TRENDS IN SCIENCES 2026; 23(11): 13765
https://doi.org/10.48048/tis.2026.13765 RESEARCH ARTICLE

Partial Least Squares Regression Combined with Various Variable Selection Techniques to Determine Pork Adulteration in Ground Beef and Lamb Using Visible Near Infrared Hyperspectral Imaging

Rudiati Evi Masithoh[1], Muhammad Fahri Reza Pahlawan[2], Juntae Kim[3], Reza Adhitama Putra Hernanda[4], Hanim Zuhrotul Amanah[1], Hoonsoo Lee [4], Widodo Widodo[5] and Byoung-Kwan Cho[2,3,*]

[1]Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia
[2]Department of Smart Agriculture Systems, College of Agricultural and Life Science, Chungnam National University, Daejeon 34134, Republic of Korea
[3]Department of Biosystems Machinery Engineering, College of Agricultural and Life Science, Chungnam National University, Daejeon 34134, Republic of Korea
[4]Department of Biosystems Engineering, College of Agriculture, Life & Environment Sciences, Chungbuk National University, Cheongju 28644, Republic of Korea
[5]Faculty of Animal Science, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia

([*]Corresponding author's e-mail: chobk@cnu.ac.kr)

Received: 10 March 2026, Revised: 2 May 2026, Accepted: 9 May 2026, Published: 10 June 2026

Abstract

Meat adulteration raises special concerns regarding quality, safety, and socio-cultural values. This study integrated visible-near-infrared hyperspectral imaging (Vis-NIR HSI) at 397.4 - 1,003.8 nm and partial least squares regression (PLSR) to determine the concentration of pork added in ground beef and lamb. Moreover, this study compared various wavelength-selection methods, i.e., Competitive Adaptive Reweighted Sampling (CARS), Monte Carlo Uninformative Variables Elimination (MCUVE), Successive Projections Algorithm (SPA), and Variable Importance in Projection (VIP). Pork at concentrations of 2%, 5%, 10%, 20%, 30%, 40%, and 50% by weight was added to ground beef and lamb, and the 34 samples were scanned using the Vis-NIR HSI. PLSR models were developed using full-wavelength and selected-wavelength data obtained from CARS, MCUVE, SPA, and VIP. Among feature-selection approaches, the SPA produced the most accurate models with coefficient determination (R^2) values of 0.974 for calibration, 0.952 for cross-validation, and 0.955 for prediction for the Beef + Pork model, as well as yielded R^2 values of 0.922 for calibration, 0.847 for cross-validation, and 0.803 for prediction for the Lamb + Pork model. Chemical maps visualized the distribution of pork adulterants, with clearer spatial patterns in beef than in lamb. The results confirm the use of Vis-NIR HSI for rapid, non-destructive detection of meat adulteration.

Keywords: Adulteration, Hyperspectral imaging, Meat, Partial least square regression, Wavelength selection

Introduction

Meat fraud, often driven by economic motives, involves substituting high-quality meat with cheaper or unauthorized substances, also known as fraudulent adulteration, and mislabeling [1]. Adulteration is regarded as an activity that is not only economic fraud but also violates health and religious or ethical standards. Adulteration is an intentional replacement of an ingredient that is specifically motivated by adding either any poisonous or deleterious substance, compared to contamination, which is considered an unintentional

activity and technically unavoidable [2]. Economically motivated adulteration (EMA) occurs when a substance in a product is intentionally substituted or added for the purpose of increasing the apparent value of the product or reducing the cost of its production [3]. Regarding health, food adulteration can cause acute or chronic illnesses, food poisoning, or allergies [4]. Meat adulteration practices have been repeatedly reported worldwide; one of them was summarized by Robsonet *et al.* [5], meat fraud in Chinese [6], or horsemeat sold as beef in Mexico [7]. Therefore, the authentication of meat products is important to protect consumers from food fraud.

Meat authentication methods have been conducted using sensory analysis, polymerase chain reaction (PCR), computer vision systems (CVS), and infrared spectroscopy, which possess advantages and limitations. Sensory parameters of meat include visual appearance, texture, flavor, and aroma [8], which were measured using sensory analysis reported by [9] and [10]. Nonetheless, sensory analysis relies on human perception, which can be subjective and influenced by several factors such as personal opinions, experiences, and competencies [10]. Moreover, low-level adulteration that closely mimics the sensory properties of the original meat may not be detectable by panelists. PCR is a DNA-based method that is preferred for meat adulteration detection because DNA remains stable under harsh thermal and chemical conditions, unlike proteins, which can denature during food processing [11]. PCR was used for identifying meat adulteration in ostrich, cat, goose, duck, chicken, horse, dog, and sheep [11], as well as beef and mutton [12]. Although accurate, PCR requires expensive equipment, long analysis times, and skilled analysts, which limit its application for routine analysis. Another technique, the computer vision system (CVS), which is based on images, has been applied to detect meat adulterants [13], [14]. CVS provides low-cost, rapid, and easy methods for detecting meat adulteration. However, conventional CVS is limited to capturing 2-dimensional spatial data and relies on visible light, which prevents it from detecting chemical constituents invisible to the human eye.

On the other hand, spectroscopic methods operating in the visible and near-infrared (Vis-NIR), near-infrared (NIR), and mid-infrared (MIR) have been applied to detect meat adulteration. Vis-NIR spectroscopy working at 350 - 1,000 nm and 1,000 - 2,500 nm was used to predict pork adulteration in beef and mutton [15], to predict pork, beef heart, and beef tallow adulteration in minced beef [16], or to estimate plant and animal protein adulteration in minced beef and pork [17]. Another spectroscopy method at longer infrared wavelengths (2,500 - 4,000 nm) was also used to detect pork, cattle, chicken, and foal meat adulteration in lamb and beef [18], to detect pork adulteration in beef, mutton, lamb, pork, camel, chicken, and veal [19], or to detect pork and duck meat adulteration in minced beef [20]. Although accurate, traditional spectroscopy is limited in that it can analyze only a small area of the sample, about 1 - 2 cm in diameter, which may not accurately represent the whole sample. Thus, it requires multiple scans to capture the entire sample. To address this limitation, hyperspectral imaging (HSI) spectroscopy enables comprehensive analysis of meat surfaces by integrating both spectral and spatial data [21]. HSI spectroscopy has been used to detect the presence of adulteration and predict chemical composition in food [22,23].

HSI spectroscopy in the NIR region has been applied in meat authentication [24,25] employed HSI-NIR (at 895 - 2,504 nm) and partial least squares regression (PLSR) to predict adulteration of pork in minced beef. [26] predicted pork adulteration in minced lamb using HSI NIR spectroscopy at 910 - 1,700 nm using PLSR and a multiple linear regression (MLR). These long infrared wavelengths are valuable for identifying fat, protein, and water in various meats [27]. Although HSI NIR spectroscopy in this range has shown success for meat adulteration detection, relatively few studies have explored its potential in the visible to near-infrared (Vis–NIR) region of 350 - 1,000 nm. The visible region (400 - 700 nm) contains absorption features of muscle pigments such as myoglobin (Mb) and deoxymyoglobin (DMb) [28], while the NIR region (700 - 1,000 nm) captures overtone bands related to water, fat, and collagen, which can be important for meat adulteration detection [27]. Therefore, this study aims to investigate the potency of Vis-NIR HSI for detecting pork adulteration in ground beef and lamb.

Moreover, the performance of PLSR models strongly depends on the selection of wavelength [29]. Previous research on meat adulteration used PLSR models commonly developed using the full spectra,

which contain strongly collinear and non-useful bands [30]. Thus, selecting a set of variables that best correlates with the analyte of interest is important for producing a simple, economical, and efficient spectroscopic system [31]. This study uses various variable selection methods to identify the best PLSR model for predicting the concentration of pork in ground beef and lamb. Additionally, the best-on-the-best PLSR model with selected variables generates visualizations of meat samples, providing qualitative assessment methods in the meat industry.

Materials and methods

Sample preparation

Ground beef, lamb, and pork samples were purchased from a meat market supplier in Daejeon, South Korea. Before sample processing, all meat samples were refrigerated at 10 °C to maintain their freshness. The meat sample was placed in 90×15 mm^2 petri dishes for spectral acquisition. Ground pork at concentrations of 2% (w/w), 5% (w/w), 10% (w/w), 20% (w/w), 30% (w/w), 40%, and 50% (w/w) was added to ground beef and lamb. Each sample weighed 70 g and was mixed manually and thoroughly. In total. Thirty-six samples were prepared in this study with 2 replications in each concentration. Images of samples are displayed in **Figure 1**, which shows that the higher the concentration of pork, the paler the color of beef and lamb. Color intensity in meat is caused by myoglobin concentration, which is high in beef, intermediate in lamb, and low in pork [32]. The samples were processed (grinding and mixing) under air-conditioned room at approximately 23 - 25 °C and finally measured in a closed Vis-NIR system room at approximately 25 - 28 °C.

