $\gamma = 0.8914$ and $\sigma^2 = 7.7791$ were obtained for establishment of LS-SVM model. The recognition rates of LS-SVM model were 100.00% and 96.67% the calibration set and prediction sets, respectively. As listed in Table S1Appendix A, one of *A. pasteurianus* colonies was predicted incorrectly among the prediction samples. *Enterobacter* sp. and *L. paracasei* were correctly identified with 100% of recognition rate. This error may be caused by similar metabolite production-molecular structure or chemical components—between two samples. The excellent recognition rates demonstrated that the three strain colonies could be completely distinguished by combining GA and LS-SVM.

Table 1 Results of colonies counting in pure cultures and mixed cultures

Sample	Species	Pure culture (CFU/mL)		Mixed cultures (CFU/mL)		Average relative error (%)
		Manual method	This method ^a	Manual method	This method ^a	
1	<i>Enterbacter sp.</i>	40	40 ± 3	41	40 ± 3	1.25
	<i>A. pasteurianus</i>	27	29 ± 2	28	29 ± 2	5.29
	<i>L. paracasei</i>	19	18 ± 1	20	18 ± 2	7.63
2	<i>Enterbacter</i>	37	38 ± 1	43	43 ± 1	1.35
	<i>A. pasteurianus</i>	24	25 ± 2	28	27 ± 3	3.87
	<i>L. paracasei</i>	28	26 ± 3	27	28 ± 2	5.42
3	<i>Enterbacter</i>	56	56 ± 3	59	58 ± 3	0.85
	<i>A. pasteurianus</i>	50	52 ± 3	53	50 ± 2	4.83
	<i>L. paracasei</i>	47	45 ± 2	43	46 ± 3	5.62
4	<i>Enterbacter</i>	40	40 ± 1	46	46 ± 1	0
	<i>A. pasteurianus</i>	62	60 ± 2	62	64 ± 1	3.22
	<i>L. paracasei</i>	40	42 ± 2	45	43 ± 1	4.72
5	<i>Enterbacter</i>	52	52 ± 3	54	55 ± 1	0.93
	<i>A. pasteurianus</i>	58	55 ± 3	61	58 ± 3	5.04
	<i>Enterbacter</i>	47	50 ± 2	49	51 ± 3	5.23

^a: Mean ± difference

3.4.3. Comparison of identification models

Table S1 Appendix A lists the recognition results of the LS-SVM and LDA models based on the selected 23 characteristic wavelengths. As observed from this table, the LS-SVM model provided recognition rates of 100% for the calibration dataset and 96.67% for the prediction dataset. The LDA model provided recognition rates of 95.00% for the calibration dataset and 90.00% for the prediction dataset. The LS-SVM model achieved higher performance than the LDA model. This is possibly because the growth of species and metabolite production are complicated and may involve nonlinear characteristics. The nonlinear LS-SVM may provide superior solutions to those of LDA. Therefore, the LS-SVM model would be used for further analysis.

4. AUTOMATED COUNT OF MICROBIAL COLONIES

A colony distribution map was developed to enumerate the confluent colonies on the agar plate by using hyperspectral imaging technology and the established LS-SVM model (**Figure 5** and **Table 1**). As seen from the colony characteristics, *Enterobacter sp.* colonies were different from the other two strain colonies. However, *A. pasteurianus*, and *L. paracasei* could not be easily distinguished by naked eyes base on the colony characteristics. Three strain colonies on hyperspectral images were identified, namely *Enterobacter sp.*, *A. pasteurianus*, and *L. paracasei* according to the LS-SVM model. To fix the position of the colonies, image binary method was used to separate colonies from background combined with the two-dimensional in hyperspectral images (**Figure 5b**). Finally, the bacterial colonies were marked in specific color, forming colony distribution map (**Figure 5c**). Red, blue, and green represented *Enterobacter sp.*, *A. pasteurianus*, and *L. paracasei* in colony distribution map and there are 40 red circles, 29 blue circles and 18 green circles (**Figure 5c**). The diluted suspension with gradient of 10^{-6} was cultured in mixed colonies plates (**Figure 5a**). That is to

say: microbial community of the mixed suspension was consisted of *Enterobacter sp.* (40×10^6 CFU/mL), *A. pasteurianus* (29×10^6 CFU/mL), and *L. paracasei* (18×10^6 CFU/mL). The quantity ratio of the three strains was 40: 29: 18. The hyperspectral image technology is capable of segmenting and counting different species of bacteria incubating on the same plate on the basis of spectral features. This cannot be easily performed through the traditional methods.

