Food Chemistry 352 (2021) 129329

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Rapid and non-destructive spectroscopic method for classifying beef freshness using a deep spectral network fused with myoglobin information

Sungho Shin [a,1], Youngjoo Lee [b,1], Sungchul Kim [b], Seungjun Choi [a], Jae Gwan Kim [b,*], Kyoobin Lee [a,*]

[a] School of Integrated Technology, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea
[b] Department of Biomedical Science & Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea

ARTICLE INFO

Keywords:
Deep spectral network
Diffuse reflectance spectroscopy
Deep learning
Myoglobin
Beef freshness

ABSTRACT

A simple, novel, rapid, and non-destructive spectroscopic method that employs the deep spectral network for beef-freshness classification was developed. The deep-learning-based model classified beef freshness by learning myoglobin information and reflectance spectra over different freshness states. The reflectance spectra (480–920 nm) were measured from 78 beef samples for 17 days, and the datasets were sorted into three freshness classes based on their pH values. Myoglobin information showed statistically significant differences depending on the freshness; consequently, it was utilized as a crucial parameter for classification. The model exhibited improved performance when the reflectance spectra were combined with the myoglobin information. The accuracy of the proposed model improved to 91.9%, whereas that of the single-spectra model was 83.6%. Further, a high value for the area under the receiver operating characteristic curve (0.958) was recorded. This study provides a basis for future studies on the investigation of myoglobin information associated with meat freshness.

1. Introduction

Beef is one of the most widely consumed foods in daily life and contains several nutrients such as proteins, vitamins, and minerals (Dixit et al., 2017; Huang, Zhao, Chen, & Zhang, 2014). Owing to the increasing concern towards beef quality and safety, consumers expect manufacturers and retailers to offer superior-quality products (Dixit et al., 2017). The most important parameter for determining beef quality and safety assessment is freshness, which is closely related to the consumers' intention of purchasing beef (Font-i-Furnols & Guerrero, 2014; Huang et al., 2014). Therefore, to satisfy consumers' expectations and improve the commercial value of beef, the development of practical methods for analyzing beef freshness is crucial (Dixit et al., 2017; Huang et al., 2014; Mustafa & Andreescu, 2018). As the storage period of beef increases, its freshness decreases owing to microbial spoilage and biochemical reactions. Many methods of measuring freshness degradation have been proposed (Cai, Chen, Wan, & Zhao, 2011; Faustman & Cassens, 1990; Tao, Peng, Li, Chao, & Dhakal, 2012). Traditional analytical methods use sensory evaluation, chemical analysis, and microbial population evaluation to assess beef freshness (Huang et al., 2014; Zhang, Tong, Chen, & Lan, 2008). However, implementing these methods is time-consuming and can damage the beef samples during the pre-treatment process. In addition, the accuracies of the measurements and results are highly dependent on the proficiency of the experimenter (Huang et al., 2014; Kuswandi & Nurfawaidi, 2017; Yamamoto & Sonehara, 1953). Hence, the purpose of this study was to develop an advanced method that can be used to analyze beef freshness expeditiously and consistently, overcoming the limitations of traditional methods.

Recently, near-infrared spectroscopy (NIRS) and other near-infrared (NIR)-related technologies (e.g., Fourier-transform (FT) NIRS) have been applied to assess parameters related to meat freshness in an efficient and non-destructive manner (Cai et al., 2011; Chen, Cai, Wan, & Zhao, 2011; Dixit et al., 2017; Liao, Fan, & Cheng, 2010). Because they

* Corresponding authors at: School of Integrated Technology, Gwangju Institute of Science and Technology, Dasan Building 215, 123 Cheomdan-gwagiro, Buk-gu, Gwangju 61005, South Korea (K. Lee); Department of Biomedical Science & Engineering, Gwangju Institute of Science and Technology, Dasan Building 418, Cheomdan-gwagiro, Buk-gu, Gwangju 61005, South Korea (J.G. Kim).
E-mail addresses: hogili89@gm.gist.ac.kr (S. Shin), mass3758@gm.gist.ac.kr (Y. Lee), sungchulkim8972@gmail.com (S. Kim), petercsj@gist.ac.kr (S. Choi), jaekim@gist.ac.kr (J.G. Kim), kyoobinlee@gist.ac.kr (K. Lee).
[1] Sungho Shin and Youngjoo Lee contributed equally to this work.

https://doi.org/10.1016/j.foodchem.2021.129329
Received 3 July 2020; Received in revised form 15 January 2021; Accepted 7 February 2021
Available online 23 February 2021

do not require any pre-treatment process and can be applied to several foods including meat (Cai et al., 2011; Chen et al., 2011; Dixit et al., 2017; Liao et al., 2010; Porep, Kammerer, & Carle, 2015), NIRS and FT-NIR have gained attention as alternative methods of meat-freshness measurement. In particular, FT-NIRS offers high resolution and accuracy with a high signal-to-noise ratio (Agelet & Hurburgh, 2010). However, this approach requires a complex and expensive system for measurement and needs to constantly update the calibration models. Furthermore, both NIRS and FT-NIR systems must be utilized in regulated conditions because the system is sensitive to measurement conditions such as temperature and vibration (Agelet & Hurburgh, 2010; Dixit et al., 2017; Salguero-Chaparro, Baeten, Fernández-Pierna, & Peña-Rodríguez, 2013). Therefore, in this study, we propose the use of diffuse reflectance spectroscopy (DRS) with deep learning to analyze beef freshness using a low-cost and simple system while retaining the advantages of NIRS and FT-NIR.

DRS is an optical technique that quantifies the composition of chromophores by illuminating the sample surface and fitting the obtained diffuse reflectance spectra to an analytical photon diffusion model in a turbid medium (Kim, Kim, & Kim, 2019; Nachabé, Hendriks, van der Voort, Desjardins, & Sterenborg, 2010; Nguyen & Kim, 2019). DRS utilizes light in the visible, NIR, and mid-infrared regions to quantify the composition of chromophores with known absorption coefficients in a simple and efficient manner (Mirabella, 1998). The DRS system does not require any calibration model updates and utilizes an affordable and easily configurable spectrometer rather than a spectrophotometer, which is used in NIRS and FT-NIRS (Agelet & Hurburgh, 2010; Mirabella, 1998; Nguyen, Kim, & Kim, 2019).

In our previous research, we utilized DRS to quantify the proportions of different myoglobin redox forms, which are the most dominant pigments in beef muscle (Nguyen et al., 2019). Myoglobin is a heme protein that delivers oxygen to muscle cells. After a cow is slaughtered, the myoglobin undergoes continuous changes in its form (Hui, Nip, & Rogers, 2001; Mancini & Hunt, 2005; Millar, Moss, & Stevenson, 1996). The various myoglobin redox forms are identified by a gradual discoloration in the beef; the most common forms are oxymyoglobin (oxy-Mb), deoxymyoglobin (deoxy-Mb), and metmyoglobin (met-Mb), associated with the colors cherry red, red-violet, and brown, respectively (Faustman & Cassens, 1990). Supplementary Figs. 2 & 4 depict this process. Generally, deoxy-Mb is predominantly found in beef before slaughter. As the beef is exposed to air after slaughter, deoxy-Mb on the surface of the beef is rapidly converted to oxy-Mb through oxygenation (Nguyen, Phan, Lee, & Kim, 2016; Richards, 2013). This process progresses from the surface to the interior of the beef over several days (Dikeman & Devine, 2014). Simultaneously, oxidation occurs from the interior of the beef, which converts oxy-Mb and deoxy-Mb into met-Mb (Mancini & Hunt, 2005). As the storage period increases, the rate of oxygenation reduces until the oxidation process occurs predominantly. As a result, met-Mb becomes the predominant form, eventually turning the color of the beef to brown (Hui et al., 2001). Further, as the beef spoilage process begins, sulfmyoglobin (sulf-Mb), which has a green color, is sporadically formed by bacteria on the surface of beef (Faustman & Cassens, 1990). Each myoglobin redox form has different absorption coefficients in the visible and NIR regions, and their proportions change during storage (represented graphically in Supplementary Figs. 2 & 3) (Mancini & Hunt, 2005; Nguyen et al., 2019). In other words, the diffuse reflectance spectra of fresh and spoiled beef differ because of the difference in proportions of myoglobin redox forms that arise due to extended storage (Supplementary Fig. 5). Therefore, the differences between the diffuse reflectance spectra of fresh and spoiled beef can enable the classification of beef freshness using DRS.