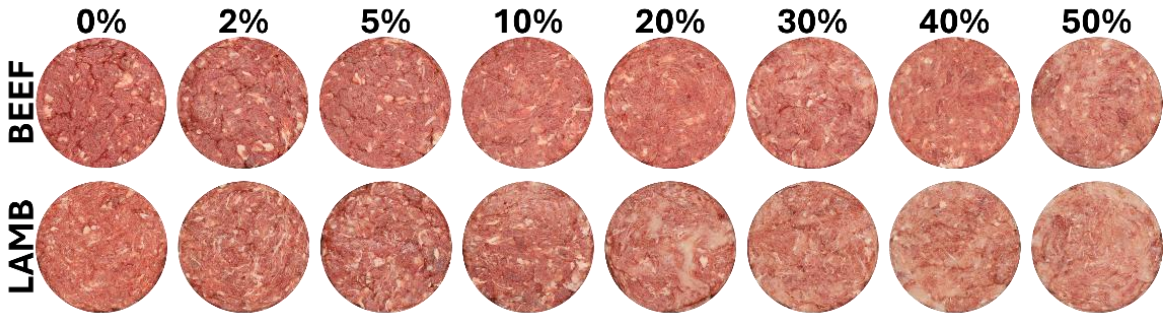


Figure 1 Color images of ground beef and lamb added with different pork concentration levels: 0% (w/w), 2% (w/w), 5% (w/w), 10% (w/w), 20% (w/w), 30% (w/w), 40% (w/w) and 50% (w/w). Captured with DSLR camera (Canon EOS Rebel T5i, Japan).

Hyperspectral data acquisitions

The workflow for detecting pork adulteration in minced beef and lamb using Vis-NIR HSI is illustrated in **Figure 2**. The visible near-infrared line-scan hyperspectral (Vis-NIR HSI) system in this research incorporated an imaging spectrograph (Headwall Photonics, 1002A-00751), an electron-multiplying charge-coupled device (EMCCD) camera (Luca DL-604 M-OEM, ANDOR Technology, MA, USA), and 6 tungsten-halogen lamps (100 W, 12 V, LS-F100HS) as the illumination source. This system recorded 128 spectral variables within the wavelength range of 397.4 - 1,003.8 nm with 4.7 nm spectral resolution (hereafter referred to as 400 - 1,000 nm). To obtain spectral and spatial data, samples were scanned with 35 cm sample-to-camera distance on a mobile platform driven by a DC motor at a speed of 3.48 mm/s with 20 ms exposure time, producing 623 scans per sample. The scanning resulted in hyperspectral images in a 3D hypercube format consisting of 2 spatial dimensions (x-axes and y-axes) and one dimension for spectral wavelength (λ), making the final dimensions of 502 (x) × 623 (y) × 128 (λ).

To obtain accurate hyperspectral data, the sensor was calibrated to account for variations in illumination and sensor response using dark and white references. The white reference image representing 100% reflectance was obtained using a white Teflon tile. In comparison, the dark reference image representing 0%

reflectance was obtained by closing the camera lens and turning off the light source. After sensor calibration, hyperspectral image acquisition was performed by scanning every 2 meat plates with random order on the HIS's moving platform.

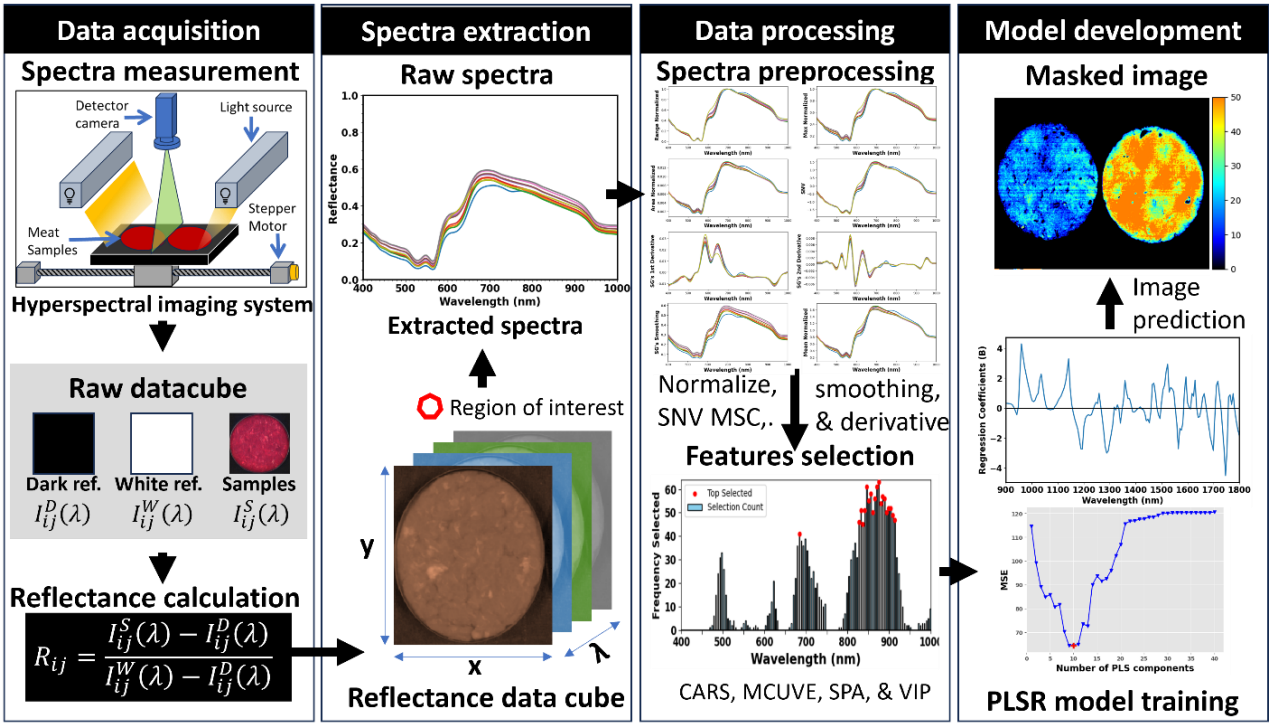


Figure 2 Workflow of pork adulteration detection in beef and lamb from Hyperspectral data acquisition to model development using Partial Least Square Regression with variable selection techniques. The processing steps include data acquisition, spectra extraction, data processing, and model development.

Hyperspectral data processing

After hyperspectral image acquisition, reflectance spectra were derived from the 3D hyperspectral cubes. Regions of interest (ROI) were manually outlined to extract spectral data for analysis. To distinguish sample spectra from background noise, a threshold wavelength of 680 nm was selected. This wavelength was selected because of 2 reasons: First, because it has the highest intensity of white reference, which indicates a high signal-to-noise ratio and highest sensitivity; and second, because it has the most distinct reflectance values between meat and background (**Figure S1**). ROI was defined via polygonal masking on x-y spatial images, with a reflectance threshold of 0.2 applied to exclude background interference. Pixel spectra within the ROI were averaged to yield representative spectra, resulting in 10 spectra per plate.

Partial Least Squares Regression (PLSR) analysis

Partial least squares regression (PLSR) was utilized to develop calibration models for the quantification of pork adulteration in beef and lamb, following the scikit-learn Python library [33]. Two models were developed: The pork adulterated beef (Beef + Pork) model and the pork adulterated lamb (Lamb + Pork) model. For each model, 180 spectra were used. The Beef + Pork model incorporated only spectra from pure beef and adulterated beef samples. Conversely, the Lamb + Pork model utilized only spectra from pure lamb and adulterated lamb samples. The full dataset was divided into a calibration set (70%, 156 spectra) and a prediction set (30%, 54 spectra). By changing the random split, 6 calibration-prediction datasets were generated. The calibration set was used for model development, while the prediction set was used to evaluate the model's predictive capability. Ten-fold

cross-validation was performed to assess the performance of the calibration model.

The PLSR model was constructed using the reflectance spectra as the independent variable (X) and the pork adulteration percentage as the dependent variable (Y), with up to 40 latent variables. To optimize model performance, several analyses were conducted, including spectral preprocessing techniques, variable selection methods, and determination of the optimal number of variables.

Spectral preprocessing techniques

To identify the best PLSR model, nine preprocessing methods were compared alongside the raw spectra. These methods included multiple scatter correction (MSC) and 5 normalization-based methods, such as standard normal variate (SNV), mean (MeN), range (RN), max (MxN), and area (AN) normalization. In addition, Savitzky-Golay smoothing (window length = 11, polynomial order = 2) and derivative (window length = 11, polynomial orders = 2 and 3, derivative = 1 and 2) were applied using the SciPy library [34]. The PLSR model was built using the full wavelength range of the raw and the 9 preprocessed spectra. The Friedman and Nemenyi post-hoc test was performed on the cross-validation performances, and a box plot was used to visualize the differences across preprocessing methods. The best preprocessing method was selected and used for further analysis.

Determination of variable selection method and number of variables

Variable selection was performed on the best-preprocessed spectra, yielding the best PLSR performance. Variable selections were used to remove the unrelated variables and focus only on variables that have a strong relationship with the target variables [35]. Applying variable selection after PLSR improves model performance and ensures consistency of selected variables, while also enhancing interpretability and understanding of the studied system [36]. Four methods were applied to select 5% to 95% (by 5% increments) of variables from the initial 128 wavelength variables, as shown in **Table 1**. The methods include Competitive Adaptive Reweighted Sampling (CARS), Monte Carlo Uninformative Variables Elimination (MCUVE), Successive Projections Algorithm (SPA), and Variable Importance in Projection (VIP). The detailed description of the variable selection methods used can be found in Talebi *et al.* [37]. All the variable selection methods were applied using the Auswahl library [38] in Python.

Some variable selection methods, such as CARS and MCUVE, exhibit low reproducibility due to the randomization step, yielding different results across runs [39]. To enhance the stability of variable selection, each method was repeated 100 times, and the number of variables was determined by their selection frequency. Three runs were performed for each calibration-validation dataset. The cross-validation performances, obtained from the 3 runs across the 6 calibration-validation datasets for each of the 19% of retained variables (**Table 1**), were then used to perform the Friedman and Nemenyi tests. The optimal model was defined as the one with the lowest percentage of retained variables that were not significantly different from the model with the highest performance. The models with the optimal percentage of retained variables obtained from each variable selection method were compared to identify the best-performing method.