Table 1 lists the counting results of pure and mixed cultures through manual counting method and counting method based on hyperspectral image technology. Thirty agar plates of pure and mixed strain were analyzed by the developed counting method. For detection, pure cultures of *Enterobacter sp.*, *A. pasteurianus*, and *L. paracasei* were inoculated in three separate petri dishes, and a mixture of *Enterobacter sp.*, *A. pasteurianus*, and *L. paracasei* with same concentration was inoculated in another petri dish. The counting results of developed method was in according with that of the manually counting method with an average relative error of 3.70%. The results of hyperspectral and manual counting were agreeable. Colonies of *L. paracasei* counting has the maximum relative error (7.63%) and colonies of *Enterobacter sp.* counting has the minimum relative error (0%). The identification of *Enterobacter sp.* was the most accurate among the identification of the three strain colonies. This result indicates that the hyperspectral image technology provides an accurate estimation of the number of colonies. The counting error might be caused by the competition of confluent colonies and the unclear colonies boundaries. Moreover, error in the dilution procedure and model discrimination error might have occurred.

5. CONCLUSION

In this study, hyperspectral imaging technology was successfully used for enumerating each species in a microbial community and monitoring the changes in the mixed fermentation. Pure and mixed

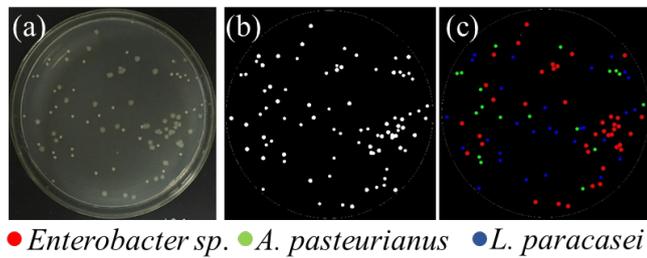


Figure 5 (a) Digital image of the three strain agar plates, (b) the segmentation binary image, and (c) colonies distribution map of each strains

colonies on agar plates were prepared and hyperspectral images were captured. Hyperspectral image data of the bacteria were processed by using SNV, GA, LDA, and LS-SVM. LDA and LS-SVM identification models were established on the basis of characteristic variables and achieved recognition rates of 90.00% and 96.67%, respectively. The optimal identification model (LS-SVM) was used to separate each bacterial colony from the background and identify strain colonies in images. Colony distribution maps were obtained displaying the three strains with different colors. The number of bacteria of each species was determined according to the colorful colonies, and the results were similar to those obtained through the manual counting method. A maximum relative error of 7.63% and an average relative error of 3.70% indicated that the hyperspectral image technology is useful for distinguishing and counting bacteria simultaneously. This suggests the feasibility of hyperspectral imaging technology to rapidly monitor microbial communities.

6. ETHICAL APPROVAL

This article does not contain any studies with human participants or animals performed by any of the authors.

7. INFORMED CONSENT

Informed consent is not applicable for the nature of this study.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENTS

This work was finally supported by the Postdoctoral Science Foundation of China (2021M691319), the Project of Agricultural Equipment Department of Jiangsu University (NZXB20200212), the Natural Science Foundation of Jiangsu Province (BE2019359), Distinguished Professor Foundation of Jiangsu Province (202074), Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

A. APPENDIX - SUPPLEMENTARY DATA

Supplementary data to this article can be found online at <https://doi.org/10.53365/efood.k/143830>.

