Fusion of spectral and textural data of hyperspectral imaging for glycine content prediction in beef using SFCN algorithms

Yu Lv
Ningxia University
Fujia Dong
Ningxia University
Jiarui Cui
Ningxia University
Ruiming Luo
Ningxia University
Songlei Wang (wangsonglei163@126.com)
Ningxia University
Argenis Rodas-Gonzalez
University of Manitoba
Jie Hao
Ningxia University
Sijia Liu
Ningxia University

Research Article

Keywords: Hyperspectral, Glycine, Fusion of spectral and textural data, SFCN algorithms, Visualization

Posted Date: August 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1925394/v1

Abstract

Glycine, the simplest free amino acid, is one of the most important factors affecting the flavor of beef. In this paper, a fast and non-destructive method combining near-infrared hyperspectral (900–1700 nm) and textural data was first proposed to determine the content and distribution of glycine in beef. On the basis of spectral information pre-processing, spectral features were extracted by the interval Variable Iterative Space Shrinkage Approach, Competitive Adaptive Reweighting algorithm and Uninformative Variable Elimination (UVE). The glycine content prediction models were established by partial least squares regression, least squares support vector machine, and the optimized shallow full convolutional neural network (SFCN). Among them, the UVE-SFCN model obtained better results with prediction set determination coefficient (R_P^2) of 0.8725). Further, textural features were extracted by the gray level co-occurrence matrix and fused with the spectral information of the best feature band to obtain an optimized UVE-FSCN-fusion model (R_P^2 = 0.9005, root mean square error = 0.3075, residual predictive deviation = 0.2688). Compared with the full spectrum and characteristic wavelength spectrum models, R_P^2 was improved by 6.41% and 3.10%. The best fusion model was visualized to visually represent the distribution of glycine in beef. The results showed that the prediction and visualization of glycine content in beef were feasible and effective, and provided a theoretical basis for the hyperspectral study of meat quality monitoring or the establishment of an online platform.

Introduction

Beef has gradually become the main edible meat product with abundant proteins and similar amino acids ratio to human's needs [1]. In the past, amino acids in meat were generally studied in the form of total nitrogen or total amino acids. For example, the total amino acid composition of different kinds of chickens was compared, and the ranking of amino acid content was obtained [2]. Basic amino acids were used to reduce the content of sodium chloride in meat [3]. Free amino acid, a single amino acid molecule in a free state, could be directly absorbed and utilized. It is the main nutrient source in beef products and is also an important odor substance. The flavor function of amino acids is related to its hydrophobicity to a certain extent. When the hydrophobicity of amino acids is low, they are mainly sweet, on the contrary, they are bitter. Amino acids with acidic side chains are one of the most significant prerequisites for sourness and freshness [4]. The flavor of broth can be replicated through different free amino acids, proving its important role in flavor generation [5, 6]. Glycine is the simplest flavor amino acid among free amino acids with low hydrophobicity, so it contributes sweet taste. It also plays a variety of vital roles in physiological processes, inducing immune regulation [7], antibacterial and bactericidal action [8], enhancing Maillard reaction, enhancing meat taste [9], and improving the body's antioxidant capacity [10]. Too much or too little glycine intake may affect the absorption balance of amino acids in the human body, leading to nutritional imbalance and poor health [11, 12]. Therefore, the quantitative detection of glycine content is particularly important.

Previous, studies on free amino acids mainly focused on the innovative applications of traditional physical and chemical detection methods, including high performance liquid chromatography (HPLC), reverse high performance liquid chromatography (RP-HPLC), and amino acid automatic analyzer. HPLC and RP-HPLC were used to determine the content of free amino acids in poultry and plasma [12, 13]. Electronic tongue and electronic nose were used to evaluated and analyzed umami amino acids in beef [14]. Amino acid analyzer was used to compare and evaluate the free amino acid content and antioxidant activity of beef according to the country of origin and marbling score [15]. Traditional testing methods are accurate and high-throughput, but they are time-consuming, destructive, tedious and cannot meet the requirements of rapid non-destructive testing.

With the development of science, hyperspectral imaging (HSI) technology, an efficient and nondestructive detection method, has been employed in food detection. HSI integrates with the main features of imaging technology and spectroscopy, and is used to obtain continuous spectral and spatial information from objects simultaneously. Defective products were identified by combining pH value and other physical properties with near-infrared (NIR) hyperspectral [14, 16]. Protein, lipid and water are the main components in the nutritional components of beef which was used for hyperspectral imaging [17]. Using VIS/NIR spectroscopy combined with a variety of methods to quickly detect the adulteration of ground meat, the accuracy was improved, compared with the original method [18]. The non-destructive detection of beef corruption index was carried out by fluorescence hyperspectral, and a model for the corruption index was established [16]. In addition, hyperspectral imaging has great potential in the study of free amino acids. Textural information, as important feature information of hyperspectral technology, was fused with spectral information for food detection. For example, predicted and visualized TVB-N content in cooked beef to quickly and non-destructively determine beef freshness [16]. A prediction model for palmitic and oleic acid content in mutton was established by integrating NIR hyperspectral with texture information [19]. However, to our knowledge, the prediction of free amino acid content in beef by hyperspectral and textural data fusion has not been reported.

In recent years, convolutional neural network (CNN) plays an important role in the field of machine learning. CNN reduced the complexity of the traditional neural networks and optimized mass data, becoming the ideal choice for analyzing and processing hyperspectral data. The mixed deep learning model based on hyperspectral remote sensing was used to predict strawberry yield with the model accuracy rate reaches 70% [20]. HSI was combined with CNN in soybean price prediction and Conv-LSTM used to improve RMSE by 31%. Fully convolutional networks (FCN) was a framework used for image semantic segmentation [21]. FCN replaced the full connection layer with the convolution layer, accepting any input size while increasing the network depth. It becomes more efficient with head of the problems of double storage and computing convolution. Remote sensing hyperspectral and FCN were used to predict the depth of mature rice, and its performance was much better than the regression model based on Vis-HIS [11]. However, the up-sampling results of SCN are fuzzy and not sensitive to details. Therefore, a lightweight shallow full convolutional neural network (SFCN) was proposed in this study. The network depth and the upsampling influence were reduced to fit in smaller datasets during the running process of SFCN [22].

In current paper, spectral data of representative beef varieties in Northwest China were collected by NIR hyperspectral (900–1700 nm) technology. After processing, the spectral data of the optimal characteristic wavelength was obtained. Textural information was got through gray co-occurrence matrix (GLCM) method. The deep learning model SFCN of glycine content was established and optimized with fusion of hyperspectral and textural information. The study may provide a reference for the rapid quantitative detection of glycine content in beef and a theoretical basis for the development of hyperspectral quality monitoring online platform; and help improve beef quality and the sustainable and healthy development of the industry. The complete test process was shown in Fig. 1.

Materials And Methods

Sample preparation

A representative breed of cattle from Northwest China was selected, with average fresh carcass weight of 285 ± 45 kg. The cattle were slaughtered in accordance with the principles and guidelines established by the Animal Care and Use Committee of the Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences. Acid was released in 24 h after death, then dissected. The three muscles, foreleg (FL), hindquarter (HQ) and longissimus dorsi (LD), were labeled, vacuumed into polyethylene bags, and transported to the Agricultural Products Non-destructive Testing Laboratory, Food College of Ningxia University, China by an ice box. Visible fat and connective tissue were removed and cut into slices approximately 4cm×3cm×1cm thick. They were then transported to the Spectral Analysis Laboratory and Food Testing and Analysis Center, Ningxia University, respectively. The glycine content was measured by an amino acid automatic analyzer, and the sample sections of the same site were measured by NIR hyperspectral. The samples were placed at room temperature for spectral measurement, and a total of 360 beef samples were obtained, of which the number of samples in each part was 120. The key to established an effective multivariate prediction model was ensure the chemical and spectral values of the same cross-section sample were measured at same time.

