Heliyon 9 (2023) e17976

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Analysis of beef quality according to color changes using computer vision and white-box machine learning techniques

Claudia N. Sánchez [a], María Teresa Orvañanos-Guerrero [a], Julieta Domínguez-Soberanes [b,*], Yenizey M. Álvarez-Cisneros [c]

[a] Universidad Panamericana. Facultad de Ingeniería. Aguascalientes, 20296, Mexico
[b] Universidad Panamericana. Escuela de Dirección de Negocios Alimentarios. Aguascalientes, 20296, Mexico
[c] Departamento de Biotecnología, Universidad Autónoma Metropolitana, Unidad Iztapalapa. Ciudad de México, 09310, Mexico

ARTICLE INFO

Keywords:
Beef color
Meat quality
Computer vision system
White-box machine learning

ABSTRACT

The quality of beef products relies on the presence of a cherry red color, as any deviation toward brownish tones indicates a loss in quality. Existing studies typically analyze individual color channels separately, establishing acceptable ranges. In contrast, our proposed approach involves conducting a multivariate analysis of beef color changes using white-box machine learning techniques. Our proposal encompasses three phases. (1) We employed a Computer Vision System (CVS) to capture the color of beef pieces, implementing a color correction pre-processing step within a specially designed cabin. (2) We examined the differences among three color spaces (RGB, HSV, and CIELab*) (3) We evaluated the performance of three white-box classifiers (decision tree, logistic regression, and multivariate normal distributions) for predicting color in both fresh and non-fresh beef. These models demonstrated high accuracy and enabled a comprehensive understanding of the prediction process. Our results affirm that conducting a multivariate analysis yields superior beef color prediction outcomes compared to the conventional practice of analyzing each channel independently.

1. Introduction

The value of meat is given by its sensory quality. The industry has reliable tools, such as evaluation by sensory and instrumental methods, that allows for prediction, characterization, and control of the quality of meat. Therefore, the first challenge is to identify the quality used by consumers to evaluate meat. These quality cues are usually meat color, cut, fat, marbling, amount of drip, and texture [1]. Meat color is the most important sensory attribute of producers and consumers for determining food quality because appearance is almost the only parameter consumers can use to judge their quality [2]. The color of the meat can depend on different factors: pH, characteristics of the muscle surface, feeding systems, conditions, and period of storage [3]. Color depends on myoglobin (Mb), a sarcoplasmic hemo protein considered one of the main proteins of meat. It possesses a monomeric structure conformed by prohyrin and iron. Prophyrin is a highly colored group [4]. The redox states of this protein in fresh meat are three: deoxymyoglobin (DMb), oxymyoglobin (MbO2), and metmyoglobin (MMb) [5,6]. (1) When the molecule does not carry oxygen, it exhibits a blue-purple color, and it is known as deoxymyoglobin (DMb) [7,8]. (2) When myoglobin is oxygenated, it is converted into oxymyoglobin (MbO2), and meat has a bright cherry color. Consumers consider it the preferred state of meat that associates this visual sensory characteristic with its

* Corresponding author. Josemaría Escrivá de Balaguer 101, Villas Bonaterra, zip code 20296, Aguascalientes, Aguascalientes, Mexico.
E-mail address: jdominguez@up.edu.mx (J. Domínguez-Soberanes).

https://doi.org/10.1016/j.heliyon.2023.e17976
Received 7 November 2022; Received in revised form 26 June 2023; Accepted 4 July 2023
Available online 15 July 2023

quality. (3) Finally, when the oxymyoglobin has been subjected to prolonged exposure to air, and in consequence, excessive oxidation has taken place, the pigment is transformed into metmyoglobin (MMb), and therefore it exhibits a brownish color [5–9]. Consumers can visually observe all these chemical changes in myoglobin. When they purchase meat, they judge it based on their vision, evaluating quality according to its references to meat quality. Consumers prefer bright red coloration in beef, mainly associated with oxymyoglobin. Based on this visual sensory analysis, consumers will make their purchase decision [9,10].