Table 1 Percentage and number of selected variables from the initial 128 variables.

No	Percentage of retained variables (%)	Number of selected variables
1	5	6
2	10	12
3	15	19
4	20	25
5	25	32
6	30	38

No	Percentage of retained variables (%)	Number of selected variables
7	35	44
8	40	51
9	45	57
10	50	64
11	55	70
12	60	76
13	65	83
14	70	89
15	75	96
16	80	102
17	85	108
18	90	115
19	95	121
20	100	128

Masked image generation

The models with the optimal percentage of retained variables from each feature selection method were used to generate masked images from the hyperspectral data. The masked images were created by predicting each pixel in the hyperspectral image using the best-performing model. The image generation process comprised pixel-wise spectral preprocessing, selection of retained variables, model prediction, and color masking based on the predicted results. Image prediction was performed for all feature selection methods across all mixture concentrations. The masked images can be used to visualize the distribution and severity of pork adulteration in ground beef and lamb based on color differences.

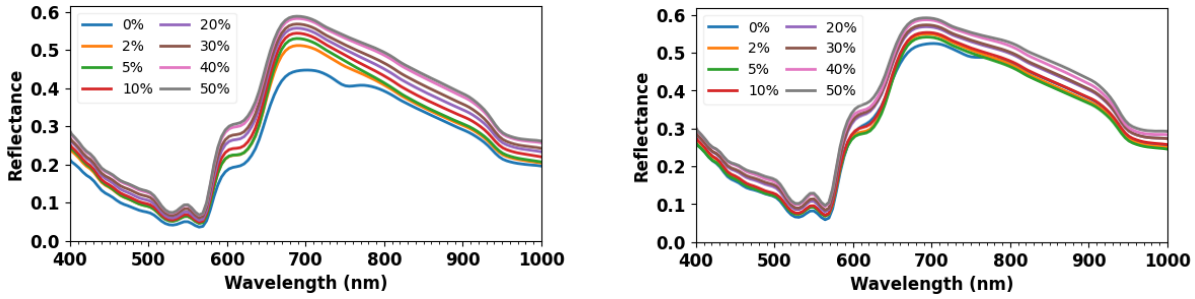
Statistical evaluation

Statistical metrics such as the coefficient of determination (R^2) and root mean square error (RMSE) were used to compare the model performance. The ideal model should have a high R^2 (Max 1.00) and a low RMSE. Statistical evaluation of significance differences was conducted using the Friedman and Nemenyi post-hoc analysis, applied across the 6 calibration-validation datasets as the replications. The Friedman test is a nonparametric method used to test differences between treatments across repeated measurements [40]. In this study, data splitting served as the replication or block for the Friedman test. When the Friedman test's p-value is < 0.05, the Nemenyi post-hoc test is used to identify the differences between treatments [41]. Friedman chi-squares were computed using the SciPy library [34], and the Nemenyi post-hoc test was performed using the Scikit-posthocs library [42].

Results and discussion

Spectra profile

Figure 3 shows the full reflectance-averaged spectra of the beef and lamb mixture with pork at various concentrations. Both figures have similar patterns to which intensity shifts as pork concentration increases, especially near 550 - 600 and 700 - 900 nm, suggesting that pigment and water-related absorptions respond to compositional change [27]. Likewise, the spectra pattern of Beef + Pork and Lamb + Pork in the shortwave near infrared region at 1,000 - 2,500 nm is also similar Masithoh *et al.* [24]. These phenomena justify the use of chemometrics for pure or adulterated meat, as they are difficult to differentiate visually.



(a) (b)

Figure 3 Averaged spectra of the a) Beef + Pork and b) Lamb + Pork mixture with pork in various concentrations ranging from 0% to 50% in Vis-NIR region (400 - 1,000 nm).

PLSR model using full wavelength variables

Tables 2 and **3** show the results of the PLSR models for Beef + Pork and Lamb + Pork, developed using different spectral preprocessing methods, including original non-preprocessed spectra, normalization, scatter correction, smoothing, and derivatives. Both tables demonstrated high predictive performance for pork adulteration in ground beef and lamb. In general, the PLSR models yielded high coefficients of determination (R^2) in calibration (R^2C = 0.982 - 0.979), cross-validation (R^2CV = 0.785 - 0.951), and prediction (R^2P = 0.796 - 0.954) for Beef + Pork. In addition, the PLSR models of Lamb + Pork resulted in R^2C of 0.893 - 0.954 for calibration, R^2CV of 0.850 - 0.888 for cross-validation, and R^2P = 0.806 - 0.859 for prediction. The results confirmed the suitability of the HSI in the VIS-NIR range (400 - 1,000 nm) for developing PLSR models for pork adulteration detection in ground beef and lamb. These results are better compared to the PLRS model for detection of intramuscular fat in minced lamb meat reported by Kazemi *et al.* [43] using VIS-NIR HSI, which contained the Ultraviolet region at 200 - 1,100 nm.

Among the preprocessing techniques, Maximum Normalization (MxN), Multiplicative Scatter Correction (MSC), and Savitzky-Golay First Derivatives (SGD1) in Beef + Pork model, and Savitzky-Golay Smoothing (SGS) in Lamb + Pork model provided relatively lower errors (RMSE) compared to the original spectra. These results indicate that appropriate preprocessing enhances the quality of spectral information by reducing scattering effects caused by baseline shifts and nonlinearity, thereby improving the stability of PLSR calibration models [44].

Table 2 Results of PLRS model for detecting percentage of pork in beef (Beef + Pork model) using various preprocessing techniques.

Preprocessing	R2C	RMSEC (%)	R2CV	RMSECV (%)	R2P	RMSEP (%)
Original	0.974	2.784	0.950	3.929	0.954	3.740
MeN	0.820	4.947	0.785	6.180	0.796	5.851
RN	0.966	3.134	0.949	3.969	0.946	4.051
MxN	0.974	2.717	0.951	3.901	0.949	3.951
AN	0.820	4.947	0.785	6.180	0.796	5.851
SNV	0.973	2.793	0.951	3.887	0.948	3.959
MSC	0.975	2.672	0.951	3.892	0.949	3.923
SGS	0.979	2.497	0.953	3.829	0.952	3.824
SGD1	0.973	2.833	0.948	3.996	0.952	3.866
SGD2	0.961	3.421	0.936	4.433	0.940	4.287

Note: MeN = mean normalization; RN = range normalization; MxN = maximum normalization; AN = average normalization; SNV = standard normal variance; MSC = multiplicative scatter correction; SGS = Savitzky Golay smoothing; SGD1 = Savitzky Golay first derivative; SGD2 = Savitzky Golay second derivative.

Table 3 Results of PLRS model for detecting percentage of pork in lamb (Lamb + Pork model) using various preprocessing techniques.

Preprocessing	R2C	RMSEC (%)	R2CV	RMSECV (%)	R2P	RMSEP (%)
Original	0.953	3.755	0.879	6.105	0.824	7.084
MeN	0.937	4.382	0.873	6.259	0.822	7.029
RN	0.943	4.001	0.873	6.255	0.810	7.554
MxN	0.919	4.882	0.869	6.364	0.806	7.556
AN	0.937	4.382	0.873	6.259	0.822	7.029
SNV	0.949	3.936	0.880	6.079	0.856	6.478
MSC	0.942	4.212	0.877	6.153	0.839	6.801
SGS	0.954	3.717	0.888	5.871	0.854	6.536
SGD1	0.954	3.765	0.873	6.260	0.859	6.565
SGD2	0.893	5.726	0.850	6.813	0.850	6.788

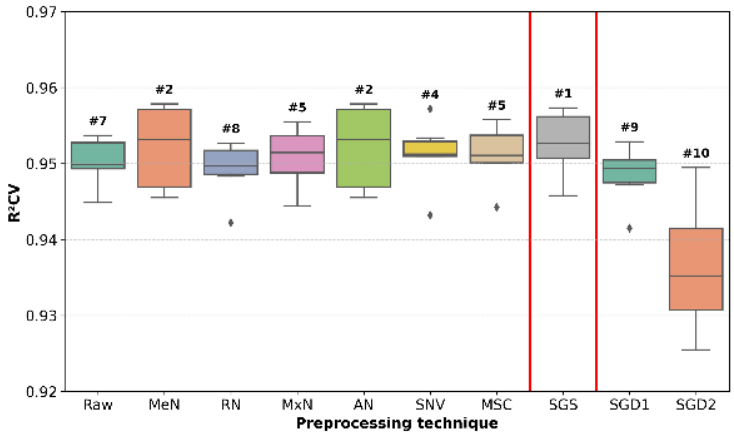
Note: MeN = mean normalization; RN = range normalization; MxN = maximum normalization; AN = average normalization; SNV = standard normal variance; MSC = multiplicative scatter correction; SGS = Savitzky Golay smoothing; SGD1 = Savitzky Golay first derivative; SGD2 = Savitzky Golay second derivative.

In the first stage of analysis, different preprocessing methods were compared in **Tables 2** and **3**. The comparison was focused on cross-validation (CV) results, rather than calibration (C) or prediction (P). Calibration used the same data on which the model was trained, leading to overly optimistic estimation while failing to account for overfitting, especially when reference data contains noise, and was unable to generalize to new and unseen data [45,46]. Moreover, external prediction results may vary depending on sample partitioning and calibration methods. Thus, selecting the best preprocessing method based on CV generalizes the model based on an independent dataset that it has not seen during the training process [47]. However, CV may still produce overestimated performance; thus, for model robustness and practical applicability, it ultimately requires external prediction [48].