Deep learning has widely been studied as a powerful algorithm for classification (Esteva et al., 2017; He, Zhang, Ren, & Sun, 2016; Russakovsky et al., 2015; Yoo, Gujrathi, Haider, & Khalvati, 2019). Previous studies have highlighted the superior performances of speed and precision of deep learning compared with the existing analytical methods in many fields such as medical science (Back et al., 2019), signal analysis (Drugman, Stylianou, Kida, & Akamine, 2015; Tekbıyık, Akbunar, Ekti, Görçin, & Kurt, 2020), and food safety and quality management (Liu, He, Cen, & Lu, 2018; Moon et al., 2020; Rodríguez, García, Pardo, Chávez, & Luque-Baena, 2018; Yu, Tang, Wu, & Lu, 2018). Liu et al. and Rodríguez et al. utilized a convolutional neural network (CNN), the most widely used deep learning algorithm, for detecting defects in vegetables and fruits from RGB images. When the training dataset includes normal and defective fruit images, CNN can automatically detect optimal features to classify normal images and defects without any prior knowledge and pre-treatment; the model also exhibits robustness in size and color of fruits if the training data include various colors and sizes. Similarly, the quality of shrimps was measured using CNN with hyperspectral imaging (Yu et al., 2018). A recent report by Moon et al. showed high accuracy in the classification of salmon, tuna, and beef using the spectral data with a CNN based machine learning algorithm (Moon et al., 2020).

Therefore, we employed a deep learning algorithm to classify beef freshness using the spectral data obtained from DRS. In this study, we developed a deep spectral network based on a CNN to fully exploit the advantages of CNN with spectral data. The deep spectral network can automatically detect optimal patterns for classifying beef-freshness using reflectance spectra without any calibration models. Myoglobin information was either fused with the input reflectance spectra during the early fusion stage or was fused with the convolutional layers during the later fusion stage (Figs. 1 & 2). Additionally, the input spectral regions that greatly influenced the classification of beef freshness were visualized using gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2017); this was done both with and without myoglobin information fusion to determine the effect of myoglobin information fusion on the overall classification performance.

2. Materials and methods

2.1. Sample preparation and storage

For the experiment, 78 beef samples collected from the round (n = 40) and sirloin (n = 38) parts of cows were used. The round samples were purchased from a local butcher shop (Gwangju, Korea) on the day of slaughter, and the sirloin samples were purchased from the same place on the day after slaughter. All samples were stored in an ice-filled cooler and transported to the laboratory within 30 min. Immediately after arrival, the samples were sliced into pieces of dimensions of 3 ($\pm$1) × 3($\pm$1) × 2($\pm$0.1) cm (length × width × thickness). The weight of each sample was approximately 20 $\pm$ 5 g. After slicing, each sample was wrapped individually in a low-oxygen-permeability polyethylene sheet and stored in a refrigerator at 0.5 °C.

2.2. Diffuse reflectance spectroscopy system and spectra acquisition

The DRS system consisted of the following main components: a broadband light source (tungsten halogen lamp, HL-2000-HP, Ocean Optics, Delray Beach, FL, USA) and a spectrometer (USB4000, Ocean Optics, Delray Beach, FL, USA) that collected reflectance spectra from the beef samples. The measurement probe in our system comprised two optical fibers (of diameter 200 μm). One fiber was connected to a tungsten halogen lamp, which illuminated the beef samples, and the other was connected to a spectrometer, which collected light within a 470–1148 nm wavelength range with a 100 ms integration time. For the deep spectral network, the spectra from 480 to 920 nm were used because the signals at other wavelength ranges had a low contribution. The distance between the two optical fibers was 2 mm. A 1-mm-thick acrylic plate was used to prevent contamination of the beef samples owing to the measurement probe. Supplementary Fig. 1 displays a schematic diagram of the experimental setup.

The diffuse reflectance spectra were measured once at 10 pm on the day the samples arrived (Day 0). Subsequently, they were measured

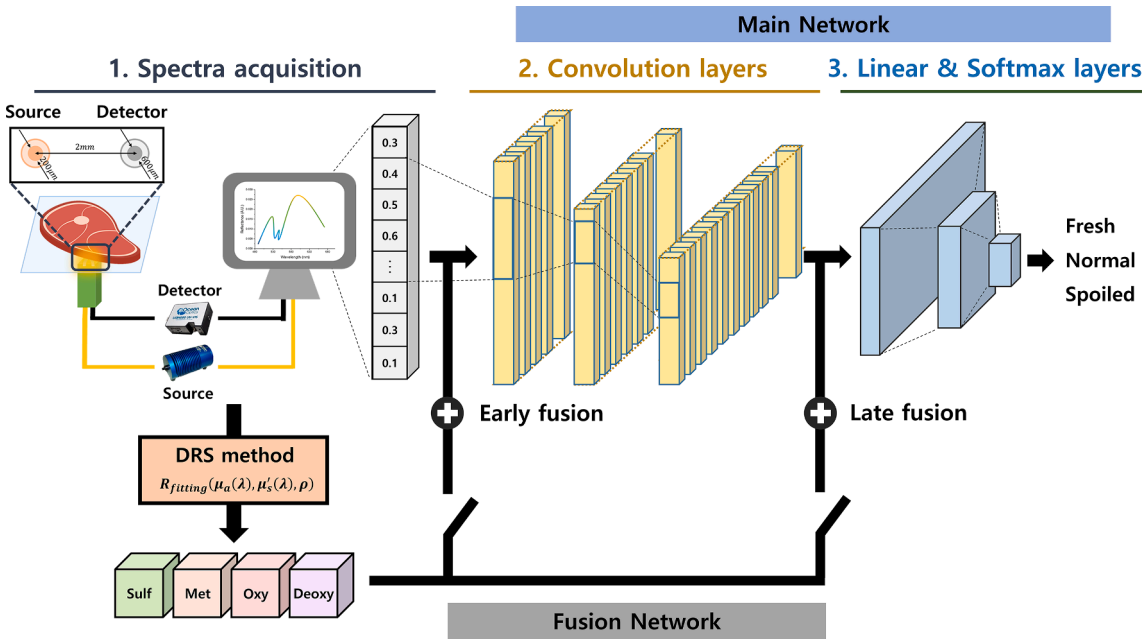

Fig. 1. Schematic diagram of the deep spectral network for beef freshness classification. The network consists of a main network for classification and a fusion network for combining the input reflectance with myoglobin information. In the fusion network, DRS is used to estimate the proportions of the myoglobin redox forms from the input reflectance spectra, and the estimated proportions are combined with the early or later stage of the main network.

every 12 h for 16 days. Accordingly, the measured values added up to a total of 2574 data points ([1 + (2 × 16)] × 78). For each measurement, the probe was covered with black fabric to avoid the influence of ambient light. The reference spectra ($R_{reference}$) were then measured using a 99% reflection plate (SRT-99-050, Labsphere, USA), and the spectra of the beef samples (R_{sample}) were divided by the reference spectra to obtain their respective diffuse reflectance spectra ($R_{measured}$). To minimize the effect of temperature on the measurements, the temperature was maintained at 25 ± 1 °C during collection of the reflectance spectra. All the measurements were performed within a few seconds, and the process of transferring the beef samples from the refrigerator to the lab benchtop using an ice tray was also performed in minutes, thereby minimizing the formation of moisture during measurement. The relationship between R_{sample}, $R_{reference}$, and $R_{measured}$ can be represented as

$$R_{measured} = \frac{R_{sample}}{R_{reference}} \quad (1)$$

2.3. pH measurement and beef freshness labeling

pH measurement, one of the chemical methods for assessing beef freshness, was used in the experiment for freshness labeling (Monin, 1998). It is important to obtain the pH value at the site of spectral measurement for training the deep learning model accurately. Furthermore, the pH values measured using different types of electrodes from the pork and beef samples do not differ significantly (Korkeala, Mäki-Petäys, Alanko, & Sorvettula, 1986). Consequently, a skin pH meter (HI-99181, HANNA instruments, Woonsocket, RI, USA) was used for the measurements in this study. The temperature was maintained at 25 ± 1 °C during the pH measurement, which was the same as that maintained during spectral data collection. The pH meter was washed after every measurement to prevent possible contamination from spreading across the beef samples. The freshness of the samples was labeled according to the pH criteria established by the Food and Agriculture Organization of the United Nations and the Ministry of Food and Drug Safety (Heinz & Hautzinger, 2007; Korea food additives code, Ministry of Food and Drug Safety, 2015). Following these criteria, the measured pH values were classified into three categories: values below 6, between 6 and 6.3, and above 6.3 were labeled "fresh" ($n = 1817$), "normal" ($n = 139$), and "spoiled" ($n = 618$), respectively.