Meat color has been measured instrumentally based on spectroscopy and colorimetry principles. Some instruments used are Minolta chroma meter, Hunter Lab colorimeter, and Dr. Lange colorimeters. However, all colorimeters have the disadvantage that the surface analyzed must be small and uniform, not considering the whole piece of meat that should be analyzed. Moreover, if many replicates are performed, a significant deviation could be obtained because of this distortion [6,9,10]. In the last years, new technologies for the evaluation of color in food products have emerged, in particular: computer vision systems or computer image analysis, which has proved to be an objective, efficient, fast, non-destructive, cost-effective, and capable of obtaining digital images that can explain color variation in all the digitalized surface of the meat product. Data can be transformed in numerous measurement systems [8–12,13]. Some authors have used Computer Vision Systems (CVS) to analyze color in pork [3,11,14–16], in beef [12,13,17,18], in chicken [9,19], and, in fish [20]. These studies reflect that CVS technology can be a powerful tool to predict meat quality in a laboratory, which would greatly interest the food industry. In addition, Tomasević et al. established that CVS color was more accurate and precise for color traits compared with a traditional colorimeter [6].

There are several methods for color representation in digital images. The color space that has been more widely used in food science is CIELab* [3,6,10,18,21], which is an international standard adopted by the Commission Internationale de l'Eclairagees, where L* refers to the lightness component (from black to white), a* is a redness parameter (from green if negative to red if positive) and b* is a yellowness parameter (from blue if negative to yellow if positive). However, recent documents [22,23] have used other color spaces, such as RGB (red, green, and blue) and HSV (hue, saturation, brightness value). This document analyzes and compares these three color spaces CIELab*, RGB, and HSV.

Most of the color analysis of beef has been performed by univariate analysis of color channels [24–27]. And some documents define ranges of color for the acceptability of beef products. Holman et al. established that beef color was considered acceptable when a* ≥ 14.5 [28]. Realini et al. defined a fresh beef steak when the color values are L*>39.5, a*>16.8, and b*>6.3 [25]. Zhang et al. defined the range of fresh beef in the ranges of L*>31.4, a*>16.4, and b*>6.5 [26]. Wang et al. established that acceptable frozen beef rolls have a*>16.4 [27]. They all established that a piece of beef with acceptable quality must have an a* value greater than 14.5 or 16.8. However, all of them analyzed the color channels separately. We consider that color analysis must be multivariate instead of univariate. Meaning that we should consider the values of two or more color channels. Some documents have proposed machine-learning