Chemical Analysis

A mixed beef sample (2.00 g) was put into a 10 mL centrifuge tube, 0.02 mol/L hydrochloric acid was added to dissolve in a constant volume. Then 5.00 mL methanol and 5.00 mL water were added respectively to activate the C_{18} pre-treated column. A 2.50 mL sample was added, followed by 1.50 mL 0.02 mol/L hydrochloric acid. After passing the column, 0.02 mol/L hydrochloric acid was added to a final volume of 5.00 mL, then filtered through a 0.45 μm membrane. Parameter settings for the amino acid automatic analyzer were: wavelength: 570 nm and 440 nm; injection volume: 20.00 μL; reaction temperature: 135 ± 5℃; separation column 57℃; chromatographic column: sulfonate cation resin separation.

Near-infrared hyperspectral system and image calibration

The NIR spectrometer used in the experiment include 256 wavelengths (from 900 nm to 1700 nm) with a spectral resolution of 5 nm. The system consisted of a hyperspectral imaging spectrometer (ImspectorN17E, Spectral Imaging Ltd., Oulu, Finland), four 35 W halogen tungsten lamps (HSIA-LS-TDIF, Zolix Instruments Co., Ltd., Beijing, China), an electronically controlled displacement platform (PSA200-11-X, Zolix Instru-ments Co., Ltd., Beijing. China), a CCD camera (Zelos-285GV, Kappa optronics Gmb H, Gleichen, Germany), and a computer (Lenovo Inter (R) Core i7-2600CPU@ 3.40 GHz, RAM 4.00G).

During spectral scanning, light source intensity, exposure time, meat surface texture shape and color may affect the size and spatial resolution of hyperspectral images collected, so reasonable scanning parameters should be set before image collection. Based on the preliminary experiment, the micro-hyperspectral scanning parameters were determined as follows: scanning speed 1.4 mm/s, back off speed 2 mm/s, exposure time 10 m/s, and gain 5. Finally, the resolution of the cubic spectral image was 320 Pix 360 Pix 256 Pix.

Black and white correction was performed on the original image before the experiment began, to eliminate the influence of instrument light source. Briefly, the white image reflectivity of a standard total reflection whiteboard (100% of the Teflon board), and an all-black image reflectivity covering the lens (0% of the reflectivity) were obtained, then the black-and-white correction image of the sample was calculated based on the following formula:

$$I = \frac{R - R_d}{R_W - R_d} \times 100\%$$

1

Where, I is the obtained calibration image, R is the original obtained image, R_d represents dark current image, and R_w represents white reference image.

Region of interest selection and spectral information extraction

The surface of beef sample was not smooth, and there was a shadowed area at the junction of sample and background in hyperspectral image collection. Meanwhile, the moisture on the surface of meat samples, fascia and light source irradiation may have interference to spectral information. To weaken these effects, the size of the image was adjusted. On the premise of retaining the complete information of the tested sample, the unnecessary pixel regions in the original hyperspectral image were clipped by ENVI 4.8 software, and part of the background image was removed to reduce spatial dimensionality. The region of interest (ROI) of the sample was extracted by threshold method (threshold value 0.14). The corresponding spectral information in ROI was extracted and its average value was calculated as the spectral value of the sample to be measured.

Spectral data analysis

A prediction model was established with spectral information. Before modeling, the data should were analyzed as follows:

a. Division of correction set and prediction set. Based on sample set partitioning and joint x-y distance (SPXY) algorithm, 75% of the samples were selected as the calibration set, the remaining were modeled for the prediction set, and the prediction error of each sample was obtained.
b. Pretreatment. Pretreatment improves the spectral resolution of the overlapping data, and minimizes the adverse effects of the instrument. The following pretreatment methods were used: Normalize, Standard Normal variable (SNV), Detrend, Baseline; and multiplication scattering correction (MSC).
c. Characteristic wavelength extraction. The whole band spectral data have high dimension and complicated information, and so may need large amount of calculations, with low model accuracy. Therefore, dimensionality reduction of the original spectra is required to extract wavelength information closely related to glycine content. In this experiment, the interval Variable Iterative Space Shrinkage Approach (iVISSA), Competitive Adaptive Reweighting algorithm (CARS) and Uninformative Variable Elimination (UVE) were used to extract eigenvalues of the spectra.

Establishment of prediction model and performance evaluation

Traditional linear regression model PLSR, nonlinear regression analysis model LSSVM and deep learning model SFCN were used to establish models for glycine content prediction, and different modeling methods were compared.

PLSR is a multi-linear regression modeling method with integrates the ascendancy of multiple linear regression, principal component analysis, and canonical correlation analysis, and is fast in modeling and sensitive to outliers. By standardizing X and Y, the correlation coefficient matrix is established, the contribution rate is calculated, and then the regression equation is made. LSSVM is used for nonlinear mapping linearization to increase accuracy and reliability, on the basis of SVM optimization of the problem constraints. Both linear and nonlinear models are traditional and classical, and have been widely used in modeling and analysis.

$$Y = X^T \cdot X$$

2

1D-CNN model has been used for spectroscopic and chemometric analysis, but it is limited by the effective features of data sets and one-dimensional convolution extraction models. In this study, a two-dimensional spectral information matrix was obtained by one-dimensional spectral self-multiplication (Formula 2). The depth, height, and width were arranged and expressed in SFCN. The network was mainly composed of an input layer, convolution layer, activation function, Pooling layer, and other parts (Fig. 2). The neurons in each layer were connected only to a small area of the previous layer, forming a Locally-Connected Layer. Rectified linear unit (ReLU) was used as the activation function of the convolution layer.

Input 2d data were convolved and pooled, and output. parameter of feature extraction module were set (Table 1). The fill value was 0, the step size was 1.

In this study, R_{CV}^2, RMSECV, prediction set determination coefficient (R_P^2), root mean square error of prediction (RMSEP) and residual prediction deviation (RPD) were used to evaluate the performance of the quantitative analysis models.

ENVI 4.8 was used to extract the ROI, Unscrambler X 10.4 was used for data preprocessing and PLSR modeling, and Matlab R2021a was used for characteristic wavelength extraction, SFCN and LSSVM modeling. All data analyses were performed on a computer with Intel Core I7-10700, 12 GB CPU, 2.9 gHz frequency, NVIDIA 3060 GPU and 16 GB memory.

Extraction of image texture information

Hyperspectral technology can not only provide important image information, but also provide rich texture information. Texture information is the projection of high-dimensional spectral information to low-dimensional space. The objects around can be effectively distinguished by measuring the difference between pixels and spatial neighborhood. GLCM is a comment method for texture extraction analysis. Texture features based on GLCM were reported to be very beneficial for distinguishing different meat conditions and predicting food attributes [23].

ENVI 4.8 software was used to perform PCA transformation on hyperspectral images, and GLCM was applied for effective PC images of ROI. Four mutually independent GLCM-based texture features, including contrast (CONT), homogeneity (HOM), energy (ENG) and correlation (CORR), were extracted from gray co-occurrence matrices with different angles (0°, 45°, 90°, and 135°), and the distance between pixel pairs was set to 1 pixel. The values in four different directions were averaged to record the texture features of each sample. Four texture factors were obtained by the calculation formula (Table 2).

Glycine visualization in beef

Hyperspectral images of characteristic wavelengths were selected, each beef sample image was segmented (ENVI 4.8 software), and the information of each pixel point was extracted (Matlab 2021a). A model was established based on the data, and the glycine content of each pixel of the sample was calculated to obtain a grayscale image. Pseudo-color processing was performed on grayscale images to predict the content of glycine in beef samples through different color bands, resulting in color visualization.

Results And Discussion

Chemical characteristics

The contents of glycine in different parts of selected beef were significantly different (Fig. 3). The content of glycine was the highest in LD, and was significantly lower in HQ (p 0.05) and FL (p 0.01). This may be

related to the content of free amino acids during catabolism. When the energy supply of free amino acids exceeds the energy required by the body, it participates in the tricarboxylic acid cycle and carbohydrate metabolism, and oxaloacetate (4C), citric acid (6C), cis aconitonic acid (6C), A-ketoglutaric acid (5C), fumaric acid (4C) and other intermediates were produced. Free amino acids in HQ and FL of beef are decomposed for energy during exercise, while LD has less movement, so it needs more free amino acids to participate in the synthesis. This is consistent with the results of fatty acid content in Tan mutton [19].