2.4. Sample temperature effects on measurement

Two conditions of sample temperature change were considered: 1) between the frozen state and unfrozen state, and 2) when the temperature increased from 2 °C to 10 °C. For the first experiment, beef samples were stored in the freezer for a day and the DRS measurements were performed immediately after the samples were taken out of the freezer until the samples reached a temperature of 2 °C on the surface. For the second experiment, the beef samples stored in the refrigerator were taken out; then, as the samples reached a temperature of 2 °C, the pH and DRS data were acquired with every 2 °C increase until the temperature of the samples reached 10 °C. The temperature change of the beef samples was observed using an infrared camera (FLIR C3, FLIR Systems, Inc., Sweden), and the locations of the pH and DRS probes on the samples were fixed to minimize the effect of inhomogeneity of beef composition. The number of beef samples used in each experiment was five.

2.5. Quantification of the myoglobin redox forms

The analytical photon diffusion model in a turbid medium, derived by Farrell et al., was used to quantify the chromophore composition (Farrell, Patterson, & Wilson, 1992). This analytical model is related to the absorption and scattering coefficients of a medium and delineates the diffuse reflected light from the illuminated position to the position detected using extrapolation boundary conditions. Its operation can be represented mathematically as follows.

$$R_{\text{fitting}}\left(\mu_a(\lambda), \mu_s'(\lambda), \rho\right) = \frac{\mu_s'}{4\pi(\mu_s' + \mu_a)}\left[z_0\left(\mu_{\text{eff}} + \frac{1}{r_1}\right)\frac{\exp(-\mu_{\text{eff}} r_1)}{r_1^2} + (z_0 + 2z_b)\left(\mu_{\text{eff}} + \frac{1}{r_2}\right)\frac{\exp(-\mu_{\text{eff}} r_2)}{r_2^2}\right] \quad (2)$$

Here, R_{fitting} is the diffuse reflectance spectra derived from the diffusion theory; μ_a is the absorption coefficient; μ_s' is the reduced

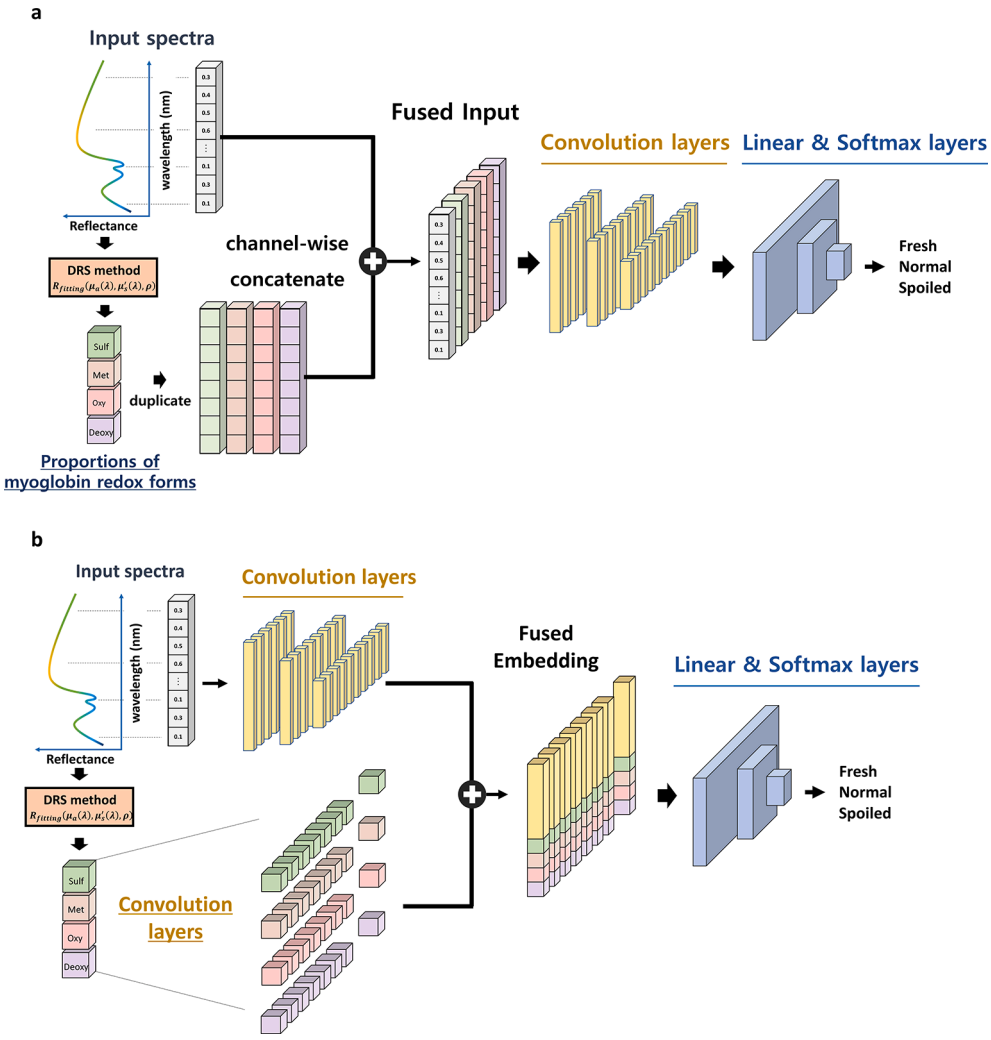

Fig. 2. Two types of fusion networks for beef freshness classification. a. Early fusion combining the estimated proportions of myoglobin redox forms with the input reflectance spectra. b. Late fusion combining the estimated proportions with the embedded reflectance spectra. To match the dimensions with the embedded spectra, the estimated myoglobin proportions are also embedded using the convolution layers.

scattering coefficient; $\mu_{eff} = \left[3\mu_a(\mu_a + \mu_s')\right]^{1/2}$ is the effective attenuation coefficient, and ρ represents the source–detector separation distance. Descriptions of the other parameters (z_0, z_b, r_1, and r_2) can be found in the report by Farrell et al. (1992).

The light absorption and scattering of the sample are described in Eqs. (3) and (4), respectively. In our previous study (Nguyen et al., 2019), only oxy-Mb, deoxy-Mb, and met-Mb were considered as the main chromophores in beef. However, in this study, as the storage period of the beef samples was longer than that in the previous experiment, sulf-Mb was also considered as one of the main chromophores. The absorption coefficients of the sample are expressed as a combination of the proportions (p_{oxy}, p_{deoxy}, p_{met}, and p_{sulf}) and extinction coefficients (ε_{oxy}, ε_{deoxy}, ε_{met}, and ε_{sulf}) of each myoglobin redox form.

$$\mu_a(\lambda) = p_{oxy}\varepsilon_{oxy}(\lambda) + p_{deoxy}\varepsilon_{deoxy}(\lambda) + p_{met}\varepsilon_{met}(\lambda) + p_{sulf}\varepsilon_{sulf}(\lambda) \quad (3)$$

Here, λ is a specific wavelength, and the extinction coefficients of the four chromophores were adopted from previous studies (Nguyen et al., 2016; Zijlstra, Buursma, & Meeuwsen-Van der Roest, 1991). The reduced scattering coefficients of the samples were expressed as combinations of Mie and Rayleigh scattering coefficients (Jacques, 2013; Nachabé et al., 2010).

$$\mu_s'(\lambda) = \alpha\left[\gamma_{MR}\left(\frac{\lambda}{\lambda_0}\right)^{-b} + (1-\gamma_{MR})\left(\frac{\lambda}{\lambda_0}\right)^{-4}\right] \quad (4)$$

Here, α is the scattering amplitude; γ_{MR} is the ratio of Mie scattering to total scattering; λ is the specific wavelength, λ_0 is the normalized wavelength (λ_0 = 500nm), and b is the Mie slope (Bi, Dong, & Lee, 2013).

To quantify the four myoglobin redox forms (i.e., oxy-Mb, deoxy-Mb, met-Mb, and sulf-Mb) using the measured diffuse reflectance spectra ($R_{measured}$), Eqs. (3) and (4) were substituted into Eq. (2). The 480–700 nm spectral range was used for fitting because the main absorbers in beef in this range are oxy-Mb, deoxy-Mb, met-Mb, and sulf-Mb. After the substitution, we obtained the seven unknown variables from the equation: p_{oxy}, p_{deoxy}, p_{met}, p_{sulf}, α, γ_{MR}, and b. Subsequently, non-linear least-squares fitting (lsqcurvefit, Matlab, MathWorks, Natick, MA, USA) was applied to solve the equation. Supplementary Fig. 2 shows the entire process of obtaining the proportions of myoglobin redox forms using DRS.

The process of determining myoglobin information is summarized as follows. First, the reference spectra ($R_{reference}$) and the diffuse reflectance spectra of the beef samples (R_{sample}) were obtained. We then divided R_{sample} by $R_{reference}$ to calculate $R_{measured}$, as expressed in Eq. (1). Finally, the myoglobin information was determined from the absorption

coefficients mentioned in Eq. (3) by fitting $R_{measured}$ and the theoretical value ($R_{fitting}$) in Eq. (2).

