Т. 87, № 2 **ЖУРНАЛ ПРИКЛАДНОЙ СПЕКТРОСКОПИИ** МАРТ — АПРЕЛЬ 2020
V. 87, N 2 **JOURNAL OF APPLIED SPECTROSCOPY** MARCH — APRIL 2020

COMBINING HYPERSPECTRAL IMAGING AND FEATURE WAVELENGTH EXTRACTION METHODS FOR THE RAPID DISCRIMINATION OF RED MEAT

D. Ding, K. Liang, B. Li, L. Liu, W. Wu, M. Shen*

College of Engineering, Nanjing Agricultural University,
Nanjing 210031, China; e-mail: dingdong@njau.edu.cn; mingxia@njau.edu.cn

A hyperspectral imaging system (400–800 nm) combined with multivariate analyses was investigated to discriminate between beef, pork, and mutton species based on the feature wavelengths of intact and minced samples. The performances of classification models constructed by combining linear discriminant analysis (LDA), partial least squares discriminant analysis (PLS-DA), or a support vector machine (SVM) with a variable selection method, such as a successive projection algorithm (SPA), regression coefficient analysis (RCA), or random frog (RF), were compared. The results clearly showed that the linear classifier was preferred to the nonlinear classifier in the identification of red meat species. Furthermore, instead of selecting different sets of feature wavelengths for different types of meat samples, only a set of optimum wavelengths including five wavebands (567, 579, 595, 624, and 732 nm) were identified as universal feature wavelengths by a comprehensive comparison of three schemes, namely, variable fusion, data merging, and cross modeling. A simplified LDA model was then established based on these important wavelengths, yielding classification accuracies of 94.20 and 98.36% in the validation set for the intact meat and minced samples, respectively. The overall results showed that the integration of hyperspectral imaging and multivariate analyses has great potential for rapid and nondestructive differentiation of common red meat species.

Keywords: *hyperspectral imaging, red meat, variable selection, feature fusion.*

СОВМЕЩЕНИЕ МЕТОДОВ ГИПЕРСПЕКТРАЛЬНОГО ИЗОБРАЖЕНИЯ И ВЫБОРА ХАРАКТЕРНЫХ ДЛИН ВОЛН ДЛЯ БЫСТРОГО РАСПОЗНАВАНИЯ ВИДОВ КРАСНОГО МЯСА

D. Ding, K. Liang, B. Li, L. Liu, W. Wu, M. Shen*

УДК 543.42:637.5

Нанкинский сельскохозяйственный университет,
Нанкин, 210031, Китай; e-mail: dingdong@njau.edu.cn; mingxia@njau.edu.cn

(Поступила 28 мая 2019)

Исследована гиперспектральная система визуализации (400–800 нм) в сочетании с многомерным анализом для различения видов говядины, свинины и баранины на основе характерных длин волн интактных и измельченных образцов. Проведено сравнение характеристик классификационных моделей, построенных путем объединения линейного дискриминантного анализа (LDA), дискриминантного анализа с проекцией на латентные структуры (PLS-DA) или метода опорных векторов (SVM), с методами выбора переменных, такими, как алгоритм последовательного проецирования (SPA), анализ коэффициента регрессии (RCA) или метод реверсивных скачков (RF). Показано, что при идентификации видов сырого мяса линейный классификатор предпочтительнее нелинейного. Путем всестороннего сравнения трех схем, в том числе синтеза переменных, слияния данных и перекрестного моделирования, определен только один набор оптимальных длин волн, включающий в себя пять диапазонов (567, 579, 595, 624 и 732 нм) в качестве универсальных и характерных, вместо выбора различных наборов характеристических длин волн для образцов различных видов мяса. На основе выбранных длин волн создана упрощенная модель LDA, позволяющая получить точность классификации 94.20 и 98.36% в валидационном наборе образцов интактного мяса и фарша. Интегрирование гиперспектральной визуализации и многомерного анализа обладает большим потенциа-

лом в решении проблемы быстрой и неразрушающей дифференциации распространенных видов сырого мяса.