The test sample (n = 360) were divided into calibration sets and prediction sets (Table 3). The best results were obtained with SPXY method, and the calibration set and the prediction set were not very different with similar data distribution. Indicating that the glycine distribution was uniform between different muscles in this sample.

Spectral features

The average spectral curves of three different parts of the beef sample in the spectral range of 900–1700 nm can be seen that LD tended to have the higher spectral reflectance than HQ and FL and the overall trend of the three parts were roughly the same (Fig. 4). This is similar to the results obtained in the near infrared spectroscopy [24]. In the range of 1050–1400 nm, the spectral reflectance values of the three parts were different. After 1400 nm, the spectral reflectance values were lower than 0.1 due to the strong water absorption in the muscle [25]. Peaks (Fig. 4a) and valleys (Fig. 4b) occurred in the entire spectral reflectance curve, mainly because chemical bonds absorb energy at specific wavelengths, the spectral intensity of which is determined by changes in the molecular dipole moment [26]. Therefore, the band in the range of 1400–1700 nm was the water absorption band of the second and first overtones related to O-H bond frequency doubling, indicating that water accounted for the largest proportion and absorbed most of the near-infrared radiation in the beef muscle. At 1050 nm, 1200 nm, and 1400 nm, O□H, C□H, and N□H overtones appeared respectively.

Prediction models using full spectral range

Normalize, SNV, Detrend, Baseline, and MSC were used to process the data, and PLSR model of pretreated spectral reflectance was established to predict the content of glycine. Comparison of the R_{CV}^2/R_P^2 and RMSEC/RMSEP values of the PLSR model and the prediction set showed that a good result was obtained by the data preprocessing (Fig. 5).

Baseline was found to be the best preprocessing method for glycine prediction models in beef samples. The data obtained by other preprocessing methods also showed an improvement trend, indicating that most preprocessing methods provided favorable trends for the spectral data. Thus, selecting the right pretreatment method may increase the reliability of the prediction model. The spectral curves of data preprocessed by different methods were similar to baseline method (Fig. 6).

Selection of effective wavelengths

In this study, iVISSA, CARS and UVE were used to extract effective feature bands in the whole band range, and then the feature wavelength extraction method with the best effect was selected through modeling.

The iVISSA algorithm extracted characteristic wavelengths through two iterative cycles of global and local search, and finally obtained a set of optimized effective wavelengths. Set the iVISSA parameter, take the maximum principal component number 20, adopt half-fold cross-validation, select "Center" as the data processing method, set the number of binary matrix samples to 500, the optimal subset ratio to 0.1. In the iterative operation, the sampling weight value of each wavelength varied with the number of iterations, and the weight ranged from 0 to 1 (Fig. 7a). After 10 iterations, the sampling weight value was basically stable, and a total of 46 characteristic wavelengths were extracted. The 46 characteristic wavelengths selected by the iVISSA method were widely and uniformly distributed in the whole waveband, indicating that the iVISSA algorithm may well preserved the synergistic effect between wavelengths (Fig. 7b).

When the CARS method was used to extract characteristic wavelengths, the variable subset with the smallest RMSECV value was selected as the optimal subset. During the running process, the number of variables and the number of runs showed an exponentially decreasing relationship (Fig. 8a). RMSECV first decreased slowly, with a minimum value at the 275th sample run, and then increased, the number of optimal variable subsets decreased, and some valid information was removed, resulting in a decrease in model accuracy (Fig. 8b). The regression coefficient paths were different for different runs, reflecting the changes in the regression coefficients (Fig. 8c). The vertical line represented the minimum RMSECV value, which was 220 runs. The extracted 19 characteristic wavelengths were mainly distributed between 1100–1165 nm and 1200–1400 nm. The selected characteristic wavelengths were little but uniform, thus preserved some of the synergistic effects between wavelengths (Fig. 8d).

The UVE method preserves the information variables, reduces the frequency band dimension, and improves the prediction accuracy of the model. The left side of the vertical line showed 233 wavelength variables under full-spectrum conditions, while the right side showed 117 random variables (Fig. 9a). Two horizontal dashed lines separate useful variables (outside the dashed line) and useless variables (inside the dashed line). The 84 characteristic wavelengths extracted by UVE method were evenly distributed, mostly in the wavelength range, and the combined effect of wavelengths was observed (Fig. 9b).

Prediction models using effective wavelengths

The three models PLSR, LSSVM, and SFCN based on different algorithms to extract feature wavelengths could effectively predict glycine contents in beef samples, and the models of all characteristic wavelengths showed good results (Table 4). In the full-band modeling of beef glycine content, the prediction set performance of PLSR method was 0.0481 higher than that of LSSVM method, indicating that the modeling effect of linear model PLSR was better than that of nonlinear model LSSVM, the linear relationship between glycine content and spectral absorption in beef samples was greater than the nonlinear relationship. The prediction set performance of SFCN method was 0.0423 higher than that of

PLSR method, this indicated that SFCN had more potential in predicting the content of glycine than traditional models.

The number of bands retained by all feature wavelength processing methods ranged from67.18%-92.57%, which could avoid data redundancy and eliminate useless information. The modeling effect of extracting characteristic wavelengths based on different methods was better than that of full-band data modeling. The model after extracting characteristic wavelengths by the VUE method achieved the best results, with R_p^2 being 0.0298 higher than that of the full-band model, indicating that extracting characteristic wavelengths retained useful information and improved model accuracy. Jiangbo et al. studied the hardness of different pear varieties based on Vis-NIR spectroscopy [27], and obtained the optimal UVE-SPA-LSSVM hardness prediction multi-variety model was obtained, and its R_P^2, RMSEP and RPD_P values of 0.94, 0.91, and 2.93, respectively. Our results suggested that the characteristic wavelengths selected by UVE were more accurate and effective. On the contrary, among the three modeling methods, CARS method was slightly worse and related to the less feature band extraction and the elimination of part of the effective information.

Prediction models using combinations of spectral and textural features

The optimal grayscale image had a critical influence on the optimal texture features. Principal component analysis was performed on beef sample images, and the cumulative contribution rate of the first three principal component images was 92.12% (Fig. 10). PC1, PC2 and PC3 explained 89.67%, 3.69% and 0.76% of the image information, respectively, suggesting that the texture information of the first principal component could be used as the representative information of the sample. Therefore, the GLCM method was used to extract the texture information of contrast (CONT), homogeneity (HOM), energy (ENG), and correlation (CORR) in four directions (0°, 45°, 90°, and 135°) in pc1 images. Pearson correlation plot of informative parameters and glycine content, positive in red and negative in blue. The darker the color, the greater the correlation. HOM and ENG were positively, and CONT and CORR were negatively correlated with glycine (GLY) content (Fig. 11). The different correlation coefficients were numerically similar, with the highest correlation between HOM and glycine content, reaching 0.47. Indicating that there was a certain correlation between texture information and glycine content, which was worthy of further study.

PLSR, LSSVM and SFCN models were established based on the fusion of extracted texture information by PC1 image and optimal spectral values (Table. 5). The R_P^2 values of the PLSR and LSSVM prediction models were increased by 0.04 and 0.05 respectively, compared with the optimal wavelength model, indicating that the linear and nonlinear relationship between glycine content and fusion information was enhanced to different degrees. The SFCN model established by the fusion of spectral information and textural information had the best performance (R_P^2=0.9005, RMSEP = 0.3075, RPD_P=0.2688). Compared with the fused data of the other two modeling methods, the R_P^2 was 0.0303 and 0.0666 higher, respectively, indicating that the SFCN method still had better predictive ability in the prediction model established after fusing the textural and spectral information of beef glycine. In summary, the fusion of

textural and spectral information can describe beef quality information more effectively. This maybe because the texture image of beef muscle reflected the arrangement characteristics or degree of density of muscle fibers, which was related to the amino acid content of beef.