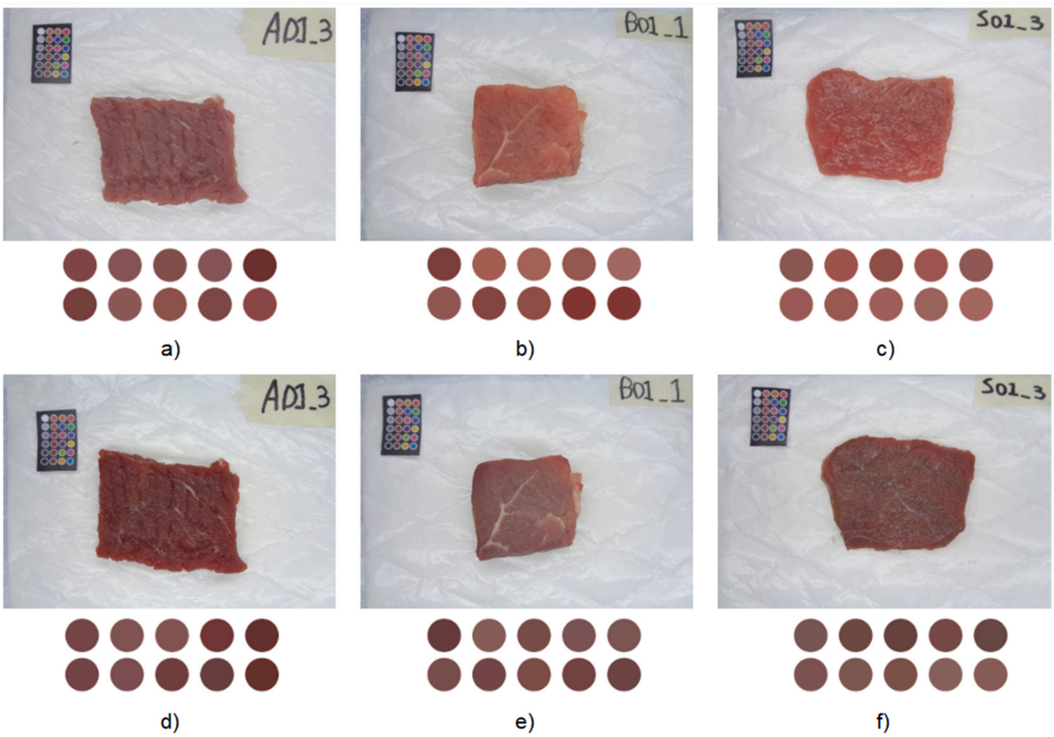

Fig. 1. Beef samples. a) and d) Inside skirt. b) and e) Knuckle. c) and f) Sirloin. a), b), and c) Images corresponding to the first day. d), e), and f) Images corresponding to the fifth day. The circles represent the median color of each beef sample. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

algorithms for meat quality prediction [22,23,29]. In all cases, the prediction results are highly accurate. Moreover, the analysis of how the prediction is made is so complex.

This research aims to comprehensively analyze beef quality by considering color changes through a multivariate approach. The objective is to utilize white-box machine learning algorithms to establish clear guidelines for determining the acceptability of beef colors. The study evaluates various beef cuts, namely inside skirt, knuckle, and sirloin. A Computer Vision System (CVS) is employed to accurately capture the color of the meat, with a comparative analysis conducted across three color spaces: CIELab*, RGB, and HSV. The primary goal is to develop a machine-learning model that automatically identifies the color threshold that distinguishes fresh and non-fresh beef samples. The models are trained using beef color samples taken on the first and fifth days following purchase. Providing the model with these reference color samples makes it proficient in differentiating and classifying the quality according to color between fresh and non-fresh samples. This approach enables a robust multivariate analysis of color, employing three white-box classifiers: Decision Tree, Logistic Regression, and Multivariate Normal Distributions. These classifiers not only yield high accuracy rates but also facilitate a comprehensive understanding of the prediction process, empowering researchers to gain insights into how the models make their predictions.

2. Materials and methods

This section describes the number and type of meat samples used for this research. We describe the CVS to obtain the color through image processing. Finally, this section presents the techniques applied to analyze beef color.

2.1. Samples of meat

The beef was bought from a local market in Aguascalientes, México. The temperature at which it arrived was 2 to 4C°. When they arrived, the pieces were cut. Thirty pieces of beef were used for this analysis, each one of 5 cm by 5 cm by 1 cm. Each cut was put in a black tray that contained a towel. For which, ten pieces of the following cuts: a) inside skirt, b) knuckles, and c) sirloin. After the meat was cut, it was kept refrigerated at 2° to 4 °C and covered with plastic, trying to simulate the natural consumption conditions. From these samples, we obtained color samples on the first and the fifth day after purchase (see Fig. 1(a–f)) using a CVS. The meat was taken out of the fridge, and image capture was performed immediately. All the images were taken around 2 p.m. to have a constant schedule and measure the changes in days.

In total, we had 60 images, 30 from the first day (beef fresh) and 30 from the fifth day (beef non-fresh). In addition to meat (see Fig. 1(a–f)), images include the sample code (up-right) and the color checker matrix for color correction (explained in Subsection 2.2).