2.6. CNN for beef freshness classification

In the CNN, convolutional layers extract the local features of the input spectra by striding the filters, and the extracted features are shared throughout the network as embedded features. Thereafter, linear layers classify the embedded features into the three aforementioned categories (fresh, normal, or spoiled). More specifically, the outputs of the linear layers pass via the softmax layer. The softmax layer converts output features of linear layers into a value between 0 and 1, which indicates the probability that a given datum belongs to a given category (i.e., the confidence score of each category).

$$P(x) = \mathrm{Softmax}(\mathrm{Linear}(\mathrm{CNN}_{main}(x))) \quad (5)$$

Here, x is the input reflectance, where $x \in \mathbb{R}^{1x981}$; $\mathrm{CNN}_{main}(\cdot)$, $\mathrm{Linear}(\cdot)$, and $\mathrm{Softmax}(\cdot)$ represent the functions of the convolutional layers in the main network, the linear layers, and the softmax layer, respectively. $P(x)$ is the probability that a given datum belongs to a given category (fresh, normal, or spoiled).

Our CNN model consists of five convolutional layers, three linear layers with dropout, and a softmax layer. Each layer follows a modified version of AlexNet (Drugman et al., 2015), which is a well-known architecture for image classification. The initial convolutional layer has a large filter size (11–13) to accommodate raw reflectance spectra, which contain a substantial amount of noise. Large filters smoothen the input spectra by averaging the input in the filter range, making our CNN model robust to noise. All models were trained with 70% of the data, and the remainder was used for evaluation (fresh ($n = 546$), normal ($n = 42$), spoiled ($n = 185$)).

The Adam optimizer, with a learning rate of 0.001, was utilized to update the model parameters, and the batch size was set to 1024. Cross entropy loss was utilized for training the deep spectral network, and the synthetic minority oversampling technique (SMOTE) was applied to deal with the unbalanced number of data (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). The SMOTE selected samples of the minority class and calculated a decision boundary based on the selected samples. Thereafter, it synthesized new samples for the minority class at a point within the decision boundary, thereby reducing the imbalance between the minority and majority classes. Each convolutional layer contained a rectified linear unit activation function to address non-linearity in the given data. All deep learning models were implemented using Python 3.7.

2.7. Mechanism of the deep spectral network

The deep spectral network extracts spectral characteristics from spectra through a CNN model composed of convolutional and linear layers, along with one softmax layer (Fig. 1). The CNN model detects optimal patterns for classifying beef freshness from the spectra. If the number of data points is small or the representation of input features is insufficient to identify these patterns, the deep spectral network cannot produce the desired results. In such cases, auxiliary features can be utilized to support the network. In the proposed network, the myoglobin information performed the role of these features. Myoglobin information corresponds to the proportions of myoglobin redox forms, which govern the discoloration associated with beef spoilage. The proportions of myoglobin redox forms were estimated from the reflectance spectra using DRS and were subsequently combined with the CNN network through a fusion network (Fig. 1).

The fusion network has two types: early fusion and late fusion (Fig. 2a & b). Early fusion combines the myoglobin information with reflectance spectra during the early stage of the deep spectral network (Fig. 2a). In Eq. (5), x is replaced with the combined input of early fusion, x_{early}.

$$P(x) = \mathrm{Softmax}(\mathrm{Linear}(\mathrm{CNN}_{main}(x_{early}))) \quad (6)$$

$$x_{early} = x \oplus \mathrm{DRS}(x) \quad (7)$$

Here, x_{early} is the fused input of the reflectance spectra and the estimated myoglobin redox forms, where $x_{early} \in \mathbb{R}^{1x985}$; $\mathrm{DRS}(\cdot)$, $\mathrm{CNN}_{main}(\cdot)$, and $\mathrm{Softmax}(\cdot)$ are functions of DRS that estimate the myoglobin redox forms, the convolutional layers in the main network, and the softmax layer, respectively. $\oplus$ indicates concatenation of two variables.

As mentioned in Section 2.5, the CNN model consists of convolutional and linear layers, along with one softmax layer for embedding the input features and classifying the embedded features. Unlike early fusion, late fusion combines the myoglobin information with the embedded features of the reflectance spectra, not the input spectra. To match the dimensions of the myoglobin proportions to those of the embedded reflectance spectra, additional convolutional layers are applied to embed the myoglobin proportions. These additional layers contain 1×1 convolutional filters, which convert the channel size of the myoglobin proportions into the same dimensions as those of the reflectance spectra. Subsequently, the embedded Mb proportions and reflectance spectra are combined before being fed to the linear layers. In Eq. (5), the output of the convolutional layers, $\mathrm{CNN}_{main}(x)$, is replaced with $\mathrm{CNN}_{late}(x)$ (Fig. 2b).

$$P(x) = \mathrm{Softmax}(\mathrm{Linear}(\mathrm{CNN}_{late}(x))) \quad (8)$$

$$\mathrm{CNN}_{late}(x) = \mathrm{CNN}_{main}(x) \oplus \mathrm{CNN}_{fusion}(\mathrm{DRS}(x)) \quad (9)$$

Here, x is the input reflectance, where $x \in \mathbb{R}^{1x981}$; $\mathrm{DRS}(\cdot)$ is the DRS function for estimating the myoglobin redox forms; $\mathrm{CNN}_{main}(\cdot)$ and $\mathrm{CNN}_{fusion}(\cdot)$ are functions of the convolutional layers in the main network and fusion network, respectively. $\mathrm{Softmax}(\cdot)$ is a function of the softmax layer, and $\mathrm{CNN}_{late}(x)$ represents the combined embedded features. $\oplus$ indicates concatenation of two variables.

2.8. Comparison metrics

The metrics for measuring the classification performances were based on sensitivity and specificity:

$$\mathrm{Sensitivity} = \sum_{i=1}^{n} \frac{1}{n} * \left(\frac{TP_i}{TP_i + FN_i} \right) \quad (10)$$

$$\mathrm{Specificity} = \sum_{i=1}^{n} \frac{1}{n} * \left(\frac{TN_i}{TN_i + FP_i} \right) \quad (11)$$

$$\mathrm{F1score} = \sum_{i=1}^{n} \frac{1}{n} * \left(\frac{2TP_i}{2TP_i + FP_i + FN_i} \right) \quad (12)$$

Here, n indicates the number of freshness classes, and i denotes a number from 1 to 3 indicating the freshness class: 1 indicates fresh, 2 indicates normal, and 3 indicates spoiled. P_i is the number of positive samples, i.e., those that are graded as class I; N_i is the number of negative samples, indicating those graded as class I; TP_i is the number of accurately predicted positive samples; and TN_i is the number of accurately predicted negative samples.FP_i is the number of negative samples predicted as positive inaccurately, and FN_i is the number of positive samples predicted as negative inaccurately. The sensitivity and specificity measured the reliability of our model in generating positive and negative results, respectively.

To evaluate the multi-class classification performance of our network, sensitivity and specificity were combined to measure the overall performance. The most widely used evaluation metrics are the AUC and F1 score. AUC is the area under the sensitivity–specificity curve, which indicates the capability of distinguishing between classes. The F1 score is the harmonic mean of precision and recall, as described

in Eq. (12), which is sensitive to biased results caused by unbalanced data. Both the AUC and F1 score use the macro-averaging method for multi-class classification. We trained each model and performed 10 iterations each, and then measured and compared the classification performances using the above metrics.

2.9. Visualizing highlighted features of the deep spectral network

To realize the effects of the myoglobin fusion on beef freshness classification, the Grad-CAM method was utilized. Grad-CAM visualizes the highly referenced input features when a deep learning model makes a decision (Selvaraju et al., 2017). It computes the gradient of the target class with respect to the given input feature. When the given input features are relevant to the target class, a large amount of gradients flow into the features during backpropagation. By visualizing the amount of gradient flow, Grad-CAM selectively highlights the features useful for classification. In this study, the Grad-CAM model highlighted certain areas of the input spectra with red, which indicated that those areas had a high positive effect on the beef freshness classification. In contrast, the blue areas signified features that were irrelevant to the classification, as shown in Fig. 5. By analyzing the differences in the highlighted areas between the deep spectral network and others, the effects of myoglobin fusion could be explained.

2.10. Statistical test

To analyze the relationship between the myoglobin proportions and beef freshness, we performed an analysis of variance (ANOVA) test to verify whether the proportions of the myoglobin redox forms differed, with statistical significance, depending on the beef freshness. When the *P*-value obtained from ANOVA was lower than the significance level α, which was set to 0.05, a post-hoc (Tukey's honestly significant difference (HSD)) test was performed to demonstrate which freshness state had statistically different proportions of myoglobin redox forms. $P \leq 0.05$, $P \leq 0.01$, and $P \leq 0.001$ were considered statistically significant, highly statistically significant, and very highly statistically significant.