At present, it was common for deep learning models to outperform traditional modeling methods. In red meat recognition and identification, the deep 3D-CNN model outperformed the PLS-DA and SVM models, and the highest recognition accuracy reached 98.6% in different spectral systems [28]. Prove the features of CNN outperform HOG features with correct classification rate (CCR) of 0.927 and 0.916 for cross-validation and test data set separately. It showed that CNN was more effective than traditional methods for fish muscle opening detection and ensures product quality. Building a graphical framework for visualizing salmon fillet classification, the best model built from the feature descriptor of CNN had achieved a high CCR of 0.927 for cross validation and 0.916 for the external test data set. It showed that CNN was more effective and suitable for fish muscle detection than traditional methods to ensure product quality [29]. The possible causes were as follows: The performance of SFCN prediction model established by deep learning method was better than that of traditional modeling method; the spectral values measured by different spectral systems may affected the training and fitting of model performance; differences breeds and muscle types may led to different results.

Visualization of glycine content prediction

On the basis of optimal spectral information and textural information fusion modeling, specific model coefficients were applied to obtain the dot product of coefficients and spectra to obtain a single pixel in the image. The predicted value of pixels was obtained by displaying distribution and visualization shows the distribution of glycine content in beef samples (Fig. 12). The color bar showed different concentration values, the concentrations of glycine was between 1.30 g/100 g and 4.97 g/100 g, and the color was not uniform, which may be caused by the uneven distribution of glycine; Some of the mutated blue and red areas may be caused by the failure to remove clean fine fascia and adipose tissue from the sample or by water loss during the experiment. It can be clearly and intuitively expressed the glycine content in different areas of samples after pseudo-color treatment. Compared with traditional prediction methods, this can help consumers to quickly and directly judged the quality of beef without damage. It can be said that these studies were effective and necessary.

Conclusions

In this study, on the basis of pre-processing, the traditional models and the SFCN model were built and their performances were compared. The results showed that the models based on the full spectrum band showed good results, indicating that the prediction model of glycine content in beef samples was feasible and effective. In addition, the UVE algorithm was used to extract 84 effective characteristic wavelengths, and the established UVE-SFCN model was the best in the characteristic band range (R_P^2 = 0.8427), which indicated that removing redundant data and retaining effective bands are positively helpful to the prediction performance of the model. Finally, the spectral data of the optimal characteristic wavelength

was fused with the textural data extracted by GRAM to establish a prediction model. The prediction results of the UVE-SFCN-fusion model were ideal, the model R_P^2 was 0.9005, RMSEP was 0.3075, RPD_P was 0.2688, which were 3.10% and 6.41% higher than the full-band model and the extracted characteristic wavelength, respectively. The fusion of textural and spectral information improved the performance of the model while obtaining better prediction results. The best fusion model was used to generate visual color images of glycine in beef samples, which provided a nondestructive and rapid detection method for the prediction of chemical values in beef. According to the fusion of spectral features and textural information, although the research on meat was slightly lacking in universality, the correlation research on other indicators in meat should be explored to improve the prediction model and provide theoretical basis for hyperspectral research on meat quality monitoring or the establishment of online platform.

Declarations

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

This work was supported by the Ningxia Hui Autonomous Region fund (2022AAC05022) and Key R & D plan of the Autonomous Region (2019BEH03002).

Author information

Authors and Affiliations

School of Food & Wine, Ningxia University, Yinchuan 750021, China

Yu Lv, Fujia Dong, Jiarui Cui, Ruiming Luo, Jie Hao, Sijia Liu

Animal Science, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

Argenis Rodas-Gonzalez

Contributions

All authors contributed to the study conception and design. Material preparation were performed by Yu Lv, Fujia Dong, and Sijia Liu. Data collection and analysis were performed by Yu Lv and Fujia Dong. The first draft of the manuscript was written by Yu Lv and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Yu Lv: Writing – original draft, Writing – review & editing, Conceptualization, Formal analysis, Investigation

Fujia Dong: Data curation, Formal analysis, Investigation

Jiarui Cui: Data curation, Formal analysis, Methodology, Software

Ruiming Luo: Funding acquisition, Resources,

Songlei Wang: Funding acquisition, Project administration, Resources

Argenis Rodas-Gonzalez: Supervision

Jie Hao: Investigation, Visualization, Writing – review & editing

Sijia Liu: Validation, Visualization, Writing – review & editing

Corresponding author

Correspondence to Songlei Wang. (ORCID: 0000-0002-5038-2401) E-mai: wangsonglei163@126.com.

Ethics declarations

Competing interests

The authors declare no competing interests.

Statement of human and animal rights

We hereby declare that this article did not contain any studies with human participants or animals.

References

1. Yda B, As A, Crcb C, Mmra B (2020) A global calibration model for prediction of intramuscular fat and pH in red meat using hyperspectral imaging ScienceDirect. Meat Sci 181:108405. https://doi.org/10.1016/j.meatsci.2020.108405
2. Wang HH, Sun DW (2002) Melting characteristics of cheese: analysis of effect of cheese dimensions using computer vision techniques. J Food Engineering 52:279 – 84. https://doi.org/10.1016/S0260-8774(01)00116-9
3. Jackman P, Sun DW (2008) Prediction of beef eating quality from color, marbling and wavelet texture features. Meat Sci 80(4):1273–81. https://doi.org/10.1016/j.meatsci.2008.06.001
4. Chong SH, Ham S (2014) Site-directed analysis on protein hydrophobicity. J Comput Chem 18:1364–70. https://doi.org/10.1002/jcc.23631

5. Golbraikh A (2014) Modelability criteria: statistical characteristics estimating feasibility to build predictive QSAR models for a dataset. Practical Aspects of Com Chem III 23631:187–230. https://doi.org/10.1002/jcc.23631
6. Kamruzzaman M, Barbin D, Elmasry G, Sun DW, Allen P (2012) Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. Inoov Food Sci Emerg 16:316–25. https://doi.org/10.1016/j.ifset.2012.07.007
7. Lu R, Park B (2008) Hyperspectral and multispectral imaging for food quality and safety. Sensing Instrumentation for Food Quality 3:131–2. https://doi.org/10.1007/s11694-008-9060-2
8. Li WY (2015) Simultaneous quantification of uronic acid, amino sugar, and neutral sugar in the acidic polysaccharides extracted from the roots of angelica sinensis (Oliv) diels by HPLC. Food Anal Method 8:2087–93. https://doi.org/10.1007/s12161-015-0096-8
9. Chen YN, Sun DW, Cheng JH, Gao WH (2016) Recent advances for rapid identification of chemical information of muscle foods by hyperspectral imaging analysis. Food Eng Rev 3:336 – 50. https://doi.org/10.1007/s12393-016-9139-1
10. Girmatsion M, Mahmud A (2021) Rapid detection of antibiotic residues in animal products using surface-enhanced raman spectroscopy: a review. Food Control 126:108019–7. https://doi.org/10.1007/978-1-4899-7445-7_7
11. Morikawa K, Igarashi T, Misumi S, Fukuda T, Ojiri H, Matsudaira H (2019) A case of pseudocystic liver metastases from an atypical lung carcinoid tumor. Radiology Case Reports 14(5):595–601. https://doi.org/10.1016/j.radcr.2019.02.022
12. Kehchiro H (1971) Effect of excess levels of individual amino acids on growth of rats fed casein diets. Laboratory of Food and Nutrition 10:1117–1125. https://doi.org/10.1093/jn/101.9.1117
13. Kucha CT, Liu L (2018) Non-destructive spectroscopic techniques and multivariate analysis for assessment fat quality in pork and pork products: a review. J Sensors 18(2):377. https://doi.org/10.3390/s18020377
14. Liu Y, Lyon D (2003) Prediction of color, texture, and sensory characteristics of beef steaks by visible and near infrared reflectance spectroscopy. Meat Sci 65(3):1107–15. https://doi.org/10.1016/S0309-1740(02)00328-5
15. Govender M, Chetty K (2007) A review of hyperspectral remote sensing and its application in vegetation and water resource studies. Water S A 33:0378–4738. https://doi.org/10.4314/wsa.v33i2.49049
16. Liu H, Ji Z, Liu X, Shi C, Yang X (2020) Non-destructive determination of chemical and microbial spoilage indicators of beef for freshness evaluation using front-face synchronous fluorescence spectroscopy. Food Chem 321(1):0308–8146. https://doi.org/10.1016/j.foodchem.2020;321(1):0308-8146
17. Weng S, Guo B, Tang P, Yin X, Pan F, Zhao Z (2020) Rapid detection of adulteration of minced beef using Vis/NIR reflectance spectroscopy with multivariate methods. Spectrochim Acta A Mol Biomol Spectrosc 230:118005. https://doi.org/10.1016/j.saa.2019.118005