Ключевые слова: *гиперспектральное изображение, сырое мясо, выбор переменных, синтез характеристик.*

Introduction. Meat and meat products are important constituents of the human diet. However, with increasing meat consumption in recent years, meat adulteration has become a growing challenge for meat manufacturers and consumers [1, 2]. Traditional analytical methods, such as sensory evaluation and biochemical analysis, continue to be widely employed to address this problem [3]. However, sensory evaluation is subjective and inconsistent, and biochemical analysis techniques are typically laborious, invasive, and technically demanding [4]. Consequently, a rapid and cost-efficient method that facilitates the accurate identification of meat from different species at each level of the supply chain is desirable in the meat industry [5].

In recent years, spectroscopy and hyperspectral imaging (HSI) have been extensively studied as nondestructive techniques [6, 7], and have already been applied to meat species identification. Cozzolino et al. [8] achieved up to 80% accuracy by implementing a principal component analysis (PCA) and partial least squares regression (PLSR) model with a monochromator in the region of 400–2500 nm to discriminate between beef, lamb, pork, and chicken samples. Mamani–Linares et al. [9] achieved 95% accuracy by employing PCA and PLS modeling within the spectral wavelength range of 400–2500 nm to discriminate between minced beef, gravy beef, alpaca, and horse meat samples. Kamruzzaman et al. [10] recognized the intact and minced muscle of pork, beef, and lamb with an accuracy of 98.67% using a 900–1700 nm HSI system.

The long-wave range has been preferred for the identification of meat samples from different species in previous studies. However, operating a spectroscopy system in the near-infrared region significantly increases the overall cost owing to the expensive detectors used at wavelengths over 900 nm. Therefore, in stark contrast to previous studies, the HSI system in the range of 400–800 nm was adopted in this study. Furthermore, feature wavelength selection has been shown to be an extremely valuable tool for the design of a simple and cost-effective multi-spectral imaging system. However, feature wavelengths of intact and minced meat samples were different in this study owing to their spectral differences, making their industrial application inconvenient. Therefore, the main objectives of the present study were to (1) establish hyperspectral imaging in the spectral region of 400–800 nm as a classification tool for red meat from three different species, (2) identify feature wavelengths for meat species identification in the intact and minced samples, and establish an optimal classification model.

Experimental and methods. *Sample preparation.* Meat samples from the longissimus dorsi muscle of pork, beef, and mutton were purchased at 24 h post-mortem from a scale abattoir (Sushi Meat Products Co., Ltd. Huaian, Jiangsu, China), and transported to the laboratory under refrigeration. The meat samples were cut into 212 slices with a thickness of 2 cm, comprising pork (70), beef (71), and mutton (71) samples. Furthermore, a total of 185 samples of pork (59), beef (60), and mutton (66) from a different batch were collected and minced to a homogeneous paste using a food processor. The minced samples were stored in a cylindrical vessel (thickness, 2 cm; diameter, 5 cm) prior to data acquisition. The Kennard–Stone algorithm was used to select the calibration set, which amounted to two-thirds of all samples from each species. The remaining one-third of samples was selected as the validation set.

Hyperspectral imaging system. A HSI system operating in reflectance mode was used to acquire spectral images of the samples within the wavelength range of 400–800 nm. The system comprised a line-scan spectrograph (ImSpector V10E, Spectral Imaging Ltd., Finland), a CCD camera (IPX-2M30-LMC, Imperx, USA), two halogen lamps as an illumination unit (MHAB-150W-C, Moritex, Japan), a translation stage powered by a stepper motor (Isuzu Optics, Taiwan, China), and a computer with data acquisition software (Isuzu Optics, Taiwan, China). All samples were bloomed for approximately 20 min before hyperspectral image acquisition. The meat samples were then put on the translation stage at room temperature and individually transported into the camera field of view at a speed of 3.3 mm/s, as determined by optimizing the camera exposure time to prevent image distortion.

Hyperspectral image correction and spectral extraction. To reduce the dark current effects associated with the CCD camera, and the nonuniformity of illumination, the relative reflectance image (R) was calculated as follows:

$$R = (R_0 - B)/(W - B), \tag{1}$$

where R_0 is the raw hyperspectral image, W is the white reference image obtained by using a high-reflectance white calibration tile, and B is the dark reference image, which was acquired with the camera lens completely covered with an opaque cap.