3. Results

3.1. Relationship between beef freshness and proportions of myoglobin redox forms

According to previous research, the proportions of the myoglobin redox forms change with storage time through oxygenation, oxidation, and microbial reactions (Supplementary Fig. 2c) (Dikeman & Devine, 2014; Hui et al., 2001; Mancini & Hunt, 2005; Millar et al., 1996; Nguyen et al., 2016; Richards, 2013). To analyze the relationship between beef freshness and proportions of the myoglobin redox forms, we utilized the ANOVA and post-hoc analysis (Tukey's HSD test). Overall, the average proportion of each myoglobin redox form showed very highly significant differences (ANOVA test; $P \leq 0.001$ for all myoglobin redox forms). Following the ANOVA test, Tukey's HSD test was performed to determine which beef freshness status had significantly different proportions for each myoglobin redox form. The average oxy-Mb proportion of the fresh class (oxy-Mb = 53.35%) was statistically higher than that of the normal (oxy-Mb = 10.00%) and spoiled (oxy-Mb = 8.64%) classes (Tukey HSD test; both $P \leq 0.001$). The proportion of oxy-Mb in the normal class was slightly higher than that in the spoiled class, although it was not significant (Tukey HSD test; $P = 0.777$) (Fig. 3a). The average met-Mb proportions in all the classes were different with statistical significance. When the fresh class decayed into the normal class, the average met-Mb proportions increased from 37.63% to 72.50% (Tukey HSD test; $P \leq 0.001$). The spoilage of beef samples from normal ones decreased the average met-Mb proportions (met-Mb = 28.73%) with very high statistical significance (Tukey HSD test, $P \leq 0.001$) (Fig. 3c).

In contrast, the average proportions of both deoxy-Mb and sulf-Mb increased as the beef decayed. The average deoxy-Mb proportion of the normal class (deoxy-Mb = 17.50%) was significantly higher than that of the fresh class (deoxy-Mb = 9.01%) but significantly lower than that of the spoiled class (deoxy-Mb = 62.10%) (Tukey's HSD test, both $P \leq 0.001$) (Fig. 3b). The average sulf-Mb proportion of the normal class (sulf-Mb = 0.03%) was slightly higher than that of the fresh class (sulf-Mb = 0.007%); however, the difference was not statistically significant (Tukey HSD test, P = 0.701). When the beef was spoiled, the average sulf-Mb proportion (sulf-Mb = 0.54%) increased significantly compared to that of the normal or fresh classes (Tukey HSD test, $P \leq 0.001$) (Fig. 3d).

3.2. CNN-based beef freshness classification

As each myoglobin redox form had a different extinction coefficient, the differences in the reflectance spectra of the beef samples were also found to depend on the freshness. As a baseline, we established a naive Bayes model for beef freshness classification based on the differences in myoglobin proportions. By simply exploiting the different proportions of the myoglobin redox forms (oxy-Mb, deoxy-Mb, met-Mb, and sulf-Mb) the model reached an average area under the receiver operating characteristic curve (AUC) of 0.943 with an accuracy of 84.7% (Table 1). However, although the baseline achieved an acceptable level of classification performance, its accuracy was low and unbalanced for the fresh and spoiled classes, resulting in a low F1 score of 0.707, as shown in Supplementary Fig. 6.

Consequently, a CNN-based deep learning model prevalently used in classification was adopted. This model classified the beef freshness based on the differences in the reflectance spectra. The spectra were used as input, and the model was evaluated under the same conditions as those of the baseline for a fair comparison. The model achieved an average AUC of 0.947 and accuracy of 83.6%, as well as an F1 score of 0.713, which are slightly higher than those of the baseline (Table 1).

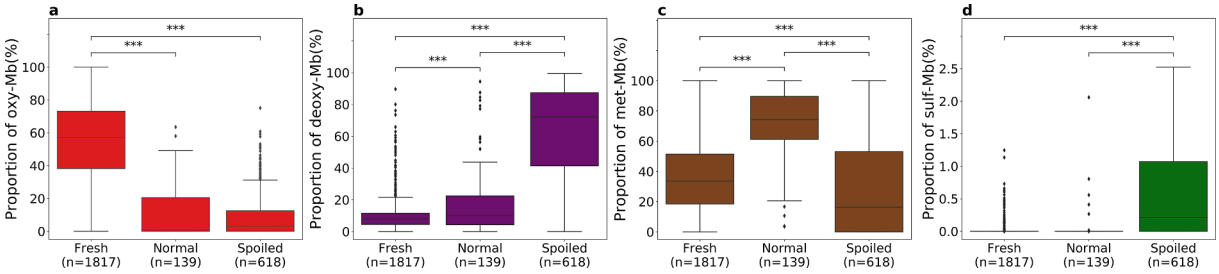

Fig. 3. Change in proportions of the myoglobin redox forms depending on beef freshness status. These plots were derived from a. oxy-Mb, b. deoxy-Mb, c. met-Mb, and d. sulf-Mb. The error bars denote 95% confidence intervals, and the diamond-shaped points indicate the outliers from the error bars. The bottom and top of each box represent the distribution range between the first quartile (25th percentile) and third quartile (75th percentile), respectively. The stars indicate the significance of the post-hoc test between two groups; *** indicates a highly significant difference with a *P*-value of 0.001.

Table 1
Performances of freshness classification models on the evaluation dataset.

Model	Data	ACC (%) (95% CI)	AUC (95% CI)	F1 score (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Naive Bayes	Mb	84.7 (84.1–85.3)	0.943 (0.940–0.946)	0.707 (0.702–0.712)	0.779 (0.772–0.786)	0.924 0.923–0.925)
CNN	Spectra	83.6 (82.9–84.3)	0.947 (0.946–0.948)	0.713 (0.710–0.716)	0.809 (0.802–0.816)	0.934 (0.933–0.935)
Early fusion	Mb + spectra	90.3 (89.7–90.8)	0.954 (0.952–0.956)	0.769 (0.760–0.778)	0.804 (0.796–0.812)	0.948 (0.947–0.949)
Late fusion	Mb + spectra	**91.9 (91.4–92.3)**	**0.958 (0.957–0.959)**	**0.789 (0.782–0.796)**	**0.815 (0.809–0.821)**	**0.954 (0.953–0.955)**

Naive Bayes, naive Bayes model; **CNN**, convolutional neural network; **ACC**, accuracy; **AUC**, area under the receiver operating characteristic curve, which uses a macro-averaging method in this environment; **CI**, confidence interval for 10 iterations of experiments; **Early fusion**, deep spectral network with early fusion; **Late fusion**, deep spectral network with late fusion; **Mb**, proportions of myoglobin redox forms; **Spectra**, reflectance spectra.
* All models were trained with 70% of the dataset and evaluated with the remaining 30%.

3.3. Deep spectral network: fusion of myoglobin information into the CNN

To improve the beef-freshness classification performance, we developed a deep spectral network by combining the CNN with myoglobin information. The proportions of the myoglobin redox forms were quantified by DRS and were fused with the reflectance spectra during the early fusion stage or were fused with the convolutional layers during the later fusion stage (Fig. 2). Both types of deep spectral networks exhibited significantly improved average AUCs and F1 scores. The average AUC of the deep spectral network was 0.954 with early fusion and 0.958 with late fusion (Fig. 4). Moreover, deep spectral networks achieved an increase of 8.8% and 11.6% in the F1 score compared to the baseline models with early and late fusion, respectively. The significant improvements in the F1 score were due to the 5.5% increase in the accuracy of fresh-class classification compared to the baseline (accuracy (ACC) of fresh class in the baseline = 89.9%; ACC of fresh class in late fusion = 94.9%), colored blue in Supplementary Fig. 6. In addition to the increase in the fresh-class accuracy, the average accuracy of the deep spectral network with late fusion also increased by 8.5% compared to the baseline. Moreover, beef freshness could be classified within 1 s in the CPU environment using the deep spectral network.

3.4. Sample temperature effects on the beef freshness classification

Supplementary Fig. 7 summarizes these results, showing no significant difference in the proportions of myoglobin redox forms between the frozen and unfrozen states. When the beef samples were removed from the refrigerator, the proportion of myoglobin redox forms showed only a slight change (<2% in the values at 2 °C until the sample reached 10 °C), indicating that the sample temperature had a minor effect on the measured value. Similarly, it was also found that the pH value change was less than 0.05 from the values at 2 °C until the temperature of the sample reached 10 °C. In addition, the deep spectral network showed consistent classification results with increasing sample temperature, and the confidence score of each result was found to vary slightly, as shown in Supplementary Fig. 8.