18. Zhang X, Lpe G (2013) Application of hyperspectral remote sensing in field of medicinal plants monitoring research. China journal of Chinese materia medica 9:03. https://doi.org/10.4268/cjcmm20130903
19. Wang C, Wang S (2020) Combination of spectra and texture data of hyperspectral imaging for prediction and visualization of palmitic acid and oleic acid contents in lamb meat. Meat Sci 169:108194. https://doi.org/10.1016/j.meatsci.2020.108194
20. Nassar L, Okwuchi IE, Saad M, Karray F, Agrawal P (2020) Prediction of strawberry yield and farm price utilizing deep learning. IJCNN https://doi.org/10.1109/IJCNN48605.2020.9206998
21. Zhang DY, Chen G, Yin X, Hu RJ, Gu CY, Pan ZG (2020) Integrating spectral and image data to detect Fusarium head blight of wheat. Comput Electron Agr 175:105588. https://doi.org/10.1016/j.compag.2020.105588
22. Xiao Q, Bai X, Gao P, He Y (2020) Application of Convolutional Neural Network-Based Feature Extraction and Data Fusion for Geographical Origin Identification of Radix Astragali by Visible/Short-Wave Near-Infrared and Near Infrared Hyperspectral Imaging. Sensors (Basel) 20(17):4940. https://doi.org/10.3390/s20174940
23. Xiong Z, Sun DW, Xie A, Han Z, Wang L (2015) Potential of hyperspectral imaging for rapid prediction of hydroxyproline content in chicken meat. Food Chem 175:417–22. https://doi.org/10.1016/j.foodchem.2014.11.161
24. Nolasco IM, Badaró AT, Barbon S, Barbon APAC, Pollonio MAR, Barbin DF (2018) Classification of Chicken Parts Using a Portable Near-Infrared (NIR) Spectrophotometer and Machine Learning. Applied Spectroscopy 72:1774–80. https://opg.optica.org/as/abstract.cfm?URI=as-72-12-1774
25. Elmasry G, Sun DW, Allen P (2011) Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging. Food Res. Int 44(9):2624–33. https://doi.org/10.1016/j.foodres.2011.05.001
26. Jijun Z, Benkang C, Zhang Y, Yang Z (2010) Variation of spectral response from cesium-covered. East China Institute of Technology 14:2561–2565. https://doi.org/10.1016/j.foodres.2011.05.001
27. Jiangbo L, Hailiang Z, Baishao Z, Yifei Z, Ruili L (2020) Nondestructive firmness measurement of the multiple cultivars of pears by Vis-NIR spectroscopy coupled with multivariate calibration analysis and MC-UVE-SPA method. Inferared Phys Techn 104:103154
28. Al-Sarayreh M, Reis MM, Wei QY (2020) Potential of deep learning and snapshot hyperspectral imaging for classification of species in meat. Food Control 117:107332. https://doi.org/10.1016/j.foodcont.2020.107332
29. Xu JL, Sun DW (2018) Computer Vision Detection of Salmon Muscle Gaping Using Convolutional Neural Network Features. Food Anal 11:34–47. https://doi.org/10.1007/s12161-017-0957-4

Tables

Table 1 Parameters of feature extraction module

Feature extraction module	Convlution1	Convlution2	Convlution3
Kernel	2*2	2*2	2*2
Stride	1	1	1
Channel	64	128	256
Activation function	ReLu	ReLu	ReLu
Padding	Yes	Yes	Yes

Table 2 Calculation formula of GLCM parameters

Feature parameters	Equation
HOM	$\sum_{i=0}^{X}\sum_{j=0}^{Y}\frac{g(i,j)}{1+(i,j)^2}$
CONT	$\sum_{i=o}^{X}\sum_{j=0}^{Y}(i-j)^2 g(i,j)$
ENG	$\sum_{i=o}^{X}\sum_{j=0}^{Y} g(i,j)$
CORR	$\sum_{i=0}^{X}\sum_{j=0}^{Y}\frac{(i-\mu_i j)(j-\mu_i j)g(i,j)}{\sigma_i\sigma_j}$
μ_i	$\sum_{j=o}^{Y} i\sum_{i=0}^{X} g(i,j)$
μ_j	$\sum_{j=o}^{Y} j\sum_{i=0}^{X} g(i,j)$
σ_i	$\sqrt{\sum_{i=0}^{X}(i-\mu_i)^2\sum_{j=0}^{Y} g(i,j)}$
σ_j	$\sqrt{\sum_{j=0}^{Y}(j-\mu_j)^2\sum_{i=0}^{X} g(i,j)}$

Where X is the column number of GLCM, Y is the row number of GLCM, and G (i, j) is the gray co-occurrence matrix.

Table 3 Descriptive statistics of glycine content in beef samples

Methods	LV_S	Calibration set		Prediction set	
		R_{CV}^2	RMSECV	R_P^2	RMSEP
SPXY	**13**	**0.8059**	**0.3897**	**0.7925**	**0.3887**
KS	12	0.8118	0.4387	0.7592	0.4511
RS	12	0.8072	0.3940	0.7312	0.4343

SPXY, Sample set partitioning based on joint X-Y distance; KS, Kennard-Stone; RS, Random select.

Table 4 Prediction models of glycine content based on different characteristic wavelengths

Model	Method	No. of variables	Calibration set		Prediction set		Cross validation	
			R_{CV}^2	RMSECV	R_P^2	RMSEP	RPD_C	RPD_P
PLSR	RAW	256	0.8776	0.4262	0.8251	0.3836	0.2532	0.2557
	UVE	84	0.8767	0.3838	0.8289	0.3941	0.2325	0.2264
	iVISSA	**46**	**0.8226**	**0.5367**	**0.8302**	**0.3371**	**0.1557**	**0.2478**
	CARS	19	0.7996	0.3893	0.7799	0.3940	0.1492	0.1474
LSSVM	RAW	256	0.7765	0.2777	0.7599	0.2930	0.3996	0.3788
	UVE	**84**	**0.8065**	**0.3780**	**0.7821**	**0.3323**	**0.3943**	**0.3607**
	iVISSA	46	0.7894	0.4997	0.7688	0.5025	0.1672	0.1662
	CARS	19	0.8490	0.2263	0.7521	0.2474	0.2561	0.2343
SFCN	RAW	256	0.8579	0.4366	0.8427	0.2746	0.2539	0.4038
	UVE	**84**	**0.8907**	**0.1425**	**0.8725**	**0.2225**	**0.4487**	**0.4154**
	iVISSA	46	0.8707	0.4521	0.8603	0.4079	0.1849	0.2049
	CARS	19	0.8880	0.3781	0.8202	0.3353	0.1500	0.1692

Table 5 Performance of the models based on optimal spectra and data fusion

Model	Types	Calibration set			Prediction set		
		R_{CV}^2	RMSECV	RPD_C	R_P^2	RMSEP	RPD_P
PLSR	Optimal spectra	0.8226	0.5367	0.3371	0.8302	0.3371	0.2478
	Data fusion	0.8779	0.2309	0.4659	0.8702	0.2341	0.4595
LSSVM	Optimal spectra	0.8065	0.3780	0.3323	0.7821	0.3323	0.3607
	Data fusion	0.8944	0.3553	0.3028	0.8339	0.4510	0.2385
SFCN	Optimal spectra	0.8907	0.1425	0.4487	0.8725	0.2225	0.4154
	Data fusion	**0.9177**	**0.2765**	**0.2990**	**0.9005**	**0.3075**	**0.2688**