After image correction, segmentation was performed with a threshold at the 550-nm band. This processing yielded a segmented image for each meat sample. The isolated lean meat was selected as the region of interest (ROI), and the adjoining fat and remaining background were ignored. A mean spectrum was then calculated by averaging the spectra of all pixels within each ROI. This procedure was repeated for all hyperspectral images originating from the three meat categories. Image correction, segmentation, and extraction of spectral information were performed using ENVI software (ITT Visual Information Solutions, USA).

Owing to the low signal-to-noise ratio at the head and tail ends of the spectrum, only the spectral range of 409–792 nm (540 bands) was used for further analysis. Notably, several preprocessing techniques, such as multiplicative scatter correction (MSC) and Savitzky–Golay smoothing, were implemented on the raw spectral data. However, as these techniques did not significantly improve the prediction accuracy, the raw spectral data were used in this study.

Feature wavelength selection and modeling. Hyperspectral data are characterized by their high dimensionality and collinearity, which consequently prevent the HSI system from being widely implemented in online industrial applications. Therefore, waveband selection is necessary to reduce the number of variables such that only a few with the most valuable information remain and are employed in effective discrimination models [11–13]. In this study, three variable selection methods, namely, the successive projections algorithm (SPA), regression coefficients analysis (RCA), and random frog (RF), were employed and evaluated. SPA is a characteristic variable forward selection algorithm that can eliminate the collinearity between numerous wavelength variables and has been widely implemented to perform wavelength selection for food quality detection [14, 15]. RCA is acknowledged as another efficient method to determine feature wavelengths [16, 3]. The regression coefficients from the PLSR model applied to the full spectra were calculated, and the variables with the highest absolute values of RC were selected as the most representative wavelengths. Finally, RF is a relatively new approach proposed by Li et al. [17] and is similar to the reversible jump Markov chain Monte Carlo algorithm. This method determines the selection probability of each variable through a few variable iterations, thereby optimizing the variable selection according to the magnitude of the probability [18].

The selected feature wavelengths of the calibration set were used as the inputs to the classification models, which were established using two typical linear classifiers and a nonlinear classifier, namely, linear discriminant analysis (LDA), partial least squares discriminant analysis (PLS-DA), and support vector machine (SVM), respectively. A tenfold cross-validation method was used to determine the number of latent variables (LVs) to avoid overfitting in the PLS-DA model. For the SVM classifier, the radial basis function kernel method was applied to simulate a nonlinear projection of the data in a higher dimension. The optimal values of the penalty parameter c and kernel parameter γ for the SVM classifiers were determined by grid search and four-fold cross-validation, respectively.

The performance of established models was evaluated in terms of the correct classification rate (CCR, %) in both the calibration set and the validation set, which was determined as follows:

$$\mathrm{CCR} = n/N \times 100, \tag{2}$$

where N is the total number of meat samples, and n is the number of correctly classified samples.

Feature wavelength fusion and optimization. The optimization of feature wavelengths was implemented to further reduce the number of feature wavelengths and identify the feature wavelengths applicable to both intact and minced meat samples without sacrificing the classification accuracy. For the former objective, stepwise regression (SR) was employed to eliminate selected feature wavelengths that did not significantly influence the classification accuracy. SR is the step-by-step iterative construction of a regression model that involves automatic variable selection [19]. This method is based on the procedure of sequentially introducing variables into the model one at a time. The number of variables retained in the final model was determined by the criterion assumed for inclusion and exclusion of variables from the model. In this study, the criteria employed for entry and elimination of feature variables were based on a significance level of 5%. Three schemes were implemented and compared to realize the latter objective. Scheme 1 was variable fusion, which entailed combining the feature wavelengths of intact and minced meat samples obtained using the same variable selection method. The SR algorithm then filters the combined feature wavelengths to obtain the optimized feature variables and establish a classification model. Scheme 2 was data merging, which entailed merging the original spectral data of the same species of intact and minced samples before processing the merged spectral data using samples partitioning, feature wavelength extraction, and SR-based reoptimization methods to obtain new feature wavelengths and build a classification model. Scheme 3 was cross modeling, which evaluated the feasibility of a non-fusion strategy. This involves feature wavelengths selected by SPA, RCA, and RF from the spectral data of intact and minced samples being refined directly with the SR method. This allows different sets of classification feature wavelengths to be obtained, each of which is applied to one type of meat sample (intact or minced samples). These feature wavelengths are used not only for classification modeling corresponding to the data set to which they belong, but also to establish the classification model of the data set for the other type of meat sample. Finally, the optimum set of feature wavelengths were identified by comparing the classification accuracy of each model in cross-modeling.