3.5. Visualizing the influence of myoglobin fusion: Grad-CAM

To visualize the wavelength regions that considerably affect beef freshness classification before and after fusion, we compared the average Grad-CAM weights of the models. In Fig. 5a, the average Grad-CAM weights of the fresh class showed high values in the 575–600 nm wavelength range before fusion. Upon fusing the myoglobin information into the main network, the wavelength range that contributed to the classification of the fresh class was broadened (Fig. 5d) (490–600 nm and 660–750 nm). Similar effects were observed in the classification of the normal class. For the normal class, the 640–660 nm wavelength range affected the decision of the model more than other wavelengths before fusion. However, after fusion, the normal class also showed relatively high Grad-CAM weights in the wavelength ranges of 490–600 nm and 660–750 nm. In contrast, the wavelength range that considerably affected the classification of the spoiled class remained in the visible region with increasing Grad-CAM weight, whereas the range in the NIR region became narrow with decreasing weight.

4. Discussion

Unlike most existing non-destructive methods of measuring the freshness of meat, our method is simple, efficient, and repeatable and has high accuracy. In general, spectroscopy-based methods (NIRS and FT-NIR) used in meat-freshness measurement focus mainly on functional groups such as O—H, C—H, S—H, and N—H (Chen et al., 2011; Huang et al., 2014); however, our method analyzes freshness based on myoglobin information, rather than on functional bonds. Thus, it was possible to use a more affordable spectrometer than those generally used

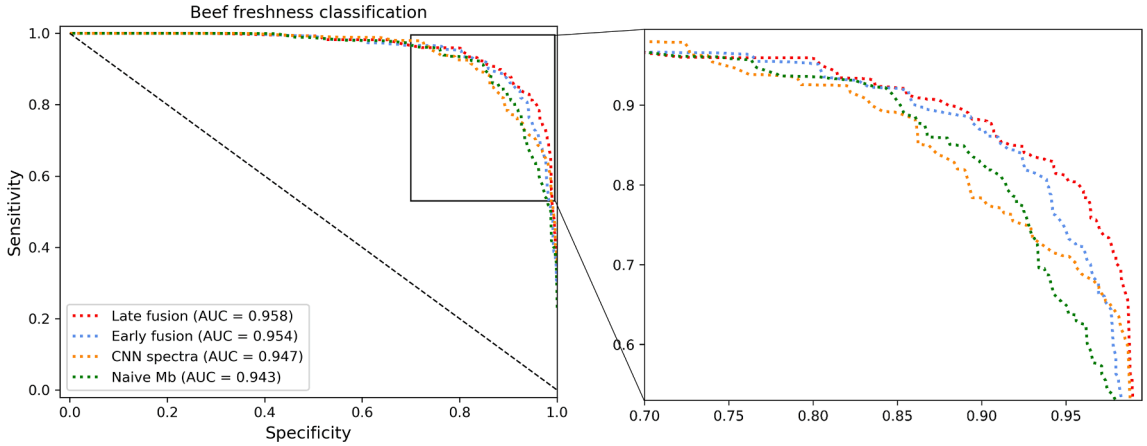

Fig. 4. Sensitivity–specificity curves for freshness classification models. AUC, area under the receiver operating characteristic curve, which uses a macro-averaging method in this experiment; naive Mb, a naive Bayes model trained with myoglobin proportions; CNN spectra, a CNN model trained with the reflectance spectra; early fusion, the deep spectral network with early fusion; late fusion, the deep spectral network with late fusion.

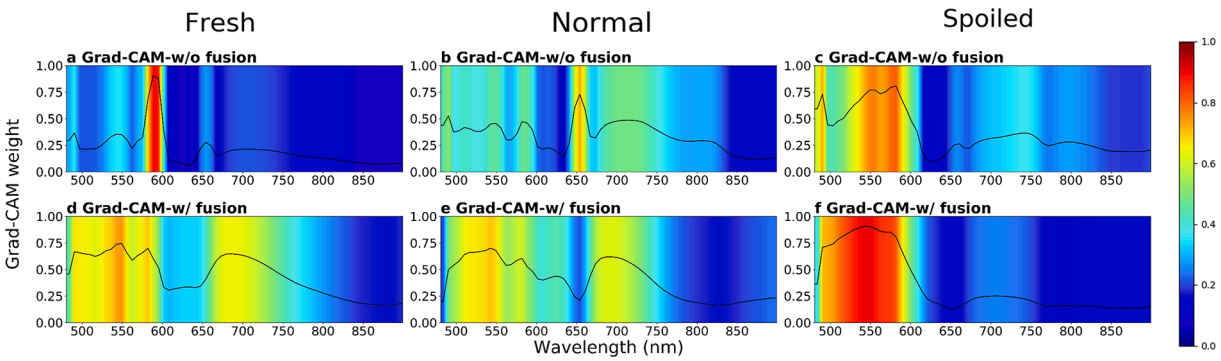

Fig. 5. Average Grad-CAM weights for each freshness class with and without fusion. a, b, and c Average Grad-CAM weights of a CNN model without fusion, and d, e, and f average Grad-CAM weights of the deep spectral network with fusion for fresh, normal, and spoiled classes, respectively. Red indicates that the area significantly affects the model decision, and blue indicates that the area was not referenced for the model decision. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in other spectroscopy-based methods (Agelet & Hurburgh, 2010). Focusing on myoglobin information instead of functional groups allows the use of relatively short wavelength ranges (VIS/NIR, 480–920 nm), unlike previous studies that required relatively long wavelength ranges (NIR, typically 750–2500 nm) to analyze meat freshness (Cai et al., 2011; Chen et al., 2011; Porep et al., 2015; Tøgersen, Arnesen, Nilsen, & Hildrum, 2003). Because water has high absorption coefficients in the wavelength range of 1000–2500 nm (Nachabé et al., 2010; Tøgersen et al., 2003; Wang, Peng, Sun, Zheng, & Wei, 2018), the moisture formed on the surface of a meat sample or the high humidity of the measurement environment makes analysis difficult. Unlike laboratory environments in which conditions are controlled for experiments, it is common for moisture to form on meat surfaces in industrial environments so that existing methods which use long wavelengths have limitations in industrial applications.

A few studies also focused on meat quality analysis by utilizing visible and near-infrared (VIS/NIR) wavelengths of light, and most of them predicted the value related to freshness by developing a regression model (Andersen, Borggaard, Rasmussen, & Houmøller, 1999; Liao, Fan, & Cheng, 2012; Savenije, Geesink, Van der Palen, & Hemke, 2006). Liao et al. utilized visible and near-infrared (350–1100 nm) reflectance to predict the quality attributes of fresh pork by developing a partial least-squares regression (PLSR) model (Liao et al., 2010). In their study, the first derivative (1st order) and multiplicative scatter correction (MSC) were used for preprocessing to eliminate baseline offset and separate the multiplicative interferences. They predicted pH values from the reflectance spectra using a regression model, while our study classified freshness with a deep learning classification model, and thus their results cannot be compared quantitatively with ours. However, our model recorded a very high accuracy of 94.9% for the pH range used in the validation set of Liao et al.'s study which was 5.06–5.98 (fresh class in our study; Supplementary Fig. 6d). In 2014, Reis & Rosenvold also predicted the pH value of beef samples with the PLSR models (VIS/NIR, 350–2500 nm) and classified beef carcasses into two classes (normal: pH $\leq$ 5.8; high: pH $\geq$ 5.8) based on the predicted pH (Reis & Rosenvold, 2014). The highest classification accuracy was 89.5% from the PLSR model, which is lower than the accuracy of this study (93.9%, fresh: pH $<$ 6, non-fresh: pH$\geq$6; Supplementary Fig. 6d). Even though Reis & Rosenvold used a wider range of wavelengths (350–2500 nm) than those used in this study (480–920 nm), a deep spectral network model fused with myoglobin information produced higher accuracy than the model used by Reis & Rosenvold.

Recently, Moon et al. reported that beef freshness can be classified by using spectral data obtained from a portable spectrometer with a CNN-based machine learning algorithm (Moon et al., 2020). VIS/NIR spectra (400–1000 nm) were obtained from beef samples (8 for the training set and another 8 for the verification set) incubated for 30 h at 25 °C and showed 92% overall accuracy of beef freshness classification (fresh, likely spoiled, and spoiled based on pH values). Compared to Moon et al.'s study, our study differs as follows. Firstly, they accelerated the process of spoilage by incubating the beef at 25 °C, while in this study, beef samples were stored in a refrigerator at 0.5 °C and spectra were acquired every 12 h for 17 days, which represents a more realistic consumer environment. Secondly, they obtained a reflectance spectrum and a pH value from a different beef sample because they clamped each beef sample with a spectrometer so that pH had to be measured from another three beef samples. The pH and reflectance spectrum can vary widely even in the same beef sample because of the inhomogeneous structure of beef. Therefore, in our study, the reflectance spectrum and pH were measured from the same position in the same sample. Lastly, Moon et al. selectively used spectral data of the "Likely Spoiled" class for the training and verification of their CNN-based model. During 30 h of the experiment, the "Likely Spoiled" state started from 2 to 3 h after incubating beef samples and lasted until the pH increased to 6.2 (~23 h). However, only the spectral data between 8 and 15 h, which have small variations in pH values, were utilized to train and verify their CNN-based model, while all spectral data at each pH value were utilized in our study. The CNN model without fusion used in this study has the same architectures as that of Moon et al. When the model was validated with our data, it showed far lower performances without fusion (ACC = 83.6%, CNN model with Spectral Data at Table 1) than the deep spectral network (ACC = 91.9%) even though the reflectance spectra were measured with a more sensitive spectrometer than the ones in the study of Moon et al. It implies that our model performs better over ranges in which the meat freshness changes, which are harder to classify and were not included in the study of Moon et al. Therefore, the applicability of our results is judged to be high because we conducted experiments on a large number of beef samples (78 vs 16) while maintaining an environment similar to the beef storage at home and in food markets.