Figures

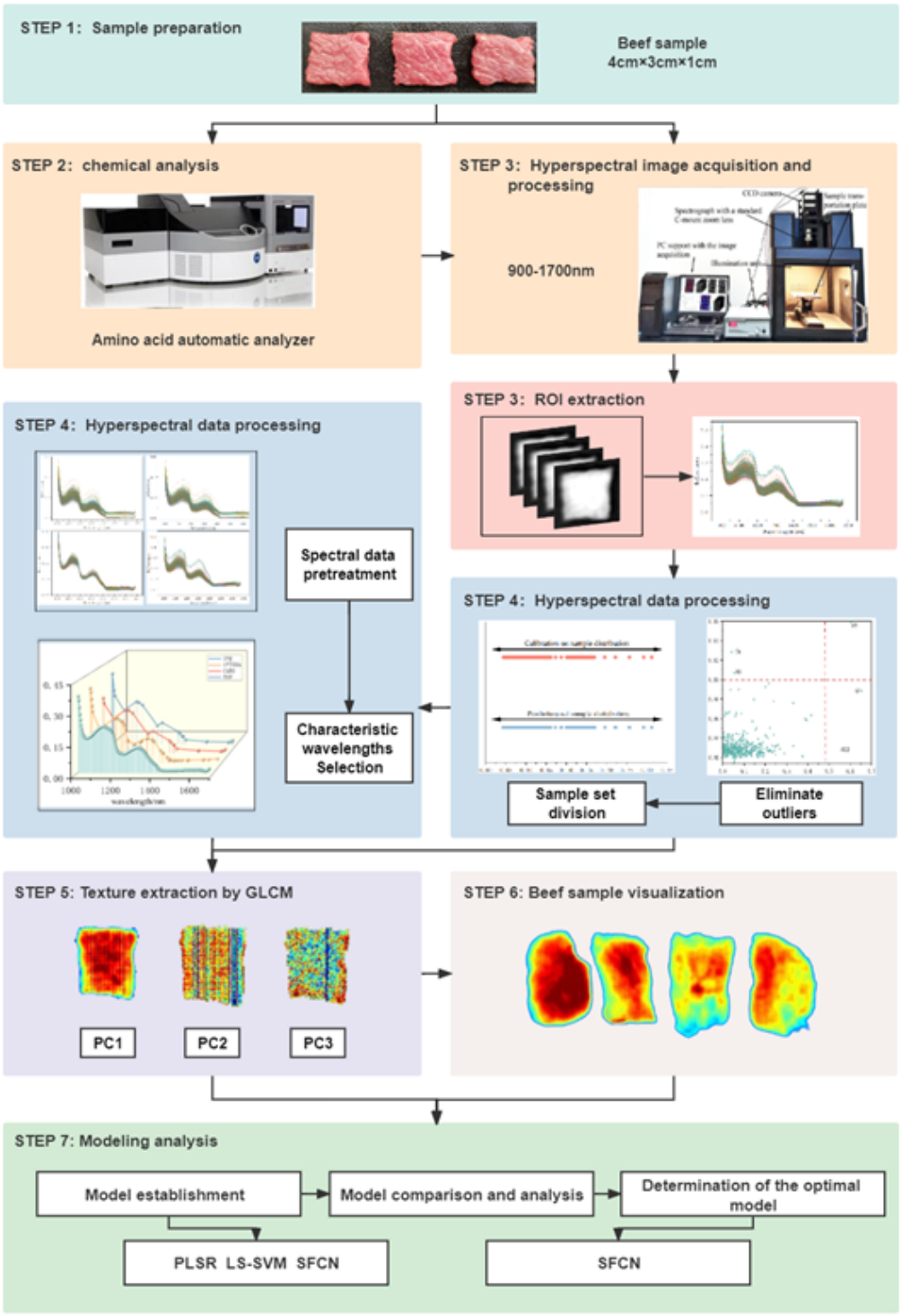

Figure 1

Complete test flow chart

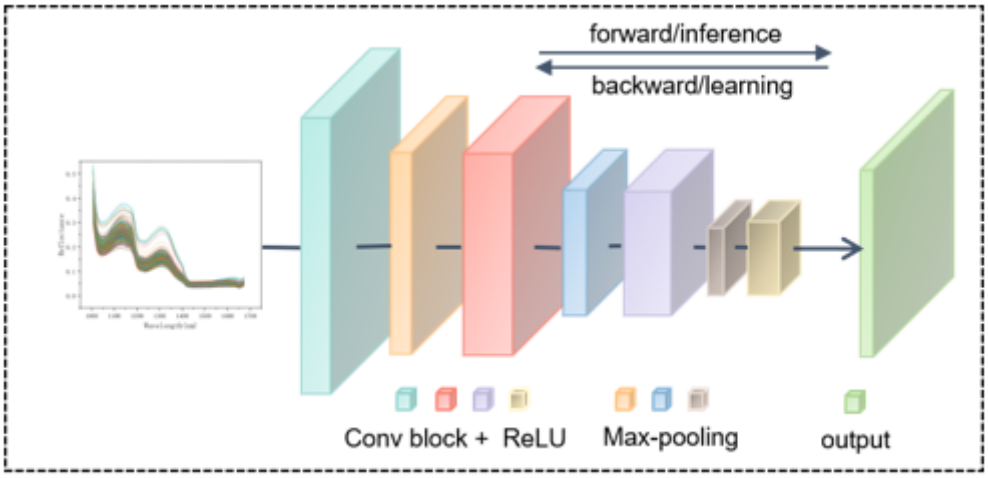

Figure 2

Schematic diagram of SFCN

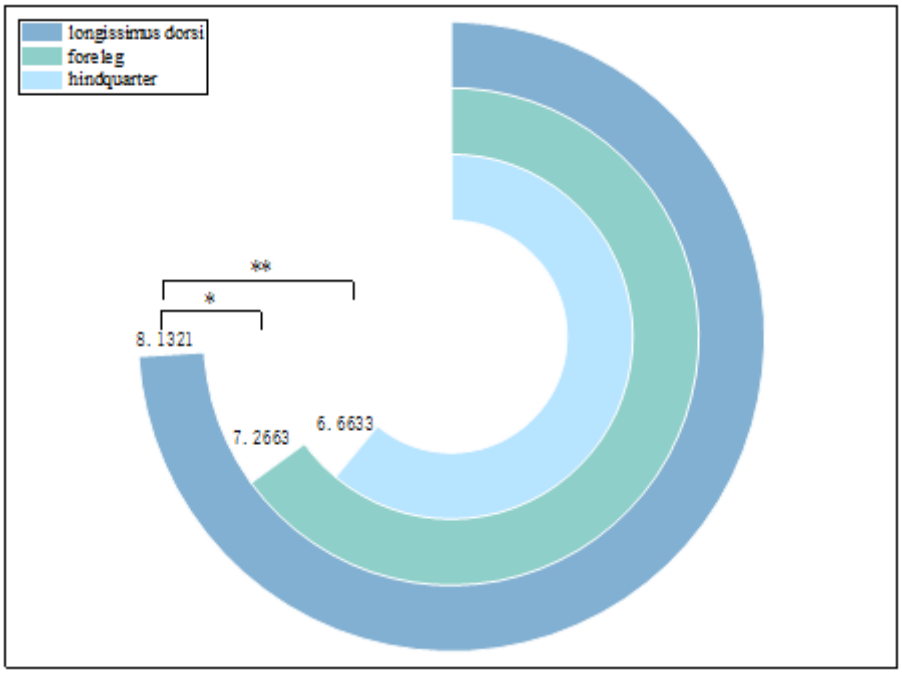

Figure 3

The contents of glycine in in different parts of beef sample

Note: ** represents extremely significant, $p < 0.01$, and * represents significant, $p < 0.05$