2.2. Obtaining beef color through a CVS

To capture the images of beef samples, we designed a cabin. The meat was placed in trays and put into the cabin. Also, we incorporated a color checker matrix (see Fig. 1(a–f) and 3(a)) to color-correct images. Finally, we segmented the image and get the median color of the lean meat.

a) b)

Fig. 2. Computer Vision System technologies. a) Cabin. b) Color checker matrix. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2.2.1. Cabin

The cabin was designed and constructed aiming to acquire images under the same light conditions. The cabin size was decided based on the measurement of the meat containers. It has a part on the top that keeps the camera in the same place. All cabin parts are washable and waterproof to allow proper hygiene. It is 3 mm thick, black acrylic. Fig. 2(a) shows the cabin that was used in this research. Its purpose is to isolate the outside light and to control the illumination. Following the proposal of Tomasevic et al. [2], we proportionate the next information about our image capture device:

- Camera: It is a Sony DSC W830, compensating the exposure brightness to +1.0 in all cases and using the autofocus function to take the pictures.
- Illumination: The cabin has two LED light tubes that keep the interior light of the booth constant at 640 lm. The temperature color is 6500K daylight white LED lights to ensure their lighting conditions are similar to daytime.
- Observation Angle: More than 10. It allows us to capture the full piece of meat. It has a mechanism that keeps the camera position and meat samples constant.

2.2.2. Color correction of images

Color checker matrices can be used to calibrate colors in images correcting differences in colors by using different devices or illumination changes. In this research, we used the color checker matrix shown in Fig. 2(b). Columns 1, 3, and 4 represent the colors of a standard color checker matrix [30]. The original color checker matrix has a column useful for calibrating colors in nature, such as human skin, vegetation, or sky. In this case, we decided to replace that column with the second column of Fig. 2(b), which contains beef color samples. The colors in the color checker matrix in hexadecimal format are #FFFFFF, #D2715D, #D67E2C, #FF0000, #C8C8C8, #BB5F4F, #505BA6, #00FF00, #A0A0A0, #A3463B, #C15A63, #0000FF, #7A7A7A, #954238, #5E3C6C, #E7C71F, #555555, #8B3D3C, #9DBC40, #BB5695, #000000, #793B2B, #E0A32E, #0885A1.

The first step for correcting the color is getting the 24 color values corresponding to the color checker matrix in the images. In this research, we decided to use the median of colors (RGB space) of the pixels in the circles. In this sense, we obtained a matrix $M_{RGB} \in \mathbb{R}^{24x3}$ that contains 24 rows and 3 columns, where each row represents a color in the color checker matrix, and the columns represent the R, G, and B channels.

In general, to perform the color correction, we need a matrix that we call $M'_{RGB} \in \mathbb{R}^{24x3}$ that represents the colors that correspond to the calibrated ones. In addition to the matrix that represents the colors in the color checker matrix of images M_{RGB}. With both matrices, we can model a function F that can transform and calibrate the colors in the image (see Equation (1)).

$$M'_{RGB} = F(M_{RGB}) \tag{1}$$

In this research, we used the library Colour [31] to implement the color correction. Two techniques were tested, Cheung [32] and Vandermonde [33].