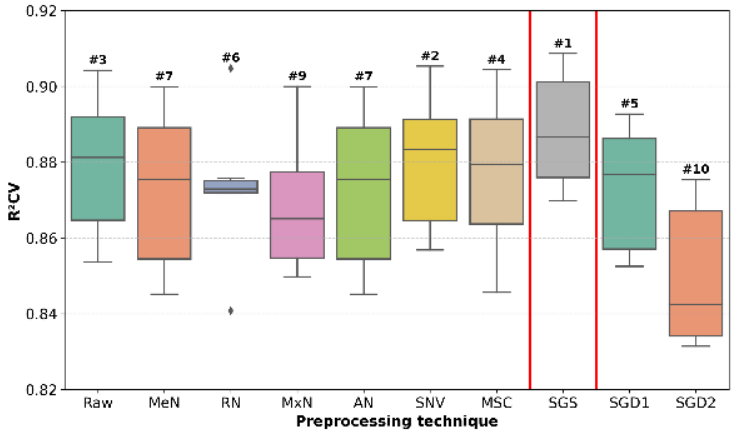
The cross-validated R^2 (R^2CV) comparison for each preprocessing technique is shown in **Figure 4**. The Friedman test yielded a p-value < 0.05 for both the Beef + Pork and Lamb + Pork models, indicating that the preprocessing method affects model performance. The boxplot illustrates the distribution of R^2CV across replications, highlighting the variability and the outliers [49]. The best preprocessing techniques were chosen based on 3 criteria: The highest median (the horizontal line in the middle of the box), the highest average value (the lowest rank), and a narrow spread (the box height). In the Beef + Pork models (**Figure 4(a)**) and the Lamb + Pork models (**Figure 4(b)**), the Savitzky-Golay smoothing (SGS) showed the highest predictive performances. On the other hand, the Savitzky-Golay second derivative (SGD2) method consistently underperformed across both datasets (Rank #10), indicating that it was not suitable for this dataset. However, the ranks from #2 to #9 vary notably between the Beef + Pork and Lamb + Pork datasets, suggesting that preprocessing methods can affect each sample differently. This is because there is no rule about whether to use or avoid a certain preprocessing method [36,50]. Therefore, comparing the model's performance across preprocessing methods is necessary.

Notably, the average model performance of Beef + Pork models was higher than that of the Lamb + Pork models. A similar trend, in which the Beef + Pork model was superior to the Lamb + Pork models, was observed in our previous study to detect the level of pork adulteration in beef and lamb using Vis-NIR spectroscopy [15] and SWIR hyperspectral imaging [24]. Although different samples and instruments were used, consistent results were obtained, suggesting that

classification between pork and lamb is more challenging. This may be due to the similarity in unsaturated fatty acids; lamb and pork contain relatively higher proportions of unsaturated fatty acids compared to beef [51]. Additionally, the similarity between lamb and pork may be because of the muscle fiber composition. Lamb and pork are usually slaughtered at a young age, which causes them to have more Type IIB fiber (fast-twitch glycolytic) compared to beef slaughtered at old age [52,53]. Moreover, similar body size may also contribute to similar muscle composition in lamb and pork, compared to the larger body size of beef [54].



(a)



(b)

Figure 4 PLSR Coefficient of determination of cross validation (R^2CV) comparison with various preprocessing method for (a) Beef+Pork and (b) Lamb+Pork (Note: the numbers with hashtag (#) show the average value rank, and the red rectangle shows the best selected method).

Feature selection

After determining the optimal preprocessing method based on CV performance (**Tables 2** and **3**), the analysis continued by evaluating the variable selection methods applied to the PLSR models. Although PLSR using spectra with the full spectral range (**Tables 2** and **3**) demonstrated strong predictive ability, variable selection remained important for further analysis. PLS with a univariate response may become unstable, and irrelevant variables can increase prediction variability in the test set [55]. Therefore, variable selection by reducing the number of wavelengths leads to a decrease in the size of variables and, consequently, the acquisition time for real-world applications [56], although there is no best variable selection method due to an interaction between the process and the properties of the data [36]. The ones yielded in this study are the best for these particular cases in the detection of pork,

beef, and lamb using the specific meat and HSI at 397.4 - 1,003.8 nm. Four methods were compared in this study, including Competitive Adaptive Reweighted Sampling (CARS), Monte Carlo Uninformative Variables Elimination (MCUVE), Successive Projections Algorithm (SPA), and Variable Importance in Projection (VIP).

Figure 5 shows the average R^2CV for the PLSR model with variable retention percentages from 5% to 95% in 5% increments. Based on the Friedman test, which yielded p-values < 0.05, the model performances were significantly affected by the number of variables for both Beef + Pork and Lamb + Pork. The lowest performance was observed with the lowest percentage of retained variables, and as the rate increased, R^2CV also increased until it reached its highest value, after which it decreased slightly. The highest R^2CV for each variable selection method is marked with the white rectangle.

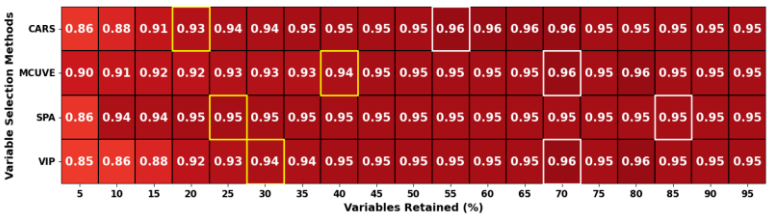
The highest R^2CV for Beef + Pork is 0.96 for CARS, MCUVE, and VIP, while SPA was slightly lower at 0.95. The corresponding percentage of retained variables that produced the highest R^2CV was 55% for Beef + Pork CARS, 70% for MCUVE and VIP, and 85% for SPA. For lamb, the highest R^2CV and the corresponding percentage of retained variables were 0.90 (65%) for CARS, 0.91 (60%) for MCUVE and VIP, and 0.88 (95%) for SPA.

Post-hoc tests are usually used to determine which group differs [57]. The Nemenyi post-hoc test was used to select the minimum percentage of retained variables that were not significantly different (p-value > 0.05) from the highest value. For Beef + Pork (**Figure 6(a)**), 20% retained variables were selected using CARS, yielding 0.93 R^2CV. The MCUVE yielded an R^2CV of 0.94 with 40% of the variables retained. The minimum percentages of 25% and 30% were selected for the VIP and SPA methods, with R^2CV values of 0.95 and 0.94, respectively. For the Lamb + Pork model (**Figure 6(b)**), the minimum percentage of retained variables was 25% using the SPA method, with an R^2CV of 0.85. For the CARS method, the minimum 30% retained variables have an R^2CV of 0.79. The minimum percentage of retained variables for MCUVE and VIP was 35%, with R^2CV values of 0.82 and 0.81, respectively.

The R^2CV gap between the highest value and the minimum percentage of retained variables appears counterintuitive, since in Lamb+Pork it can reach 0.1 R^2CV. However, the result is expected and statistically supported. This may be attributed to the large variance across replications, which was sufficient for the test to conclude that they were not significantly different.

Selected variables

Feature selection methods often include a randomization step in their algorithm, such as CARS and MCUVE [58,59]. This randomization step is crucial for selecting highly adaptable variables that do not depend on the training set [58]. However, this randomization causes variability. The algorithm's results are a set of different variables in each run. To evaluate the stability and reproducibility of these algorithms, they should be run multiple times [37]. In this study, each algorithm was executed 100 times, and the final set of variables was determined by selecting variables with the highest frequency of selection across runs. This approach aimed to avoid picking up artifacts of random sampling and to focus on the truly important variables.



(a)

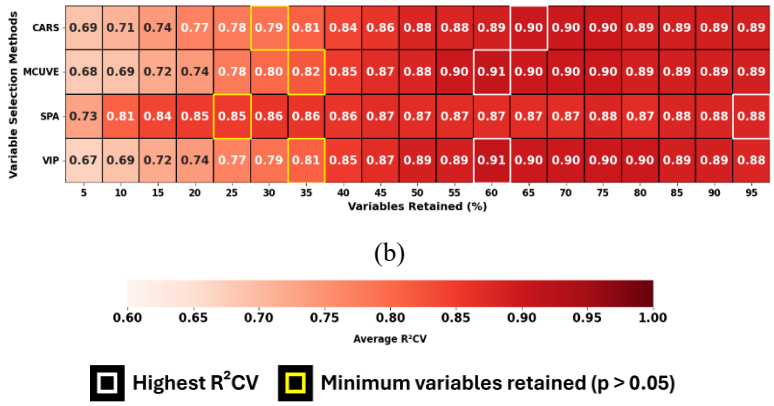


Figure 5 Comparison of PLSR performance (R^2CV) across different percentages of retained variables and feature selection methods for (a) Beef + Pork and (b) Lamb + Pork. White squares indicate the highest R^2CV, while yellow squares indicate the selected variable retention with the minimum number of variables that is not significantly different from the highest R^2CV.