REFERENCES

- [1] Zhang Q. Monitoring microbial succession and metabolic activity during manual and mechanical solid-state fermentation of Chinese cereal vinegar. *LWT*. 2020;133:109868. doi:10.1016/j.lwt.2020.109868.
- [2] Liu SN, Han Y, Zhou ZJJFRI. Lactic acid bacteria in traditional fermented Chinese foods. *Food Research International*. 2011;44:643–651. doi:10.1016/j.foodres.2010.12.034.
- [3] Gan X, Tang H, Ye D, Li P, Luo L, Lin W. Diversity and dynamics stability of bacterial community in traditional solid-state fermentation of Qishan vinegar. *Diversity and dynamics stability of bacterial community in traditional solid-state fermentation of Qishan vinegar*. 2017;67:703–713. doi:10.1007/s13213-017-1299-6.
- [4] Fang GY. Deciphering the succession patterns of bacterial community and their correlations with environmental factors and flavor compounds during the fermentation of Zhejiang rosy vinegar. *International Journal of Food Microbiology*. 2021;341:109070. doi:10.1016/j.ijfoodmicro.2021.109070.
- [5] Wei X. Analysis of the microbial diversity and function during acetic acid fermentation process of Zhenjiang Aromatic Vinegar. Wuxi: Jiangnan University; 2011.
- [6] Li S, Li P, Feng F, Luo LX. Microbial diversity and their roles in the vinegar fermentation process. *Applied microbiology and biotechnology*. 2015;99:4997–5024. doi:10.1007/s00253-015-6659-1.
- [7] Fernández-Pérez R, Torres C, Sanz S, Ruiz-Larrea F. Strain typing of acetic acid bacteria responsible for vinegar production by the submerged elaboration method. *Food microbiology*. 2010;27:973–978. doi:10.1016/j.fm.2010.05.020.
- [8] Jia JW, Ma YK, Zhang FF, Chen FS. Culture-dependent and culture-independent analysis of lactic acid bacteria from Shanxi aged vinegar. *Annals of Microbiology*. 2012;62(4):1825–1830. doi:10.1007/s13213-011-0396-1.
- [9] Thomas L, Larroche C, Pandey A. Current developments in solid-state fermentation. *Biochemical Engineering Journal*. 2013;81:146–161. doi:10.1016/j.bej.2013.10.013.
- [10] Rajeev M, Sushmitha T, Pandian SK. Culture-Dependent and-Independent Strategies in Bacterial Diversity Appraisal. *Microbial Systematics*. 2020;p. 1–27.
- [11] Demirci T, Göktepe ÇK, Öztürk Hİ. Prevalence and fingerprinting of lactic acid bacteria community during 180 days of ripening in traditional Turkish goatskin bag Tulum cheeses produced in the mountainous region of Karaman using culture-dependent and-independent methods. *International Dairy Journal*. 2021;118:105041. doi:10.1016/j.idairyj.2021.105041.
- [12] Salfinger Y, Tortorello ML. *Compendium of methods for the Microbiological Examination*. Washington: American Public Health Association; 2018. doi:10.2105/MBEF.0222.
- [13] Chen WB, Zhang C. An automated bacterial colony counting and classification system. *Information Systems Frontiers*. 2009;11(4):349–368. doi:10.1007/s10796-009-9149-0.
- [14] Ferrari A, Lombardi S, Signoroni AJPR. Bacterial colony counting with convolutional neural networks in digital microbiology imaging. *Pattern Recognition*. 2017;61:629–640. doi:10.1016/j.patcog.2016.07.016.
- [15] Shi J. Noise-free microbial colony counting method based on hyperspectral features of agar plates. *Food Chemistry*.