Among these three schemes, the optimal scheme was determined by performing a comprehensive comparison analysis, which analyzed and compared the number of feature variables and the discriminant accuracy of the simplified models. All multivariate statistical analysis operations were performed using MATLAB R2010b software (The MathWorks, Inc., Natick, MA, USA).

Results and discussion. *Spectral characteristics of meat samples.* The mean reflectance spectra extracted from the ROIs in the hyperspectral images of the meat samples are shown in Fig. 1. The spectra of the three red meat species were observed to have similar profiles, with the only significant difference being the magnitude of reflectance. The reflectance values for pork meat were clearly the highest, mainly owing to the pork samples being visibly lighter and brighter than the beef and mutton samples. Furthermore, two distinct absorption bands were observed at approximately 545 and 580 nm. This observation indicated a certain absorption effect shown by all three red meat species on the incident light that reduces the reflectance of these two bands, which are each related to both myoglobin and oxymyoglobin [9].

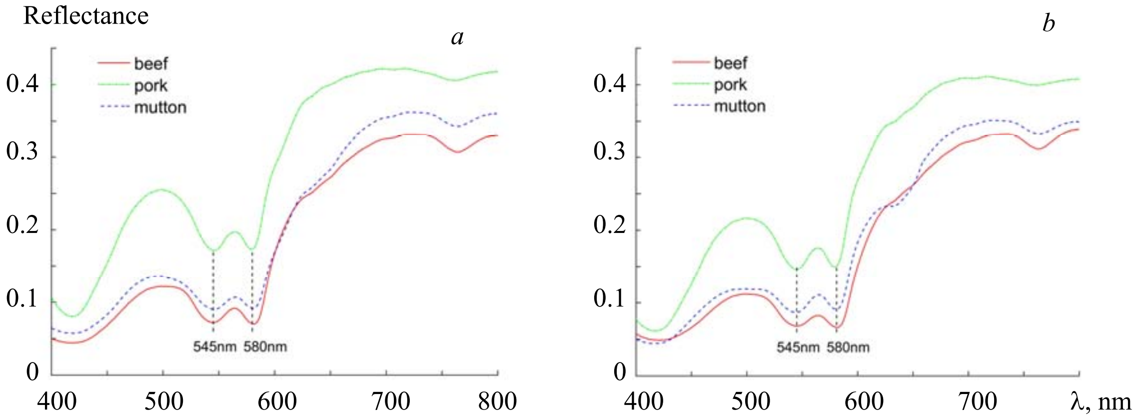

Fig. 1. Mean reflectance spectra for (a) intact and (b) minced meat samples.

Feature wavelength selection and classification models. SPA, RCA, and RF were conducted to extract the feature wavelengths from the full-wavelength spectral data. The number of feature wavelengths using SPA was determined using the principle of root mean square error (RMSE) minimization, with the RMSE results shown in Fig. 2. Increasing the number of feature variables above nine did not significantly reduce the respective RMSE values of 0.1717 and 0.0757 for intact and minced samples, respectively.

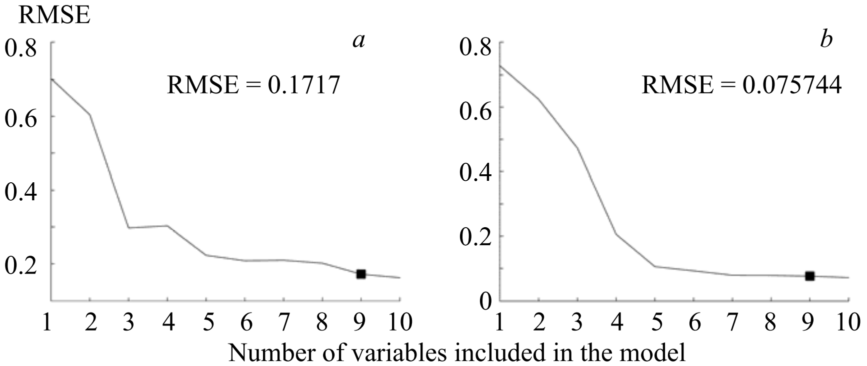

Fig. 2. RMSE plots of the number of selected variables for (a) intact and (b) minced samples by SPA.