Model robustness and uniformity are significant challenges that need to be improved in spectroscopy-based methods. In particular, the sample temperature is one of the environmental factors that affects model robustness. The effects of sample temperature on the NIR spectra (1100–2500 nm) were reported by Tøgersen et al. (2003). In their study, most areas of the spectra (1100–2500 nm) were found to be influenced by changes in sample temperature. This implies that if there is a difference between the temperature of the sample used to develop a calibration model and the temperature of the measured sample, inaccurate results will be derived. To avoid inaccurate results due to sample temperature differences, it is necessary to selectively use a wavelength that is relatively insensitive to temperature using a filter or to re-develop a calibration model (Tøgersen et al., 2003; Wang et al., 2018). In this study, however, the use of relatively short wavelength ranges (VIS/NIR, 480–920 nm) minimized the effects of temperature on the classification

results, and the confidence score was maintained, as shown in Supplementary Fig. 8. The confidence score is the probability that determines the reliability of the results of the deep spectral network. Small changes in input data can lead to a significant difference in confidence score although the classification results would remain the same. However, there was no significant difference in the confidence score during the change of temperature in the results (<3% from the values at 0 °C and < 0.01% from the values at 2 °C, shown in Supplementary Fig. 8a & b, respectively); this proves the robustness of our method against changing temperature.

Existing spectroscopy-based methods require a calibration model that is specialized for a specific spectrometer; consequently, they are often incompatible with other spectrometers due to a lack of uniformity between instruments. Rebuilding and re-verifying the calibration model is time-consuming and laborious work and increases the difficulty in the application (Agelet & Hurburgh, 2010; Wang et al., 2018). However, because the DRS-based deep learning method used in this study utilizes the spectra obtained by dividing the reflectance spectra of the sample by the previously acquired reference spectra as input data, it can offset differences between spectrometers. The reference spectra can be acquired simply using an integrating sphere or a standard reflection plate; therefore, any spectrometer can be used in our method as long as it has a reasonable sensitivity in the wavelength range of 480–920 nm.

Deep learning is like a black box, and how the decision is made during the process of classification remains unknown. Therefore, explainable artificial intelligence gets more attention from researchers. In this study, Grad-CAM analysis was performed to identify the wavelength regions that highly affected the decision making of the model before and after fusion with myoglobin information. Without the fusion of myoglobin information, both the fresh and normal classes showed narrow wavelength ranges that contributed to the freshness classification, whereas the spoiled class had a relatively wide wavelength range (520–600 nm) showing a high relevance to the classification (Fig. 5a-c). In contrast, the fusion of myoglobin information significantly broadened the wavelength range, contributing to the decisions of the model. All the freshness classes showed relatively high Grad-CAM weights in the wavelength range 490–600 nm, as shown in Fig. 5d-f, which includes the noticeable bands of deoxy-Mb and oxy-Mb reported in previous studies (Supplementary Fig. 3) (Krzywicki, 1979; Liao et al., 2010; Millar et al., 1996; Suman & Joseph, 2013). These results support the possibility that myoglobin information can be used as an indicator of meat freshness and demonstrate that it contributed decisively to freshness classification. Interestingly, sulf-Mb, which has the highest absorption peak at approximately 625 nm, appeared to barely affect the classification performance, although its level was significantly higher in the spoiled class than those in the fresh and normal classes (Fig. 3d). It was mainly due to the small proportion of sulf-Mb in beef compared to those of other Mb redox forms; therefore, the Grad-CAM weights at 625 nm were also low in the spoiled class, both with and without myoglobin information fusion. From the biochemical aspects, sulf-Mb is not always formed when the beef freshness state changes from normal to spoiled (Nicol, Shaw, & Ledward, 1970), which could be another reason that the sulf-Mb did not play a significant role in the classification.

Although the proposed method has advantages over other conventional methods, it has a limitation in the dataset. To achieve high performance in the classification, it is necessary to utilize a large volume of balanced data as input for the deep learning network. However, the sample size of the normal class in our study ($n = 139$) is much smaller than those of other classes owing to the fast conversion from fresh ($n = 1817$) to spoiled status ($n = 618$). Because the amount of training data is unbalanced, the results of the deep neural network can be biased towards the majority class. To prevent biased learning, SMOTE was applied during the training process, after which, the classification accuracy of normal samples was significantly improved compared to the value without SMOTE application, and biased learning could be prevented, as shown in Supplementary Fig. 9. In most food freshness studies including studies on meat, obtaining a balanced dataset is labor-intensive and time-consuming; thus, applying SMOTE to studies that use deep learning is expected to yield unbiased results.

5. Conclusion

A DRS-based deep spectral network was developed in this study for accurate, efficient, and non-destructive classification of beef freshness. The network classified the freshness of beef by utilizing both diffuse reflectance spectra and myoglobin information, and its validity was demonstrated with data obtained using 78 beef samples measured over 17 days of observation, showing an accuracy of 91.9%. Unlike other spectroscopy-based methods, our method includes myoglobin information; thus, it was robust to environmental factors such as sample temperature and humidity, which is a great advantage for industrial applications. With statistical analysis, it was confirmed that the myoglobin information showed a significant difference depending on the freshness of beef; further, it was proven through Grad-CAM that this information contributed decisively to the improvement of model performance. From this study, we demonstrate that a simple spectrometer based DRS system can provide high accuracy in the classification of beef freshness by employing a deep learning algorithm, which confirms the great potential for artificial intelligence in food safety applications.

Author contributions

S.S. established the architecture of the deep spectral network by designing the myoglobin fusion network and trained the network for classifying beef freshness with a fast and simple method. S.C. helped summarize the results. Y.L. suggested the idea of myoglobin fusion and designed the experimental setup. S.K. and Y.L. utilized the diffuse optical theory and codes for the quantification of myoglobin redox forms. K.L. and J.K. directed and supervised the project. All authors discussed the results and contributed to the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the GIST Research Institute (GRI) grant funded by the GIST (Gwangju Institute of Science and Technology) in 2019 and 2020, and the Basic Science Research Program through the National Research Foundation of Korea (Grant 2018R1A2B6006797). This work was also supported by the Technology Innovation Program (20005096, Development of intelligent meal assistant robot with easy installation for the elderly and disabled) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2021.129329.

References

Agelet, L. E., & Hurburgh, C. R., Jr (2010). A tutorial on near infrared spectroscopy and its calibration. *Critical Reviews in Analytical Chemistry, 40*(4), 246–260.

Andersen, J. R., Borggaard, C., Rasmussen, A. J., & Houmøller, L. P. (1999). Optical measurements of pH in meat. *Meat Science, 53*(2), 135–141.

Back, S., Lee, S., Seo, H., Park, D., Kim, T., & Lee, K. (2019). Intra-and Inter-epoch Temporal Context Network (IITNet) for Automatic Sleep Stage Scoring. arXiv preprint arXiv:1902.06562.

Bi, R., Dong, J., & Lee, K. (2013). Deep tissue flowmetry based on diffuse speckle contrast analysis. *Optics Letters, 38*(9), 1401–1403.

Cai, J., Chen, Q., Wan, X., & Zhao, J. (2011). Determination of total volatile basic nitrogen (TVB-N) content and Warner-Bratzler shear force (WBSF) in pork using Fourier transform near infrared (FT-NIR) spectroscopy. *Food Chemistry, 126*(3), 1354–1360.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

Chen, Q., Cai, J., Wan, X., & Zhao, J. (2011). Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy. *LWT-Food Science and Technology, 44*(10), 2053–2058.

Dikeman, M., & Devine, C. (2014). Encyclopedia of meat sciences: 3-volume set: Academic Press, London.

Dixit, Y., Casado-Gavalda, M. P., Cama-Moncunill, R., Cama-Moncunill, X., Markiewicz-Keszycka, M., Cullen, P., & Sullivan, C. (2017). Developments and challenges in online NIR spectroscopy for meat processing. *Comprehensive Reviews in Food Science and Food Safety, 16*(6), 1172–1187.