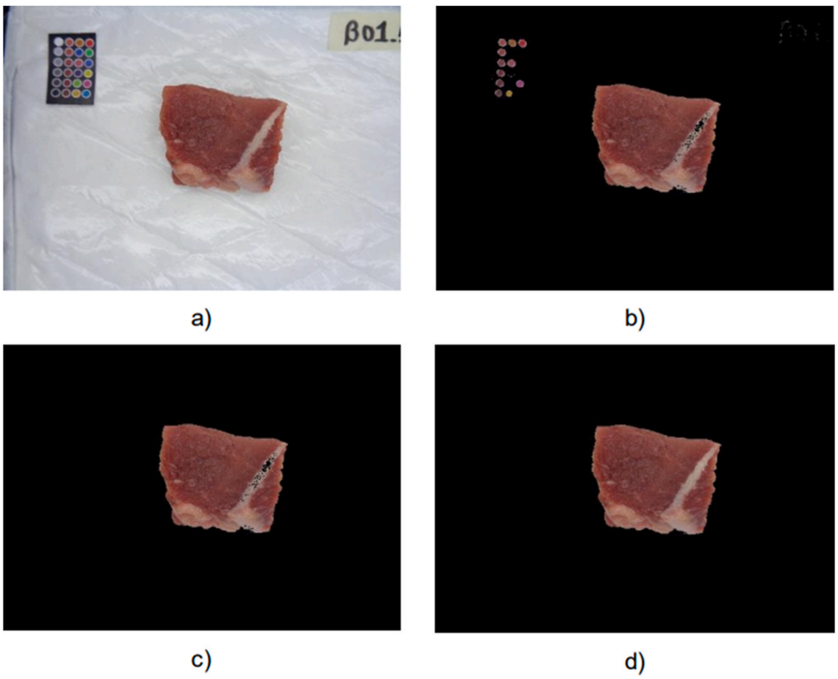

Fig. 3. Image segmentation. a) Original image. b) Binary classifier. c) Connected components. d) Binary fill holes.

2.2.3. Image segmentation and color extraction

Image segmentation divides an image into regions with similar features [34]. In this case, the goal is to keep only the meat region. The image segmentation process is described as follows (see Fig. 3(a–d)). First, we implemented a binary classifier using a Support Vector Machine (SVM) [35]. On the one hand, we had the pixels that belonged to meat; on the other hand, the pixels that belonged to the color checker matrix and the absorbent paper towels. The results are shown on Fig. 3(b). This model allows identifying the pixels with colors like meat. Then, we calculated the connected components, where in all cases, the meat corresponded to the bigger one (Fig. 3(c)). Finally, we used binary fill holes to obtain the whole meat area (Fig. 3(d)).