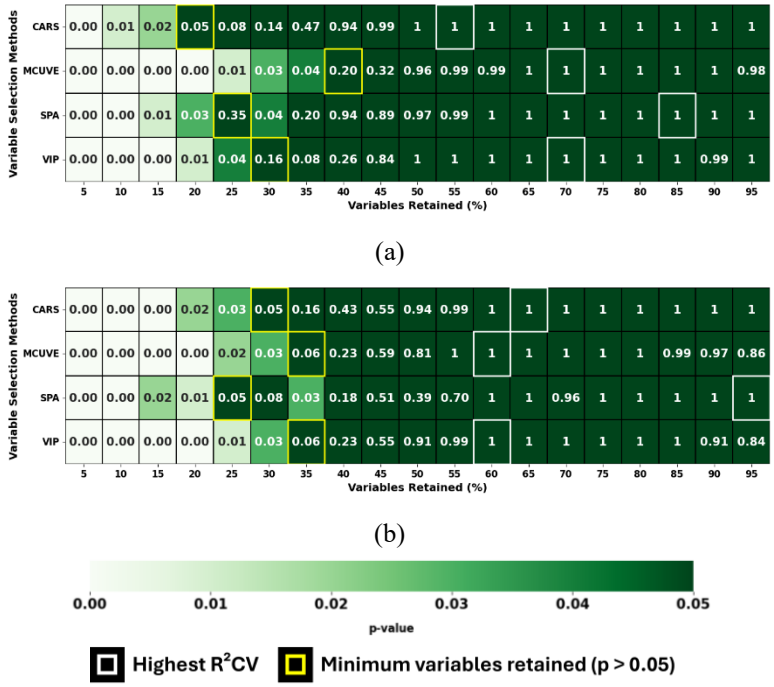


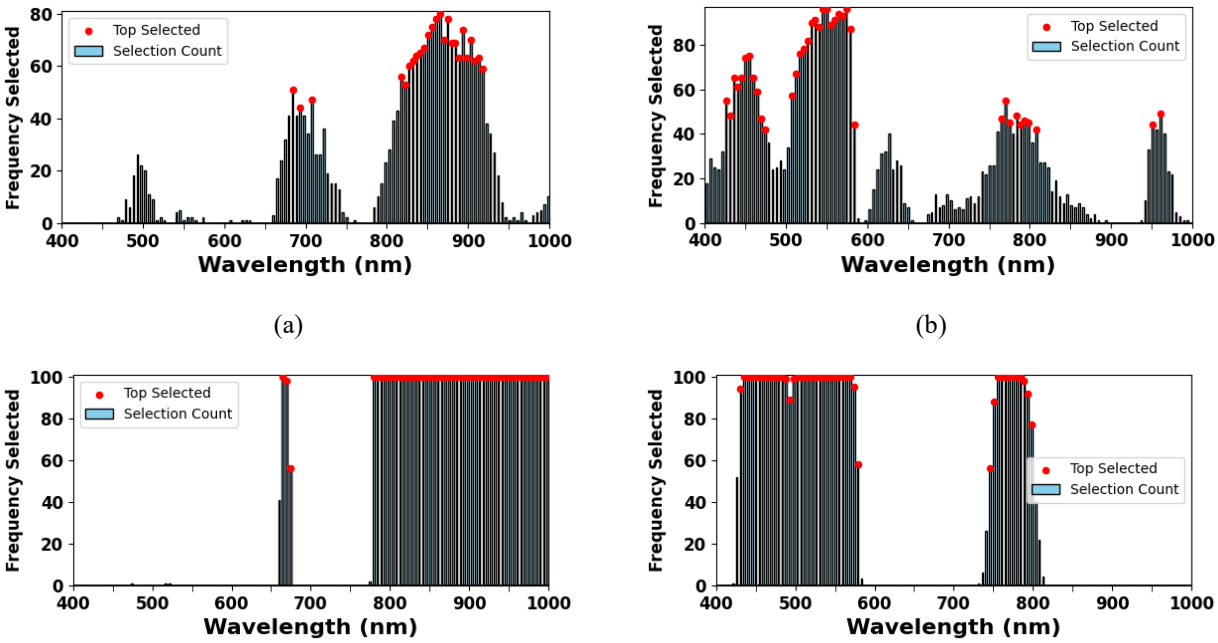
Figure 6 Nemenyi test *p*-values for the model with the highest R^2CV, compared across different percentages of retained variables and feature selection methods for (a) Beef + Pork and (b) Lamb + Pork. White squares indicate the p-value corresponding to the highest R^2CV, while yellow squares indicate the minimum variable set not significantly different from the highest R^2CV.

Figure 7 shows the distribution of selected variables across the wavelength range, along with the frequencies over 100 runs for each feature-selection method. The selection frequency highlights the algorithm's tendency to choose specific wavelength ranges. High-frequency ranges indicate that these wavelengths are likely associated with meaningful features of the pork-adulteration-related constituent. Therefore, these repeatedly selected spectral ranges have strong discriminatory power for detecting pork adulteration in beef and lamb.

The visualization result (**Figure 7**) shows that each feature-selection method has a distinct distribution of selected variables and reproducibility. Among the methods, CARS yielded the lowest reproducibility, reflected in the dispersed feature selections and generally low selection frequencies. The MCUVE showed much higher stability. Most selected variables reached a frequency of 100 selections and were concentrated in 2 wavelength clusters, although the specific ranges differed between the Lamb + Pork and Beef + Pork models. These wavelength clusters were broadly aligned with those selected by the VIP method, which exhibited perfect reproducibility, indicating that these wavelength ranges possess strong discriminative power for pork detection. In contrast to the clustered selections of MCUVE and VIP, SPA produced a perfectly stable pattern. The wavelengths selected by SPA spanned the entire wavelength range, yet each was chosen with a frequency of 100.

The selected variables by each feature-selection method for Beef + Pork and Lamb + Pork modes were different. For SPA, the selected wavelengths spanned the full spectral range (400 - 1,000 nm). Although the distribution was broad, both models showed a noticeable concentration of selected wavelengths that appear at 400 - 450 nm, which are affected by Soret absorption [60]. Additional wavelengths were selected at 450 - 600 nm, which are highly influenced by myoglobin pigments such as deoxymyoglobin, metmyoglobin, and oxymyoglobin [27]. Furthermore, several more wavelengths were spread at 600 - 1,000 nm.

In contrast to SPA, which selected similar wavelength ranges for both Beef + Pork and Lamb + Pork, the other chosen methods are substantially different wavelength clusters between the 2 models. In Beef + Pork models, a small group of wavelengths was selected around 650 - 700 nm, while most of the selected variables were concentrated between 780 and 1,000 nm. On the other hand, in Lamb + Pork models, small clusters of wavelengths were selected around 750 - 800 nm, while the remaining selected wavelengths were located around 430 - 600 nm, which were affected by Soret absorption and myoglobin [60]. Additionally, CARS method selected a few wavelengths at around 950 - 970 nm for Lamb + Pork model. The absorbances at 650 - 700 nm are affected by oxymyoglobin formation [61], while at 750 - 800 nm was attributed to Myoglobin (Deoxymyoglobin) [62,63]. The wavelengths at 800 - 1,000 were attributed to the meat components, such as protein, at 800 - 850 nm [61], fat at 850 - 950 nm [61,[64], and water at 950 - 1,000 nm [65].



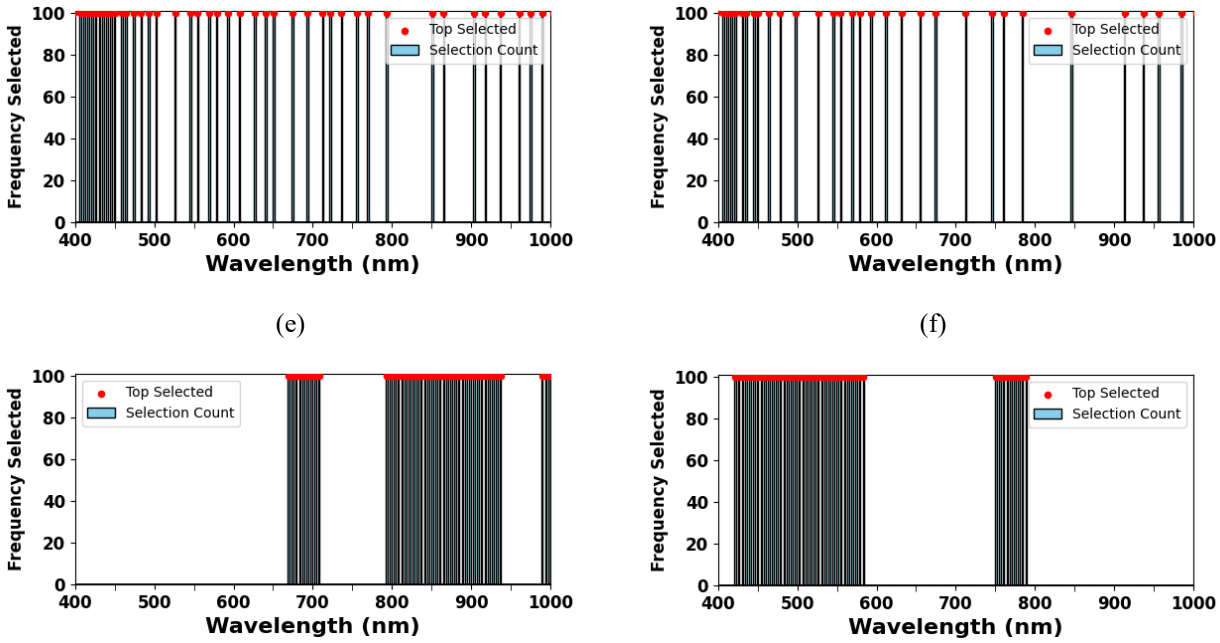


Figure 7 Meat spectra and the selected variables across various feature selection methods (CARS. MCUVE, SPA, and VIP); (a) Beef + Pork CARS (25 variables), (b) Lamb + Pork CARS (38 variables), (c) c. Beef + Pork MCUVE (51 variables), (d) Lamb + Pork MCUVE (44 variables), (e) Beef + Pork SPA (32 variables), (f) Lamb + Pork SPA (32 variables), (g) Beef + Pork VIP (38 variables) and (h) Lamb + Pork VIP (44 variables). Red scatter points indicate the variables with the highest selection frequency across 100 runs.