RCA was implemented during PLS regression, with plots of the regression coefficient shown in Fig. 3. Nine and eleven wavelengths from the intact and minced samples, respectively, were determined to have significantly large absolute RC values and were selected as the feature wavelengths.

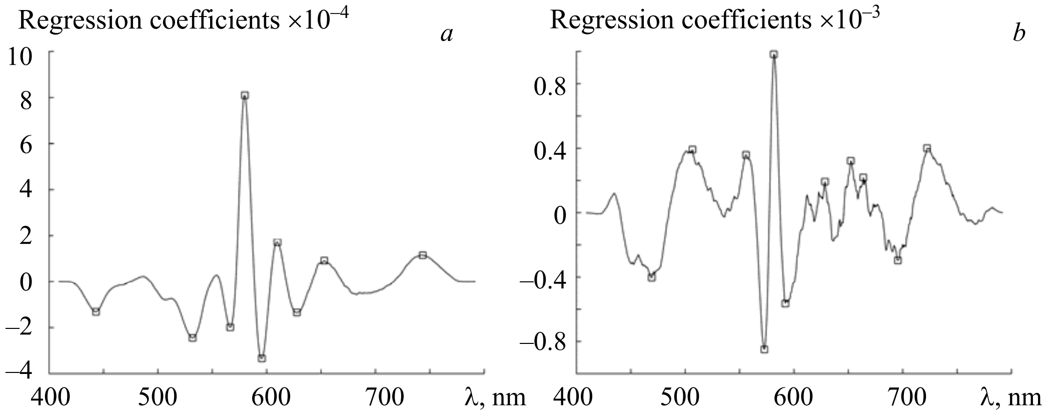

Fig. 3. Regression coefficients of each wavelength for (a) intact and (b) minced meat samples by RCA.

RF was conducted ten times in this study to offset the inherent instability of the algorithm, which caused the results to slightly vary. The average selection probability values are shown in Fig. 4. The threshold was set as 0.2 and 0.35 for intact and minced samples, respectively, and wavelengths with a selection probability higher than the threshold were selected as the feature wavelengths.

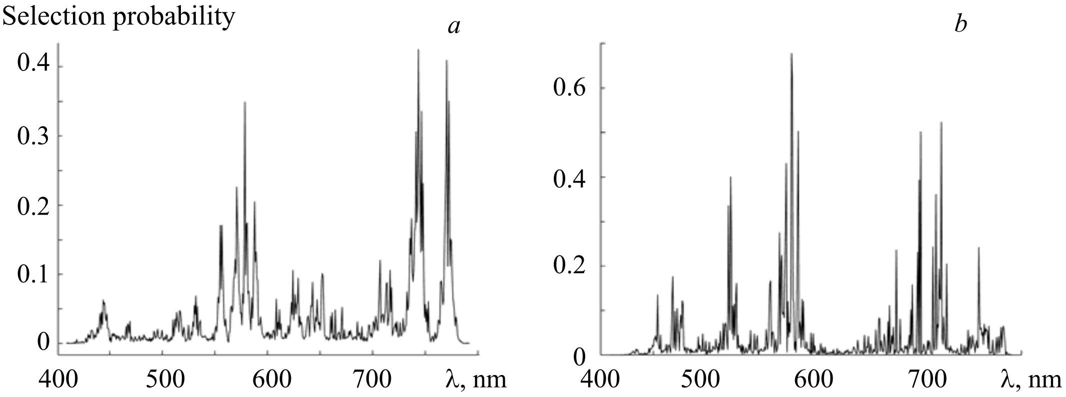

Fig. 4. Mean selection probability of each wavelength for (a) intact and (b) minced samples by RF.

Table 1 shows the detailed results of feature wavelengths obtained by SPA, RCA, and RF. Notably, all three variable selection methods significantly reduced the number of wavebands. Some feature wavelengths selected by different methods were similar to each other. RF also tended to yield feature wavelengths that were directly adjacent to each other, which could lead to data collinearity.