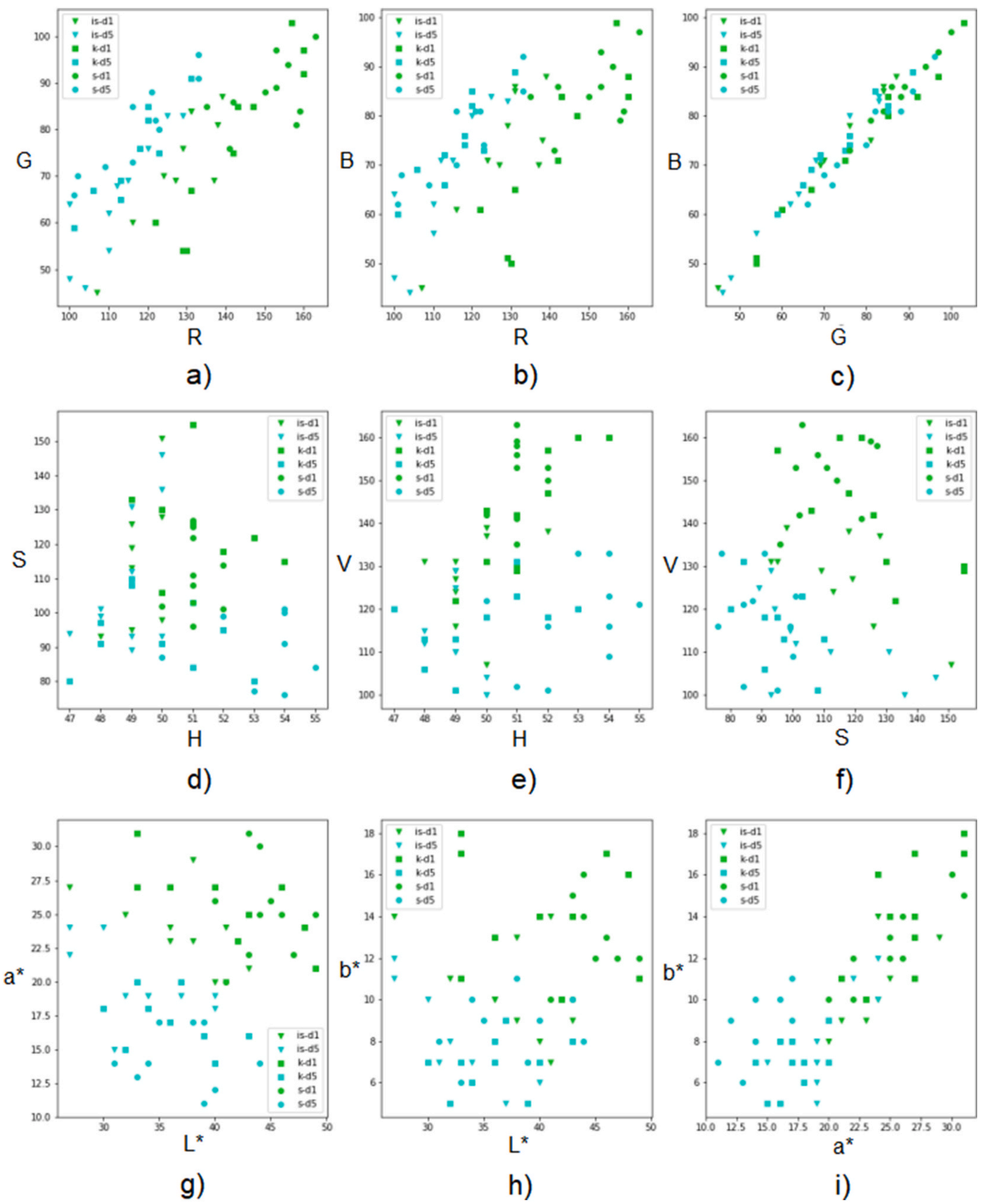

Fig. 4. Beef color values in the three color spaces: RGB, HSV, and CIELab*. a) Analysis of R and G channels. b) Analysis of R and B channels. c) Analysis of G and B channels. d) Analysis of H and S channels. e) Analysis of H and V channels. f) Analysis of S and V channels. g) Analysis of L* and a* channels. h) Analysis of L* and b* channels. i) Analysis of a* and b* channels. Green and blue points represent the beef m colors on the first (d1) and the fifth (d5) day after purchase, respectively. Triangles, squares, and circles were used to represent the inside skirt (is), knuckle (k), and sirloin (s) samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Once the image contains only the pixels corresponding to the meat, we calculated the median of the pixels' colors that correspond to lean meat without fat and connective tissue, whose saturation value is higher than 50 (S > 50).

2.3. Color analysis

Once we have the calibrated images, we use the OpenCV library [36] to obtain the values of the different color spaces. We analyzed the beef color values by calculating each channel's mean values, in addition to using scatter plots and principal component analysis.

Scatter plot: This is a visual representation of samples in a two-dimensional space. It allows for analyzing the dependency among features (see Fig. 4(a–i)).

Principal component analysis (PCA): It is an orthogonal projection of the data onto a lower-dimensional space such that the variance of the projected data is maximized [37]. It helps visualize the multidimensional data and analyze its patterns. When it is used to visualize the data in a two-dimensional space, adding the explained variance percentage is common to identify how much information is represented in the graph (see Fig. 5(a–d)).