PLSR model using selected wavelength variables

The PLSR model's performance in predicting pork adulteration in beef and lamb using selected variables is presented in **Table 4**. Model performance is represented by high coefficients of determination (R^2) and low root mean square error (RMSE). The best PLSR model for predicting the percentage of pork in beef (Beef + Pork model) was obtained by applying the SPA method, yielding R^2 values of 0.974, 0.952, and 0.955, and RMSE values of 2.842, 3.840, and 3.752 for calibration, validation, and prediction, respectively. The PLSR for Lamb + Pork showed similar results, with the SPA selection method producing the best performance. For predicting pork concentration in lamb, the PLSR model using SPA yielded R^2 values of 0.922, 0.847, and 0.803, and RMSE values of 4.891, 6.862, and 7.685 for calibration, validation, and prediction, respectively. The PLSR results using HSI at the Vis-NIR region at 400 - 1,000 nm were lower compared to the results of the same samples (Beef + Pork and Lamb + Pork) when using HSI at the full-wavelength of the NIR region at 1,000 - 2,500 nm [24].

Table 4 PLSR performances with the selected variables.

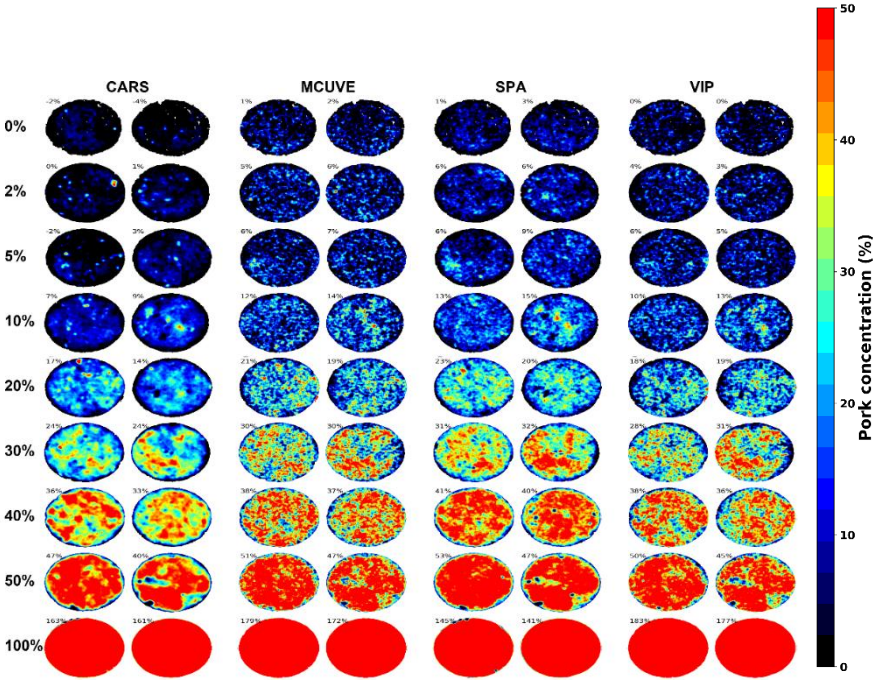
Features selection method	Retained variables (%)	Number of variables	LV	R^2C	RMSEC	R^2CV	RMSECV	R^2P	RMSEP
Beef + Pork									
CARS	20	25	9	0.945	4.110	0.927	4.749	0.938	4.360
MCUVE	40	51	11	0.961	3.469	0.942	4.254	0.951	3.891
SPA	**25**	**32**	**19**	**0.974**	**2.842**	**0.952**	**3.840**	**0.955**	**3.752**
VIP	30	38	10	0.959	3.566	0.941	4.284	0.946	4.099

Features selection method	Retained variables (%)	Number of variables	LV	R^2C	RMSEC	R^2CV	RMSECV	R^2P	RMSEP
Lamb + Pork									
CARS	30	38	19	0.885	5.953	0.793	7.998	0.755	8.604
MCUVE	35	44	23	0.910	5.263	0.816	7.540	0.764	8.433
SPA	**25**	**32**	**22**	**0.922**	**4.891**	**0.847**	**6.862**	**0.803**	**7.685**
VIP	35	44	24	0.907	5.352	0.805	7.757	0.727	9.075

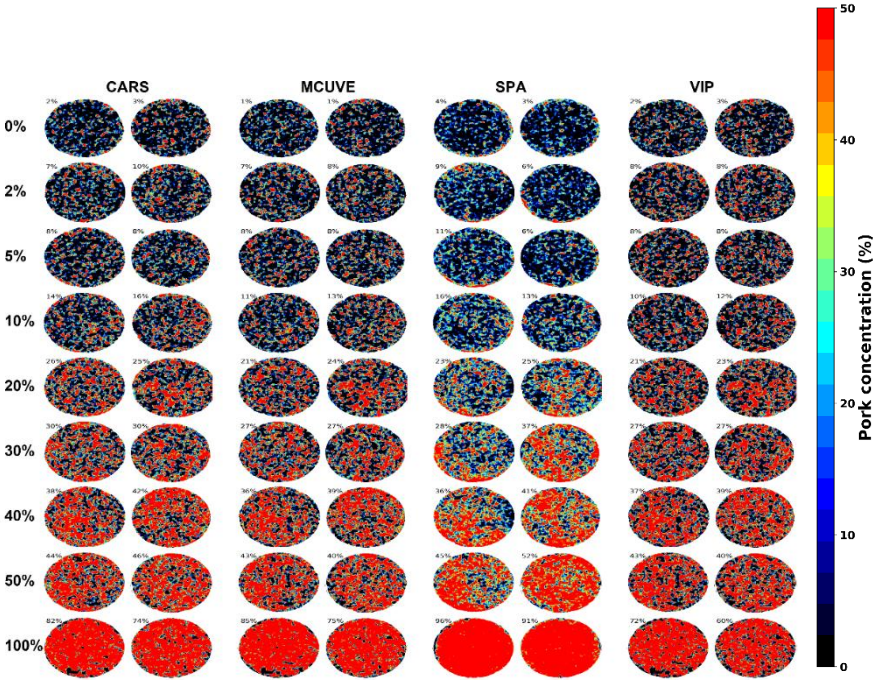
Masked image results

The beta coefficients obtained from the PLSR models were used to develop chemical images of pure and pork-adulterated beef and lamb. Pixel-wise masked images were constructed for each beef and lamb sample using various variable selection methods (CARS, MCUEVE, SPA and VIP), as shown in 0, to visualize differences in pork concentrations. The masked images were generated using a linear color scale to show variations in pork levels. The masked image enabled the determination of pork concentration, providing a rapid and easy method for quality assessment. The meat adulteration maps have been reported for identifying horsemeat in minced beef [66], pork in minced lamb and duck [26,67], and pork in beef and lamb [24]. As shown in our masked images (**Figure 8**), the pork was colored red while beef and lamb were colored blue; thus, the higher concentration of pork, the higher red color on the images.

According to **Figure 8**, the clearest images for portraying pork concentration in beef and lamb were generated using the SPA wavelength selection method. This result was in line with the PLSR results, which showed that the SPA method achieved the best performance for both the Pork + Beef and Pork + Lamb models (**Table 4**). However, as shown in 0, weaker chemical image performance was observed in the Lamb + Pork model than in the Beef + Pork model when using spectra from 400 - 1,000 nm. Unlike the masked images developed from the spectra obtained from hyperspectral imaging at 1,000 - 2,500 nm, as reported by Masithoh *et al.* [24]. The phenomenon occurs because there is not much spectral contrast in the 400 - 1,000 nm range, which is mostly contain pigment-related features and water absorptions [68]. This region is sufficient for distinguishing beef and pork, but is less effective for lamb and pork, since the similarity in pigment composition and surface chemistry of lamb and pork [69,70]. On the other hand, in the 1,000 - 2,500 nm region, stronger absorptions associated with water, protein, and especially fat are present [68] which amplifies the spectral distinction between lamb and pork, thus improving the mapping quality. Similarly, Wu and Sun [71] reported that hyperspectral imaging in the 900 - 1,700 nm range showed higher predictive ability than in the 400 - 1,000 nm range.



(a)



(b)

Figure 8 Chemical map of (a) beef + pork and (b) lamb + pork generated from PLSR models using different feature selection methods (CARS, MCUVE, SPA, and VIP). Each row represents a different pork concentration level (0% - 100%), while each column corresponds to a feature selection method. The color scale indicates the predicted pork concentration (%) at the pixel level.

Conclusions

This study confirms the potency of hyperspectral imaging in the visible and near-infrared (397.4 - 1,003.8 nm) region, combined with PLSR, as a rapid and non-destructive approach for detecting pork adulteration in beef and lamb. However, the performance of PLSR models is strongly influenced by spectral preprocessing and wavelength selection methods. The Savitzky-Golay

smoothing and SPA wavelength selection method produced the most robust results for both Beef + Pork and Lamb + Pork samples. Higher performance in PLSR was obtained from the Beef + Pork model compared to the Lamb + Pork model due to a stronger spectral contrast between beef and pork within the Vis-NIR region compared to lamb and pork. The constructed chemical maps visualized the distribution of pork adulterants in beef and lamb according to the level of pork concentration, thus offering an operative tool for quality inspection.