TABLE 1. Feature Wavelength Selection Results

Sample	Method	Feature wavelength, nm
Intact	SPA	551 567 579 595 608 624 652 699 732
	RCA	443 532 567 580 596 610 628 653 743
	RF	571 579 588 741 744 745 746 748 769 770 771 772 773
Minced	SPA	562 573 580 590 611 636 652 670 719
	RCA	469 507 556 573 581 592 629 652 664 696 722
	RF	521 524 576 581 582 587 588 703 704 719 724

Table 2 shows the performances of the three classifiers based on the selected feature wavelengths. As highlighted in Table 2, PLS-DA showed the best overall classification performance for intact meat samples compared with the other classifiers, achieving a minimum CCR of approximately 95%. Among the intact-meat PLS-DA models, the RF-PLS-DA model with 13 feature wavelengths showed the best performances with a CCR of 98.60 and 98.55% in the calibration and validation sets, respectively. The SPA-LDA model was also able to achieve an average CCR above 97% with only nine feature wavelengths. In contrast, applying the RCA-PLS-DA and SPA-PLS-DA models to the minced meat calibration and validation sets yielded a CCR of 100%. Notably, the average CCR for minced meat samples was significantly higher than that for the intact meat samples. This was primarily due to the classification data being derived from the average spectral features of the meat surface, and excluded all spatial image information. Therefore, no loss of information occurred for the minced samples during spectral extraction, whereas the ROI of the intact samples might contain intramuscular fat or fascia that would lead to incorrect information being introduced. Finally, among the three classifiers, PLS-DA and LDA performed significantly better than SVM, which suggested that linear classifiers were preferred for the classification of common red meat species.

TABLE 2. CCR (%) of the Calibration and Validation Sets Based on Feature Wavelengths

Sample	Method	nSV[a]	LDA		PLS-DA		SVM	
			Calibration	Validation	Calibration	Validation	Calibration	Validation
Intact	SPA	9	97.90	97.10	98.60	95.65	86.01	81.16
	RCA	9	92.31	91.30	95.10	95.65	88.11	84.06
	RF	13	87.41	73.91	**98.60**	**98.55**	90.91	91.30
Minced	SPA	9	99.19	100.00	**100.00**	**100.00**	89.52	85.25
	RCA	11	97.58	100.00	**100.00**	**100.00**	90.32	90.16
	RF	11	99.19	100.00	98.39	100.00	92.74	91.80

[a] Number of selected variables.

Feature wavelength fusion and optimization. Although the feature wavelength-based classification models yielded good results, the number of feature wavelengths remained undesirably large. Furthermore, the classification feature wavelengths for intact and minced meat differed, showing that further optimization was needed. As the linear classifiers yielded the highest recognition rates, only LDA and PLS-DA were used to establish classification models in the three schemes described above. Various combinations of the feature variable selection algorithms and classifiers were present in the optimization schemes, with the best results from each scheme listed in Table 3.

TABLE 3. Highest CCR (%) of Each Scheme at Optimized Wavelengths
for the Calibration and Validation Sets

Scheme	Method	nSV[a]	nLV[b]	Intact samples		Minced samples	
				Calibration	Validation	Calibration	Validation
1	SPA+SR+PLS-DA	11	9	98.60	98.55	100.00	100.00
2	RCA+SR+PLS-DA	9	9	97.90	97.10	99.19	100.00
3	SPA+SR+LDA	5		95.10	94.20	97.58	98.36

[a] Number of selected variables.
[b] Number of latent variables of PLS.

As indicated in Table 3, the fusion strategy adopted in schemes 1 and 2 did not achieve the expected results in terms of the number of feature wavelengths, although the classification accuracy of these two schemes was still high. After optimization, the number of universal feature wavelengths determined by schemes 1 and 2 was eleven and nine, respectively, which were not lower compared with Table 2. In contrast, scheme 3 achieved relatively satisfactory results, in which the SPA and SR-based LDA model achieved CCRs of 95.10 and 94.20% for the intact meat calibration and validation sets, and 97.58 and 98.36% for the minced meat calibration and validation sets, respectively. Although the classification accuracy of scheme 3 was slightly lower than that of the other schemes, five wavelengths (567, 579, 595, 624, and 732 nm) extracted from the spectral data of intact samples using a coupled SPA and SR algorithm were identified as optimal feature wavelengths, which accounted for only 0.93% of the original variables, and, more importantly, were effective for the classification of minced pork, beef, and mutton. In contrast, the feature wavelengths extracted from minced samples were not suitable for species classification of intact samples.