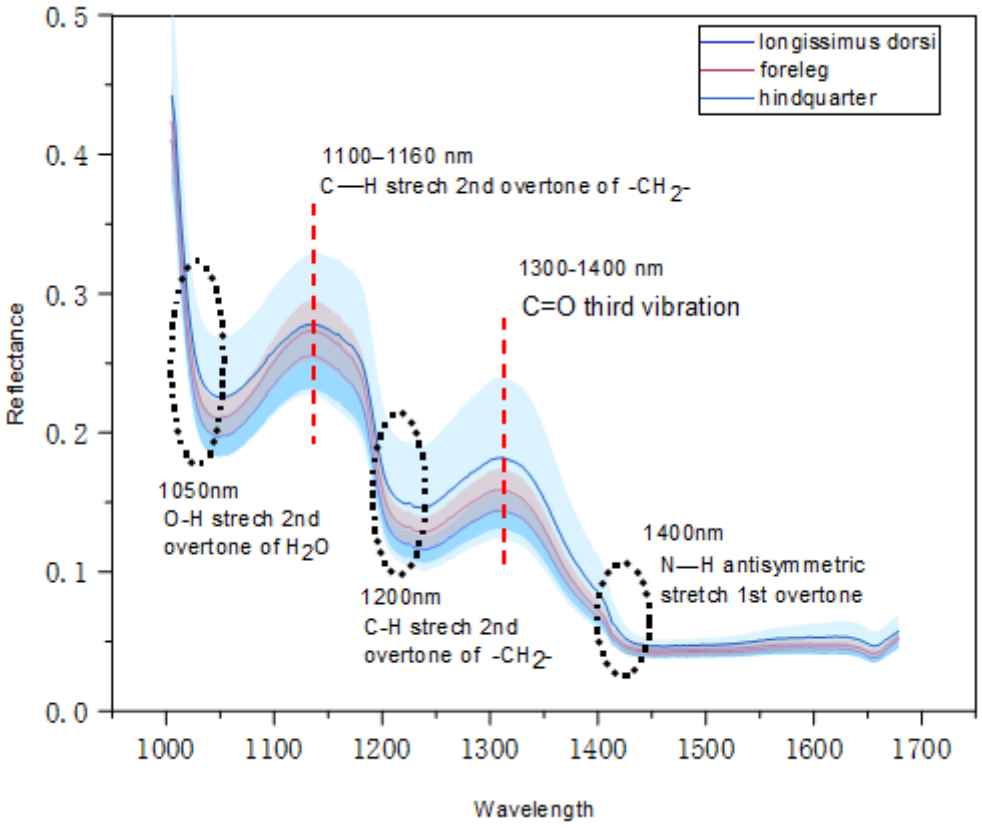

Figure 4

Average spectral curves of beef samples from different parts in the spectral range of 900-1700 nm

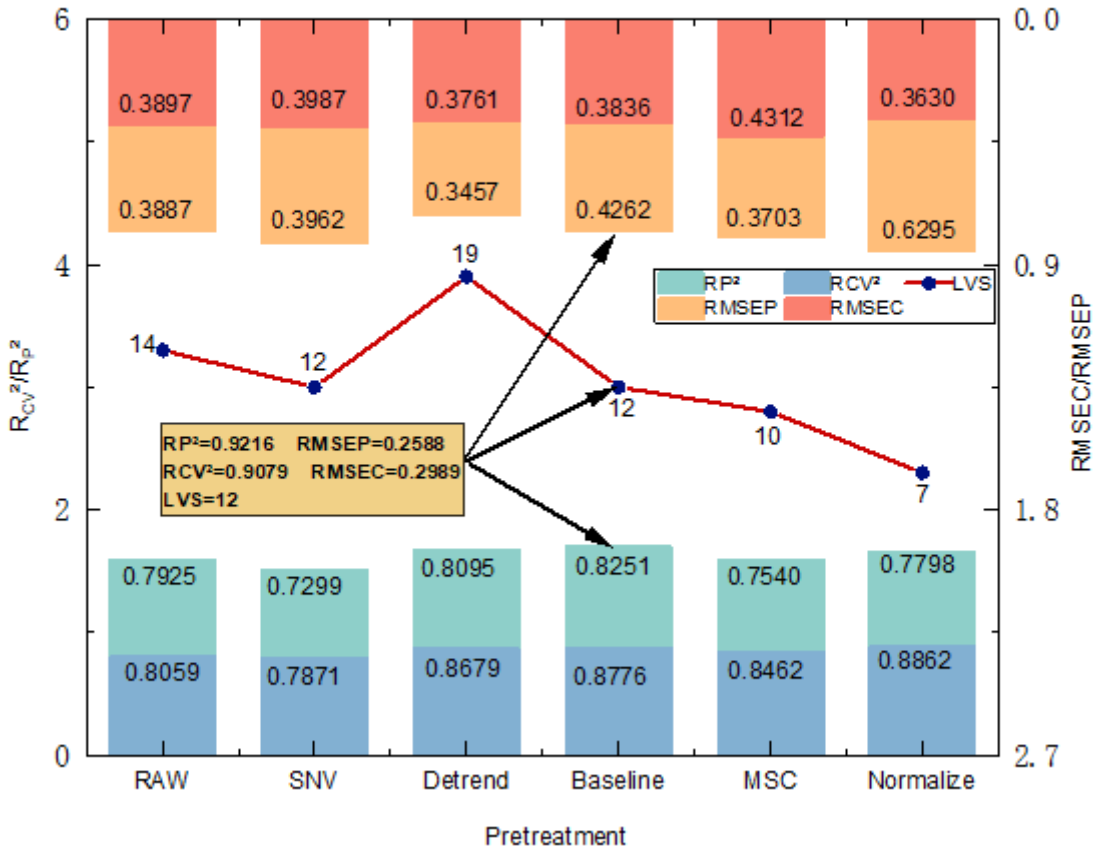

Figure 5

Performance of glycine content prediction models after pretreatment by different methods

Figure 6

Spectral curves of data preprocessed by Baseline method

Figure 7

iVISSA algorithm running diagram (a) iVISSA weight coefficient diagram (b) spectral characteristic wavelength of glycine content extracted by iVISSA method