Although Vis-NIR hyperspectral imaging is effective for meat authentication, it has limited contrast in differentiating lamb and pork. Therefore, incorporating longer wavelengths in the NIR regions could enhance detection capabilities. In addition, the sample that was used in this study was under laboratory-controlled conditions and has not been validated using an independent dataset or real-world datasets. Future work will focus on evaluating the applicability of the proposed system under more realistic conditions. All in all, the study demonstrated that, with appropriate variable selection and a compact HSI system, the system can support efficient meat-fraud monitoring in food industries.

Acknowledgment

This work was supported by the Directorate of Research, Universitas Gadjah Mada, Indonesia, under Grant "Program Peningkerroratan Academic Excellence Skema A Tahun 2025".

Declaration of Generative AI in Scientific Writing

During the preparation of this work, the authors used generative AI tools (QuillBot and ChatGPT by OpenAI) to improve language, grammar, and structure. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the published article.

CRediT Author Statement

Rudiati Evi Masithoh: Conceptualization; Data curation; Funding acquisition; Supervision; Validation; Writing - Original draft. **Muhammad Fahri Reza Pahlawan:** Formal analysis; Investigation; Software; Visualization; Writing - Original draft. **Juntae Kim:** Investigation. **Reza Adhitama Putra Hernanda:** Investigation. **Hanim Zuhrotul Amanah:** Data curation; Investigation; Validation. **Honsoo Lee:** Data curation; Validation; Writing - Review. **Widodo Widodo:** Data curation; Writing - Review. **Byoung-Kwan Cho:** Resources; Software; Writing - Review.

References

[1] MM Bhat, H Jalal, PA Para, SA Bukhari, S Ganguly, AA Bhat, R Wakchaure and K Qadri. Fraudulent adulteration/Substitution of meat: A review. *Journal of Recent Research and Applied Studies* 2015; **2(12)**, 21-33.

[2] L Manning and JM Soon. Food safety, food fraud, and food defense: A fast evolving literature. *Journal of Food Science* 2016; **81(4)**, R823-R834.

[3] R Johnson. *Food fraud and 'Economically motivated adulteration' of food and food ingredients*. Congressional Research Service, Washington DC, 2014.

[4] Z Chen. Food adulteration: Understanding the risks and safeguarding our health. *Journal of Food Microbiology, Safety & Hygiene* 2023; **8(1)**, 1-2.

[5] K Robson, M Dean, S Brooks, S Haughey and C Elliott. A 20-year analysis of reported food fraud in the global beef supply chain. *Food Control* 2020; **116**, 107310.

[6] X Li, M Zang, D Li, K Zhang, Z Zhang and S Wang. Meat food fraud risk in Chinese markets 2012-2021. *npj Science of Food* 2023; **7(1)**, 12.

[7] MR Lozano, JH Chávez, FR López, RM Medina, ED Suárez, RM Medina and TM Ngapo. Horse meat sold as beef and consequent clenbuterol residues in the unregulated Mexican marketplace. *Food Control* 2020; **110**, 107028.

[8] M Font-i-Furnols and L Guerrero. Consumer preference, behavior and perception about meat and meat products: An overview. *Meat Science* 2014; **98(3)**, 361-371.

[9] M Tura, MA Gagliano, E Valli, M Petracci and TG Toschi. A methodological review in sensory analyses of chicken meat. *Poultry Science* 2024; **103(11)**, 104083.

[10] C Ruiz-Capillas and AM Herrero. Sensory analysis and consumer research in new product development. *Foods* 2021; **10(3)**, 582.

[11] C Yang, G Zhong, S Zhou, Y Guo, D Pan, S Wang, Q Liu, Q Xia and Z Cai. Detection and

characterization of meat adulteration in various types of meat products by using a high-efficiency multiplex polymerase chain reaction technique. *Frontiers in Nutrition* 2022; **9**, 979977.

[12] C Alamprese, JM Amigo, E Casiraghi and SB Engelsen. Identification and quantification of turkey meat adulteration in fresh, frozen-thawed and cooked minced beef by FT-NIR spectroscopy and chemometrics. *Meat Science* 2016; **121**, 175-181.

[13] M Modzelewska-Kapituła and S Jun. The application of computer vision systems in meat science and industry-A review. *Meat Science* 2022; **192**, 108904.

[14] IC Setiadi, AM Hatta, S Koentjoro, S Stendafity, NN Azizah and WY Wijaya. Adulteration detection in minced beef using low-cost color imaging system coupled with deep neural network. *Frontiers in Sustainable Food Systems* 2022; **6**, 1073969.

[15] RE Masithoh, A Zahroh, MFR Pahlawan and L Rahmawati. Pork adulteration in beef and mutton detection using visible near-infrared (Vis-NIR) spectroscopy and chemometrics. *Pertanika Journal of Science & Technology* 2025; **33(85)**, 109-119.

[16] S Weng, B Guo, P Tang, X Yin, F Pan, J Zhao, L Huang and D Zhang. Rapid detection of adulteration of minced beef using Vis/NIR reflectance spectroscopy with multivariate methods. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 2020; **230**, 118005.

[17] A Rady and A Adedeji. Assessing different processed meats for adulterants using visible-near-infrared spectroscopy. *Meat Science* 2018; **136**, 59-67.

[18] A López-Maestresalas, K Insausti, C Jarén, C Pérez-Roncal, O Urrutia, MJ Beriain and S Arazuri. Detection of minced lamb and beef fraud using NIR spectroscopy. *Food Control* 2019; **98**, 465-473.

[19] F Mabood, R Boqué, AY Alkindi, A Al-Harrasi, IS Al Amri, S Boukra, F Jabeen, J Hussain, G Abbas, Z Naureen, QMI Haq, HH Shah, A Khan, SK Khalaf and I Kadim. Fast detection and quantification of pork meat in other meats by reflectance FT-NIR spectroscopy and multivariate analysis. *Meat Science* 2020; **163**, 108084.

[20] T Leng, F Li, L Xiong, Q Xiong, M Zhu and Y Chen. Quantitative detection of binary and ternary adulteration of minced beef meat with pork and duck meat by NIR combined with chemometrics. *Food Control* 2020; **113**, 107203.

[21] M Manley. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews* 2014; **43(24)**, 8200-8214.

[22] Y Liu, H Pu and DW Sun. 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.

[23] WH Su and DW Sun. Evaluation of spectral imaging for inspection of adulterants in terms of common wheat flour, cassava flour and corn flour in organic Avatar wheat (*Triticum* spp.) flour. *Journal of Food Engineering* 2017; **200**, 59-69.

[24] RE Masithoh, MFR Pahlawan, J Kim, MAA Arief, H Kurniawan, RAP Hernanda, HZ Amanah and BK Cho. Shortwave infrared hyperspectral imaging for the determination of pork adulteration in minced beef and lamb. *Food Control* 2024; **166**, 110736.

[25] RE Masithoh, RAP Hernanda, MFR Pahlawan, J Kim, HZ Amanah and BK Cho. Shortwave near infrared-hyperspectral imaging spectra to detect pork adulteration in beef using partial least square regression coupled with VIP wavelength selections method. *Optics* 2025; **6(1)**, 1.

[26] M Kamruzzaman, DW Sun, G ElMasry and P Allen. Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta* 2013; **103**, 130-136.

[27] M Peyvasteh, A Popov, A Bykov and I Meglinski. Meat freshness revealed by visible to near-infrared spectroscopy and principal component analysis. *Journal of Physics Communications* 2020; **4(9)**, 095011.

[28] MM Hasan, MMA Chaudhry, C Erkinbaev, J Paliwal, SP Suman and A Rodas-Gonzalez. Application of Vis-NIR and SWIR spectroscopy

for the segregation of bison muscles based on their color stability. *Meat Science* 2022; **188**, 108774.

[29] SD Osborne, RB Jordan and R Künnemeyer. Method of wavelength selection for partial least squares. *Analyst* 1997; **122(12)**, 1531-1537.

[30] J Jin and Q Wang. Evaluation of informative bands used in different PLS regressions for estimating leaf biochemical contents from hyperspectral reflectance. *Remote Sensing* 2019; **11(2)**, 11-15.

[31] J Fu, HD Yu, Z Chen and YH Yun. A review on hybrid strategy-based wavelength selection methods in analysis of near-infrared spectral data. *Infrared Physics & Technology* 2022; **125**, 104231.

[32] SC Seidemen, HR Cross, GC Smith and PR Durland. Factors associated with fresh meat color: A review. *Journal of Food Quality* 1984; **6**, 211-237.

[33] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, A Müller, J Nothman, G Louppe, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot and É Duchesnay. Scikit-Learn: Machine learning in Python. *Journal of Machine Learning Research* 2011; **12**, 2825-2830.

[34] P Virtanen, R Gommers, TE Oliphant, M Haberland, T Reddy, D Cournapeau, E Burovski, P Peterson, W Weckesser, J Bright, SJ van der Walt, M Brett, J Wilson, KJ Millman, N Mayorov, ARJ Nelson, E Jones, R Kern, E Larson, …, P van Mulbregt. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods* 2020; **17(3)**, 261-272.

[35] L Zhang, J Sun, X Zhou, A Nirere, X Wu and R Dai. Classification detection of saccharin jujube based on hyperspectral imaging technology. *Journal of Food Processing and Preservation* 2020; **44(8)**, e14591.