Conclusions. In this study, hyperspectral imaging performed in the spectral range of 400–800 nm was combined with various chemometric methods for the discrimination of beef, pork, and mutton. Three classifiers (LDA, PLS-DA, and SVM) coupled with three variable selection methods (SPA, RCA, and RF) were used to establish the classification models. Comparison of the accuracies showed that linear classifiers were the preferred choice for identifying meat species compared with nonlinear classifiers, and that the average classification accuracy for the minced meat samples was significantly higher than that for the intact meat samples. Furthermore, the feasibility of determining the feature wavelengths for identification of meat species applicable to both intact and minced samples was studied using three schemes, namely, variable fusion, data merging, and cross modeling. A comprehensive comparison of the three schemes showed that the fusion strategy could not further reduce the number of feature variables, while the feature wavelengths extracted from the data of intact samples could be effectively employed to classify minced meat samples. Five wavelengths (567, 579, 595, 624, and 732 nm) were identified as universal feature wavelengths and a simplified LDA model was established to yield overall satisfactory classification performance, with an average CCR exceeding 94%. The results obtained in this study indicated that the development of a low-cost multi-spectral imaging system to identify common meat species was feasible.

Acknowledgments. This study is supported by the Agricultural Machinery Administration Research Fund of Jiangsu Province (gxz14004).

REFERENCES

1. Y. Kumar, S. C. Karne, *Trends Food Sci. Technol.*, **62**, 59–67 (2017).
2. M. Kamruzzaman, Y. Makino, S. Oshita, *Anal. Chim. Acta*, **853**, 19–29 (2015).
3. M. Kamruzzaman, D. W. Sun, G. El Masry, P. Allen, *Talanta*, **103**, N 2, 130–136 (2013).
4. C. H. Feng, Y. Makino, S. Oshita, J. F. G. Martin, *Food Control*, **84**, 165–176 (2018).
5. D. J. Troy, K. S. Ojha, J. P. Kerry, B. K. Tiwari, *Meat Sci.*, **120**, 2–9 (2016).
6. J. H. Cheng, B. Nicolai, D. W. Sun, *Meat Sci.*, **123**, 182–191 (2017).
7. M. M. Reis, R. V. Beers, M. Al-Sarayreh, P. Shorten, W. Q. Yan, W. Sayers, R. Klette, C. Craigie, *Meat Sci.*, **144**, 100–109 (2018).
8. D. Cozzolino, I. Murray, *LWT-Food Sci. Technol.*, **37**, N 4, 447–452 (2004).
9. L. W. Mamani-Linares, C. Gallo, D. Alomar, *Meat Sci.*, **90**, N 2, 378–385 (2012).
10. M. Kamruzzaman, D. Barbin, G. El Masry, D.-W. Sun, P. Allen, *Innov. Food Sci. Emerg. Technol.*, **16**, N 39, 316–325 (2012).
11. J. Qin, K. Chao, M. S. Kim, R. Lu, T. F. Burks, *J. Food Eng.*, **118**, N 2, 157–171 (2013).
12. D. Liu, D. W. Sun, X. A. Zeng, *Food Bioprocess Technol.*, **7**, N 2, 307–323 (2014).
13. H. Pu, M. Kamruzzaman, D. W. Sun, *Trends Food Sci. Technol.*, **45**, N 1, 86–104 (2015).
14. Z. Xiong, D. W. Sun, H. Pu, Zh. Zhu, M. Luo, *LWT-Food Sci. Technol.*, **60**, N 2, 649–655 (2015).
15. X. Wu, X. Song, Z. Qiu, Y. He, *Meat Sci.*, **113**, 92–96 (2016).
16. M. Kamruzzaman, Y. Makino, S. Oshita, *Food Chem.*, **196**, N 3, 1084–1091 (2016).
17. H. D. Li, Q. S. Xu, Y. Z. Liang, *Anal. Chim. Acta*, **740**, 20–26 (2012).
18. M. H. Hu, Q. L. Dong, B. L. Liu, U. L. Opara, L. Chen, *Postharvest Biol. Technol.*, **106**, 1–10 (2015).
19. S. R. Jammalamadaka, *Am. Stat.*, **57**, N 1, 67–69 (2012).