Drugman, T., Stylianou, Y., Kida, Y., & Akamine, M. (2015). Voice activity detection: Merging source and filter-based information. *IEEE Signal Processing Letters, 23*(2), 252–256.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118.

Farrell, T. J., Patterson, M. S., & Wilson, B. (1992). A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo. *Medical Physics, 19*(4), 879–888.

Faustman, C., & Cassens, R. (1990). The biochemical basis for discoloration in fresh meat: a review. *Journal of Muscle Foods, 1*(3), 217–243.

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

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778).

Heinz, G., & Hautzinger, P. (2007). *Meat processing technology for small-to medium-scale producers*. Bangkok: FAO Regional Office for Asia and the Pacific.

Huang, L., Zhao, J., Chen, Q., & Zhang, Y. (2014). Nondestructive measurement of total volatile basic nitrogen (TVB-N) in pork meat by integrating near infrared spectroscopy, computer vision and electronic nose techniques. *Food Chemistry, 145*, 228–236.

Hui, Y. H., Nip, W.-K., & Rogers, R. (2001). *Meat science and applications*. CRC Press.

Jacques, S. L. (2013). Optical properties of biological tissues: a review. *Physics in Medicine & Biology, 58*(11), R37.

Kim, S., Kim, M., & Kim, J. G. (2019). Development of simple diffuse optical metabolic spectroscopy for tissue metabolism measurement. *Biomedical Optics Express, 10*(6), 2956–2966.

Korkeala, H., Mäki-Petäys, O., Alanko, T., & Sorvettula, O. (1986). Determination of pH in meat. *Meat Science, 18*(2), 121–132.

Krzywicki, K. (1979). Assessment of relative content of myoglobin, oxymyoglobin and metmyoglobin at the surface of beef. *Meat Science, 3*(1), 1–10.

Kuswandi, B., & Nurfawaidi, A. (2017). On-package dual sensors label based on pH indicators for real-time monitoring of beef freshness. *Food Control, 82*, 91–100.

Liao, Y.-T., Fan, Y.-X., & Cheng, F. (2010). On-line prediction of fresh pork quality using visible/near-infrared reflectance spectroscopy. *Meat Science, 86*(4), 901–907.

Liao, Y., Fan, Y., & Cheng, F. (2012). On-line prediction of pH values in fresh pork using visible/near-infrared spectroscopy with wavelet de-noising and variable selection methods. *Journal of Food Engineering, 109*(4), 668–675.

Liu, Z., He, Y., Cen, H., & Lu, R. (2018). Deep feature representation with stacked sparse auto-encoder and convolutional neural network for hyperspectral imaging-based detection of cucumber defects. *Transactions of the ASABE, 61*(2), 425–436.

Mancini, R., & Hunt, M. (2005). Current research in meat color. *Meat Science, 71*(1), 100–121.

Millar, S., Moss, B., & Stevenson, M. (1996). Some observations on the absorption spectra of various myoglobin derivatives found in meat. *Meat Science, 42*(3), 277–288.

Ministry of Food and Drug Safety. (2015). Korea Food Additives Code.

Mirabella, F. M. (1998). *Modern techniques in applied molecular spectroscopy*. John Wiley & Sons.

Monin, G. (1998). Recent methods for predicting quality of whole meat. *Meat Science, 49*, S231–S243.

Moon, E. J., Kim, Y., Xu, Y., Na, Y., Giaccia, A. J., & Lee, J. H. (2020). Evaluation of salmon, tuna, and beef freshness using a portable spectrometer. *Sensors, 20*(15), 4299.

Mustafa, F., & Andreescu, S. (2018). Chemical and biological sensors for food-quality monitoring and smart packaging. *Foods, 7*(10).

Nachabé, R., Hendriks, B. H., Desjardins, A. E., Van Der Voort, M., Van Der Mark, M. B., & Sterenborg, H. J. (2010). Estimation of lipid and water concentrations in scattering media with diffuse optical spectroscopy from 900 to 1600 nm. *Journal of Biomedical Optics, 15*(3).

Nachabé, R., Hendriks, B. H., van der Voort, M., Desjardins, A. E., & Sterenborg, H. J. (2010). Estimation of biological chromophores using diffuse optical spectroscopy: Benefit of extending the UV-VIS wavelength range to include 1000 to 1600 nm. *Biomedical Optics Express, 1*(5), 1432–1442.

Nguyen, T., & Kim, J. G. (2019). A simple but quantitative method for non-destructive monitoring of myoglobin redox forms inside the meat. *Journal of Food Science and Technology, 56*(12), 5354–5361.

Nguyen, T., Kim, S., & Kim, J. G. (2019). Diffuse reflectance spectroscopy to quantify the met-myoglobin proportion and meat oxygenation inside of pork and beef. *Food Chemistry, 275*, 369–376.

Nguyen, T., Phan, K. N., Lee, J.-B., & Kim, J. G. (2016). Met-myoglobin formation, accumulation, degradation, and myoglobin oxygenation monitoring based on multiwavelength attenuance measurement in porcine meat. *Journal of Biomedical Optics, 21*(5).

Nicol, D., Shaw, M., & Ledward, D. (1970). Hydrogen sulfide production by bacteria and sulfmyoglobin formation in prepacked chilled beef. *Applied Microbiology, 19*(6), 937–939.

Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology, 46*(2), 211–230.

Reis, M. M., & Rosenvold, K. (2014). Early on-line classification of beef carcasses based on ultimate pH by near infrared spectroscopy. *MeatS, 96*(2), 862–869.

Richards, M. P. (2013). Redox reactions of myoglobin. *Antioxidants & Redox Signaling, 18*(17), 2342–2351.

Rodríguez, F. J., García, A., Pardo, P. J., Chávez, F., & Luque-Baena, R. M. (2018). Study and classification of plum varieties using image analysis and deep learning techniques. *Progress in Artificial Intelligence, 7*(2), 119–127.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Bernstein, M. (2015). Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252.

Salguero-Chaparro, L., Baeten, V., Fernández-Pierna, J. A., & Peña-Rodríguez, F. (2013). Near infrared spectroscopy (NIRS) for on-line determination of quality parameters in intact olives. *Food Chemistry, 139*(1-4), 1121–1126.

Savenije, B., Geesink, G., Van der Palen, J., & Hemke, G. (2006). Prediction of pork quality using visible/near-infrared reflectance spectroscopy. *Meat Science, 73*(1), 181–184.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (pp. 618-626).

Suman, S. P., & Joseph, P. (2013). Myoglobin chemistry and meat color. *Annual Review of Food Science and Technology, 4*, 79–99.

Tao, F., Peng, Y., Li, Y., Chao, K., & Dhakal, S. (2012). Simultaneous determination of tenderness and Escherichia coli contamination of pork using hyperspectral scattering technique. *Meat Science, 90*(3), 851–857.

Tekbıyık, K., Akbunar, Ö., Ekti, A. R., Görçin, A., & Kurt, G. K. (2020). Real-World Considerations for Deep Learning in Wireless Signal Identification Based on Spectral Correlation Function. arXiv preprint arXiv:2003.08359.

Tøgersen, G., Arnesen, J., Nilsen, B., & Hildrum, K. (2003). On-line prediction of chemical composition of semi-frozen ground beef by non-invasive NIR spectroscopy. *Meat Science, 63*(4), 515–523.

Wang, W., Peng, Y., Sun, H., Zheng, X., & Wei, W. (2018). Spectral detection techniques for non-destructively monitoring the quality, safety, and classification of fresh red meat. *Food Analytical Methods, 11*(10), 2707–2730.

Yamamoto, M., & Sonehara, M. (1953). An assay method for freshness of fishes by the estimation of pH value. *Bulletin of the Japanese Society for the Science of Fish, 9*, 761–765.

Yoo, S., Gujrathi, I., Haider, M. A., & Khalvati, F. (2019). prostate cancer Detection using Deep convolutional neural networks. *Scientific Reports, 9*.

Yu, X., Tang, L., Wu, X., & Lu, H. (2018). Nondestructive freshness discriminating of shrimp using visible/near-infrared hyperspectral imaging technique and deep learning algorithm. *Food Analytical Methods, 11*(3), 768–780.

Zhang, Z., Tong, J., Chen, D.-H., & Lan, Y.-B. (2008). Electronic nose with an air sensor matrix for detecting beef freshness. *Journal of Bionic Engineering, 5*(1), 67–73.

Zijlstra, W., Buursma, A., & Meeuwsen-Van der Roest, W. (1991). Absorption spectra of human fetal and adult oxyhemoglobin, de-oxyhemoglobin, carboxyhemoglobin, and methemoglobin. *Clinical Chemistry, 37*(9), 1633–1638.