- 2019;274:925–932. doi:10.1016/j.foodchem.2018.09.058.
- [16] Liu Y, Pu H, Sun DW. Hyperspectral imaging technique for evaluating food quality and safety during various processes: A review of recent applications. *Trends in Food Science & Technology*. 2017;69:25–35. doi:10.1016/j.tifs.2017.08.013.
- [17] Foca G, Ferrari C, Ulrici A, Sciotto G, S P, S M. The potential of spectral and hyperspectral-imaging techniques for bacterial detection in food: A case study on lactic acid bacteria. *Talanta*. 2016;153:111–119. doi:10.1016/j.talanta.2016.02.059.
- [18] Feng YZ, Yu W, Chen W, Peng KK, Jia GF. Invasive weed optimization for optimizing one-agar-for-all classification of bacterial colonies based on hyperspectral imaging. *Sensors and Actuators B: Chemical*. 2018;269:264–270. doi:http://doi.org/10.1016/j.snb.2018.05.008.
- [19] Shi J, X H, X Z, Holmes M, Tahir HE, Huang X, et al. Rapid identification of *Lactobacillus* species using near infrared spectral features of bacterial colonies. *Journal of Near Infrared Spectroscopy*. 2019;27:302–313. doi:10.1177/0967033519852012.
- [20] Shi J, Wang Y, Liu C, Li Z, Huang X, Guo Z, et al. Application of spectral features for separating homochromatic foreign matter from mixed congee. *Food Chemistry: X*. 2021;11:100128. doi:10.1016/j.fochx.2021.100128.
- [21] JiyongShi, Hu X, Zou X, Zhao J, Zhang W, Holmes M, et al. A rapid and nondestructive method to determine the distribution map of protein, carbohydrate and sialic acid on Edible bird's nest by hyper-spectral imaging and chemometrics. *Food Chemistry*. 2017;229:235–241. doi:10.1016/j.foodchem.2017.02.075.
- [22] Qureshi R, Uzair M, Khurshid K, Yan H. Hyperspectral document image processing: Applications, challenges and future prospects. *Pattern Recognition*. 2019;90:12–22. doi:10.1016/j.patcog.2019.01.026.
- [23] Zhang C, Liu F, Yong H. Identification of coffee bean varieties using hyperspectral imaging: Influence of preprocessing methods and pixel-wise spectra analysis. *Scientific Reports*. 2018;8:2166. doi:10.1038/s41598-018-20270-y.
- [24] Stefansson P. Fast method for GA-PLS with simultaneous feature selection and identification of optimal preprocessing technique for datasets with many observations. *Journal of Chemometrics*. 2020;34:3195–3195. doi:10.1002/cem.3195.
- [25] Hassanat A, Almohammadi K, Alkafaween E, Abunawas E, Hammouri A, Prasath VBS. Choosing mutation and crossover ratios for genetic algorithms-a review with a new dynamic approach. *Information*. 2019;10:390. doi:10.3390/info10120390.
- [26] Jia S. Feasibility of analyzing frost-damaged and non-viable maize kernels based on near infrared spectroscopy and chemometrics. *Journal of Cereal Science*. 2016;69:145–150. doi:10.1016/j.jcs.2016.02.018.
- [27] Bianco MJ. Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*. 2019;146:3590–3628. doi:10.1121/1.5133944.
- [28] Shen T, Xiaobo Z, Jiyong S, et al. Determination Geographical Origin and Flavonoids Content of Goji Berry Using Near-Infrared Spectroscopy and Chemometrics. *Food Analytical Methods*. 2016;9(1):68–79. doi:10.1007/s12161-015-0175-x.
- [29] Lee H, Kim MS, Lim HS, Park E, Lee WH, Cho BK. Detection of cucumber green mottle mosaic virus-infected watermelon seeds using a near-infrared (NIR) hyperspectral imaging system: Application to seeds of the “Sambok Honey” cultivar. *Biosystems Engineering*. 2016;148:138–147. doi:10.1016/j.biosystemseng.2016.05.014.
- [30] Li H, Kutsanedzie F, Zhao J, Chen Q. Quantifying total viable count in pork meat using combined hyperspectral imaging and artificial olfaction techniques. *Food Analytical Methods*. 2016;9:3015–3024. doi:10.1007/s12161-016-0475-9.
- [31] Yeh YH, Chung WC, Liao JY, Chung CL, Kuo YF, Lin TT. Strawberry foliar anthracnose assessment by hyperspectral imaging. *Computers and Electronics in Agriculture*. 2016;122:1–9. doi:10.1016/j.compag.2016.01.012.
- [32] Li Y, Xiaobo Z, Tingting S, Jiyong S, Jiewen Z, Holmes M. Determination of Geographical Origin and Anthocyanin Content of Black Goji Berry (*Lycium ruthenicum* Murr.) Using Near-Infrared Spectroscopy and Chemometrics. *Food Analytical Methods*. 2016;10(4):1–11. doi:10.1007/s12161-016-0666-4.
- [33] Sitorus A, Muslih M, Cebro IS, Bulan R. Dataset of adulteration with water in coconut milk using FTIR spectroscopy. *Dataset of adulteration with water in coconut milk using FTIR spectroscopy*. 2021;36:107058–107058. doi:10.1016/j.dib.2021.107058.
- [34] Stefansson P, Liland KH, Thiis T, Burud I. Fast method for GA-PLS with simultaneous feature selection and identification of optimal preprocessing technique for datasets with many observations. *Journal of Chemometrics*. 2020;34:3195. doi:10.1002/cem.3195.
- [35] Patra SS. Emerging Healthcare Problems in High-Dimensional Data and Dimension Reduction. *Advanced Prognostic Predictive Modelling in Healthcare Data Analytics*. 2021;p. 25–49. doi:10.1007/978-981-16-0538-3_2.