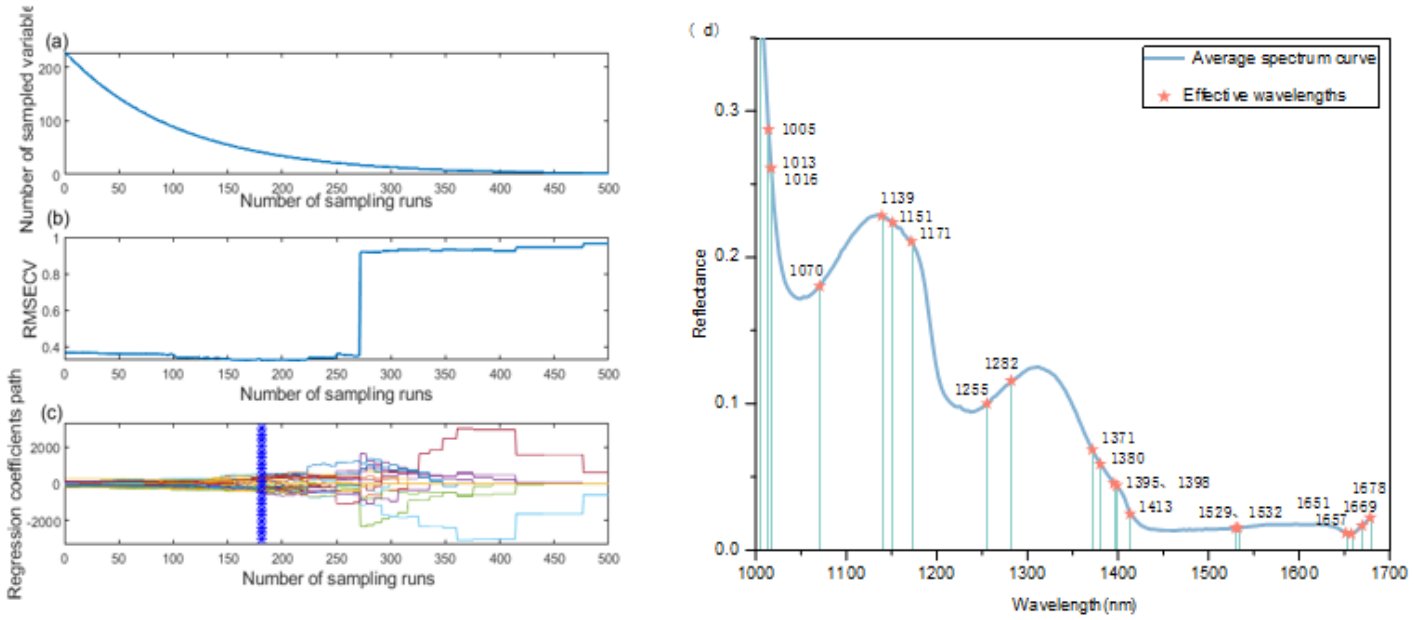

Figure 8

CARS algorithm running diagram (a) Plot of the number of variables versus the number of runs

(b) Optimal number of variable subsets (c) Variation of regression coefficients (d) Characteristic wavelength of glycine content spectrum extracted by CARS method

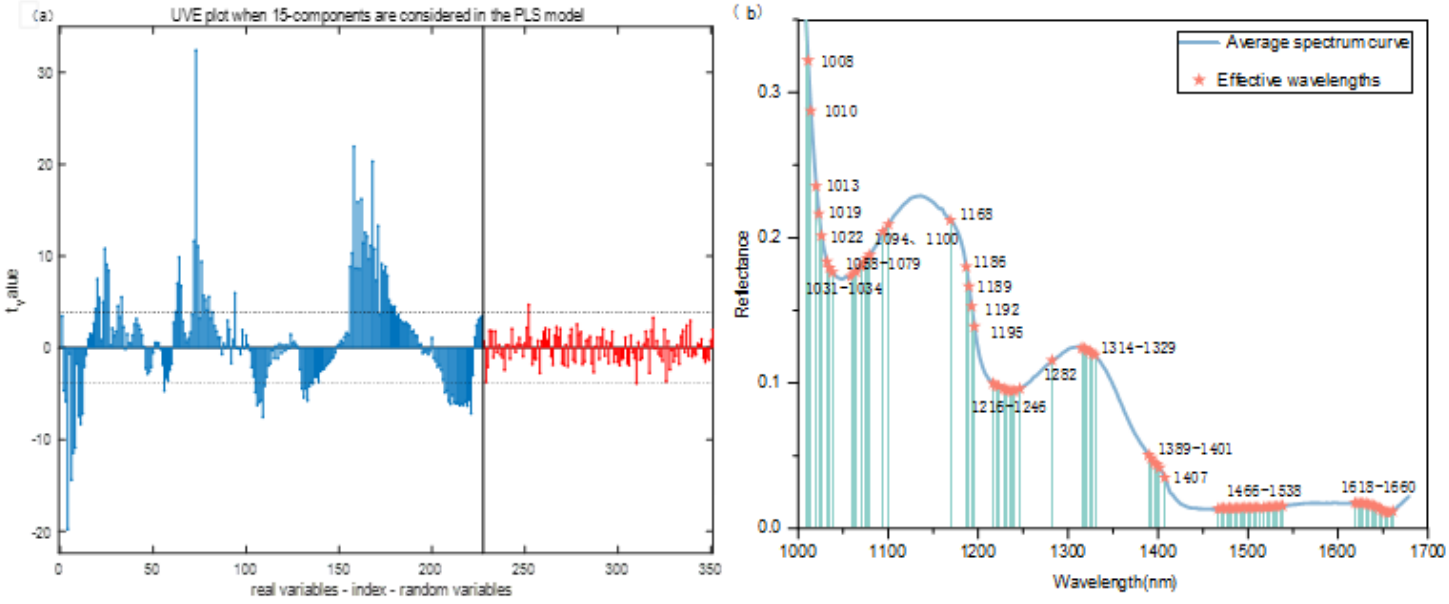

Figure 9

UVE algorithm running diagram (a) Distribution curve of UVE algorithm (b) Characteristic wavelength of glycine content spectrum extracted by UVE method

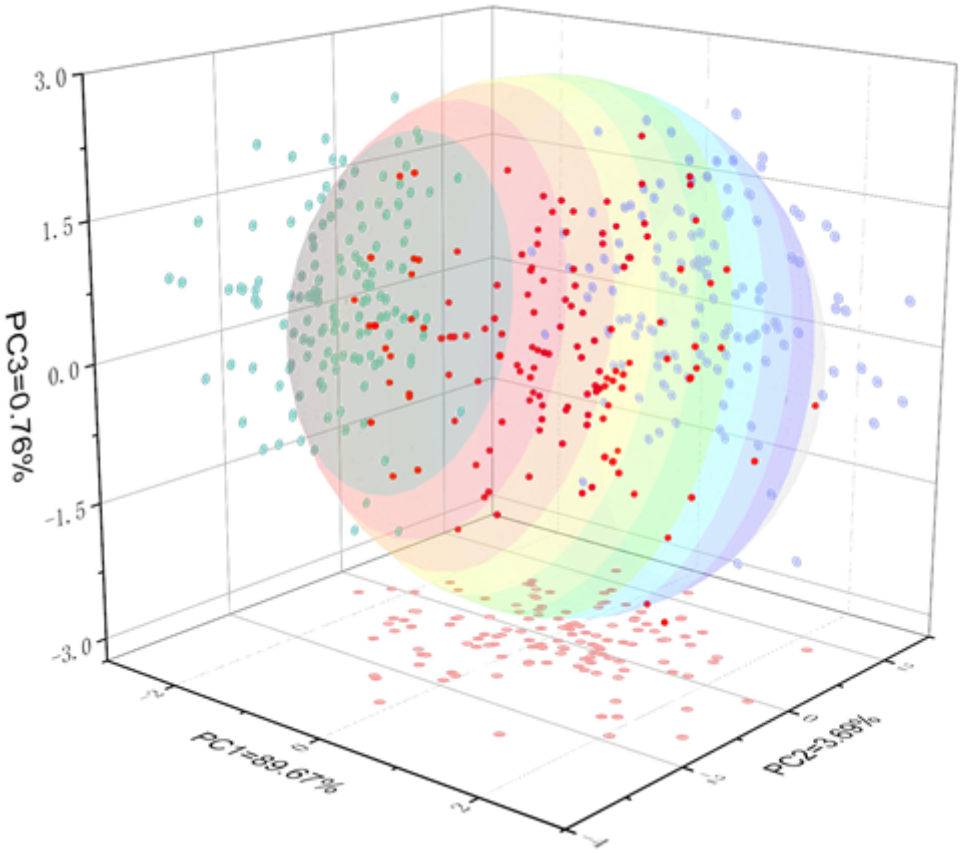

Figure 10

Images of the first three principal components of principal component analysis in beef samples

Figure 11

Correlation of glycine content with texture parameters extracted using GLCM

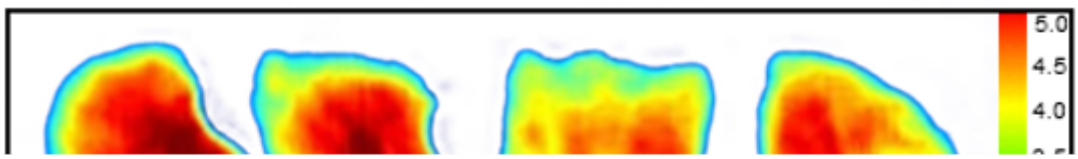

Figure 12

Visualization map of glycine in beef sample