[36] T Mehmood, KH Liland, L Snipen and S Sæbø. A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* 2012; **118**, 62-69.

[37] M Talebi, G Schuster, RA Shellie, R Szucs and PR Haddad. Performance comparison of partial least squares-related variable selection methods for quantitative structure retention relationships modelling of retention times in reversed-phase liquid chromatography. *Journal of Chromatography A* 2015; **1424**, 69-76.

[38] F Buckermann, S Heil and A Krause. AUSWAHL - A scikit-learn compatible package for wavelength selection, Available at: https://auswahl.readthedocs.io/en/latest/index.html, accessed January 2026.

[39] H Liu, Y Yuan, G Wang, W Xu, L Zhao and J Li. An advanced variable selection method based on information gain and Fisher criterion reselection iteration for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems* 2023; **235**, 104796.

[40] DW Gee, R Phitayakorn, A Khatri, K Butler, JT Mullen and ER Petrusa. A pilot study to gauge effectiveness of standardized patient scenarios in assessing general surgery milestones. *Journal of Surgical Education* 2016; **73(6)**, e1-e8.

[41] EA Elamir. A graphical approach for Friedman test: Moments approach. *arXiv* 2022; **2202**, 09131.

[42] M Terpiłowski. scikit-posthocs: Pairwise multiple comparison tests in Python. *Journal of Open Source Software* 2019; **4(36)**, 1169.

[43] A Kazemi, A Mahmoudi, H Veladi, A Javanmard and M Khojastehnazhand. Rapid identification and quantification of intramuscular fat adulteration in lamb meat with VIS - NIR spectroscopy and chemometrics methods. *Journal of Food Measurement and Characterization* 2022; **16(3)**, 2400-2410.

[44] Å Rinnan, F van den Berg and SB Engelsen. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry* 2009; **28(10)**, 1201-1222.

[45] I Bilbao, J Bilbao and C Feniser. Adopting some good practices to avoid overfitting in the use of machine learning. *WSEAS Transactions on Mathematics* 2018; **17**, 274-279.

[46] J Rodrigues, A Alves, H Pereira, DDS Perez, G Chantre and M Schwanninger. NIR PLSR results obtained by calibration with noisy, low-precision reference values: Are the results acceptable? *Holzforschung* 2006; **60(4)**, 402-408.

[47] CPJ Chen, RR White and R Wright. Common pitfalls in evaluating model performance and strategies for avoidance in agricultural studies.

Computers and Electronics in Agriculture 2025; **234**, 110126.

[48] SS Vasanthadev and P Shanthi. Influence of data pre-processing techniques for PLSR model to predict blood glucose by NIR spectroscopy. *Optics and Spectroscopy* 2022; **130(5)**, 613.

[49] RA Klein, KA Ratliff, M Vianello, RB Adams, Š Bahník, MJ Bernstein, K Bocian, MJ Brandt, B Brooks, CC Brumbaugh, Z Cemalcilar, J Chandler, W Cheong, WE Davis, T Devos, M Eisner, N Frankowska, D Furrow, EM Galliani, …, BA Nosek. Investigating variation in replicability: A 'many labs' replication project. *Social Psychology* 2014; **45(3)**, 142-152.

[50] J Engel, J Gerretzen, E Szymańska, JJ Jansen, G Downey, L Blanchet and LM Buydens. Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry* 2013; **50**, 96-106.

[51] M Enser, K Hallett, B Hewitt, GAJ Fursey and JD Wood. Fatty acid content and composition of English beef, lamb and pork at retail. *Meat Science* 1996; **42(4)**, 443-456.

[52] A Listrat, B Lebret, I Louveau, T Astruc, M Bonnet, L Lefaucheur, B Picard and J Bugeon. How muscle structure and composition influence meat and flesh quality. *The Scientific World Journal* 2016; **2016(1)**, 3182746.

[53] L Lefaucheur. A second look into fibre typing - Relation to meat quality. *Meat Science* 2010; **84(2)**, 257-270.

[54] S Schiaffino and C Reggiani. Fiber types in mammalian skeletal muscles. *Physiological Reviews* 2011; **91(4)**, 1447-1531.

[55] T Mehmood, S Sæbø and KH Liland. Comparison of variable selection methods in partial least squares regression. *Journal of Chemometrics* 2020; **34(6)**, e3226.

[56] Z Xiaobo, Z Jiewen, MJW Povey, M Holmes and M Hanpin. Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta* 2010; **667(1-2)**, 14-32.

[57] CE Agbangba, ES Aide, H Honfo and RG Kakai. On the use of post-hoc tests in environmental and biological sciences: A critical review. *Heliyon* 2024; **10(3)**, e25131.

[58] H Li, Y Liang, Q Xu and D Cao. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta* 2009; **648(1)**, 77-84.

[59] W Cai, Y Li and X Shao. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems* 2008; **90(2)**, 188-194.

[60] H Ayaz, M Ahmad, A Sohaib, MN Yasir, MA Zaidan, M Ali, MH Khan and Z Saleem. Myoglobin-based classification of minced meat using hyperspectral imaging. *Applied Sciences* 2020; **10(19)**, 6862.

[61] A Aït-Kaddour, M Loudiyi, O Boukria, J Safarov, S Sultanova, D Andueza, A Listrat and Y Cahyana. Beef muscle discrimination based on two-trace two-dimensional correlation spectroscopy (2T2D COS) combined with snapshot visible-near infrared multispectral imaging. *Meat Science* 2024; **214**, 109533.

[62] MA Hashem, A Ahmmed and MM Hasan. Non-destructive discrimination of fresh, aged, and frozen-thawed beef using portable near-infrared spectroscopy combined with explainable artificial intelligence. *LWT* 2025; **236**, 118659.

[63] N Morsy and DW Sun. Robust linear and non-linear models of NIR spectroscopy for detection and quantification of adulterants in fresh and frozen-thawed minced beef. *Meat Science* 2013; **93(2)**, 292-302.

[64] G ElMasry, DW Sun and P Allen. Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging. *Food Research International* 2011; **44(9)**, 2624-2633.

[65] D Cozzolino and I Murray. Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. *LWT - Food Science and Technology* 2004; **37(4)**, 447-452.

[66] M Kamruzzaman and S Oshita. Hyperspectral imaging in tandem with multivariate analysis and image processing for non-invasive detection and visualization of pork adulteration in minced beef. *Analytical Methods* 2015; **7**, 7496-7502.

[67] X Zheng, Y Li, W Wei and Y Peng. Detection of adulteration with duck meat in minced lamb meat by using visible near-infrared hyperspectral imaging. *Meat Science* 2019; **149**, 55-62.

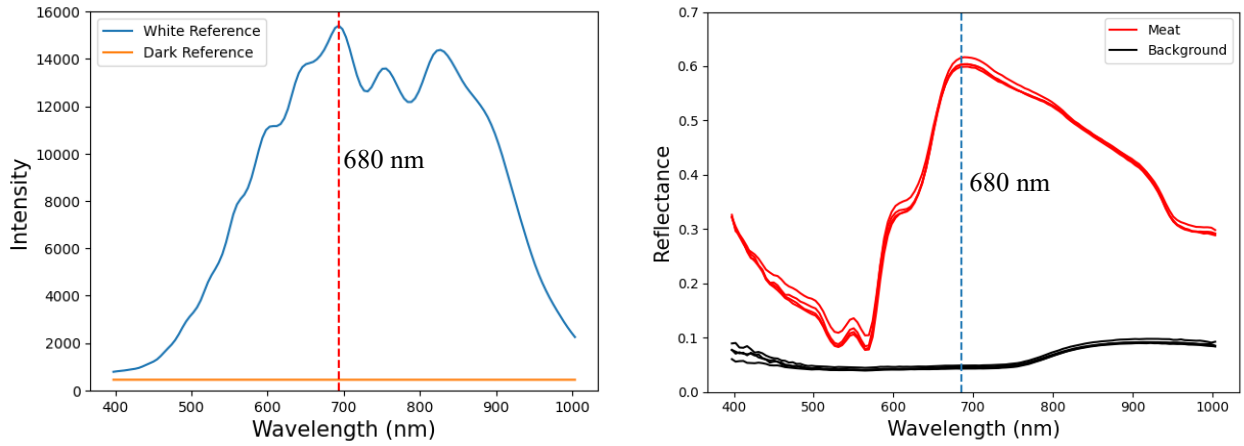
[68] S León-Ecay, K Insausti, A López-Maestresalas and N Prieto. Non-destructive spectroscopy-based technologies for meat and meat product discrimination - A review. *Meat Science* 2025; **228**, 109893.

[69] A Dashti, J Müller-Maatsch, E Roetgerink, M Wijtten, Y Weesepoel, H Parastar and H Yazdanpanah. Comparison of a portable Vis-NIR hyperspectral imaging and a snapscan SWIR hyperspectral imaging for evaluation of meat authenticity. *Food Chemistry: X* 2023; **18**, 100667.

[70] A Dashti, J Müller-Maatsch, Y Weesepoel, H Parastar, F Kobarfard, B Daraei, MHS AliAbadi and H Yazdanpanah. The feasibility of two handheld spectrometers for meat speciation combined with chemometric methods and its application for halal certification. *Foods* 2021; **11(1)**, 71.

[71] D Wu and D Sun. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review - Part II: Applications. *Innovative Food Science & Emerging Technologies* 2013; **19**, 15-28.

Supplementary Materials



(a) References intensity comparison (b) Meat and background spectra comparison

Figure S1 The optimal background separation wavelength (680 nm) is based on (a) the maximum intensity of the white references and (b) the maximum difference between meat and background.