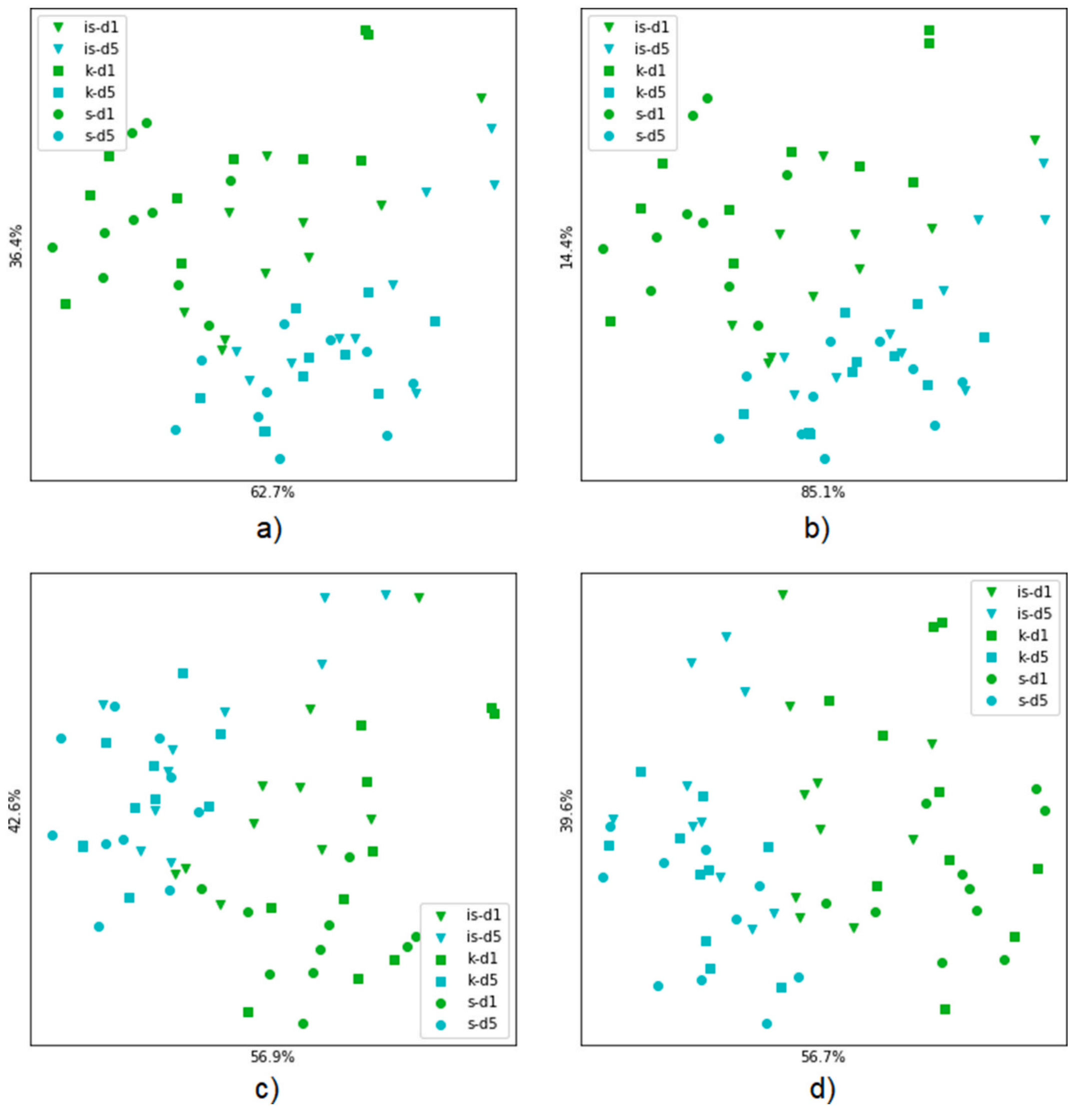

Fig. 5. Principal component analysis. Each one of the maps represents the meat samples using different color spaces a) Three color spaces: RGB, HSV, and CIELab*. b) RGB. c) HSV. d) CIELab*. Green and blue points represent the beef colors on the first (d1) and the fifth (d5) day after purchase, respectively. Triangles, squares, and circles were used to represent the inside skirt (is), knuckle (k), and sirloin (s) samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2.4. *White-box machine learning classifiers*

Once the colors are obtained, we can calculate a model for predicting if a piece of meat presents a good quality according to the color comparison between fresh and non-fresh beef. In Machine Learning, the models that predict categories based on available features are known as classifiers. In this case, we want to create a binary classifier fitted with the color of the lean meat on the first and the fifth day based on the purchased date (see Fig. 1(a–f)). The model's input data is the median color of the lean meat. We proposed three white-box machine learning models: decision trees, logistic regression, and multivariate normal distribution. White-box algorithms are understandable and they can be represented as simple equations or graphs.

2.4.1. *Decision trees*

Classification and regression trees are machine-learning methods for constructing prediction models from data [38]. The models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition (see Fig. 6(a–c)). The main idea is that samples with similar values tend to concentrate in the same region. In this research, the decision tree minimizes the Gini impurity metric that is defined as Equation (2), where x goes for all the categories of the random variable X, and P_x represents the probability of the category x.

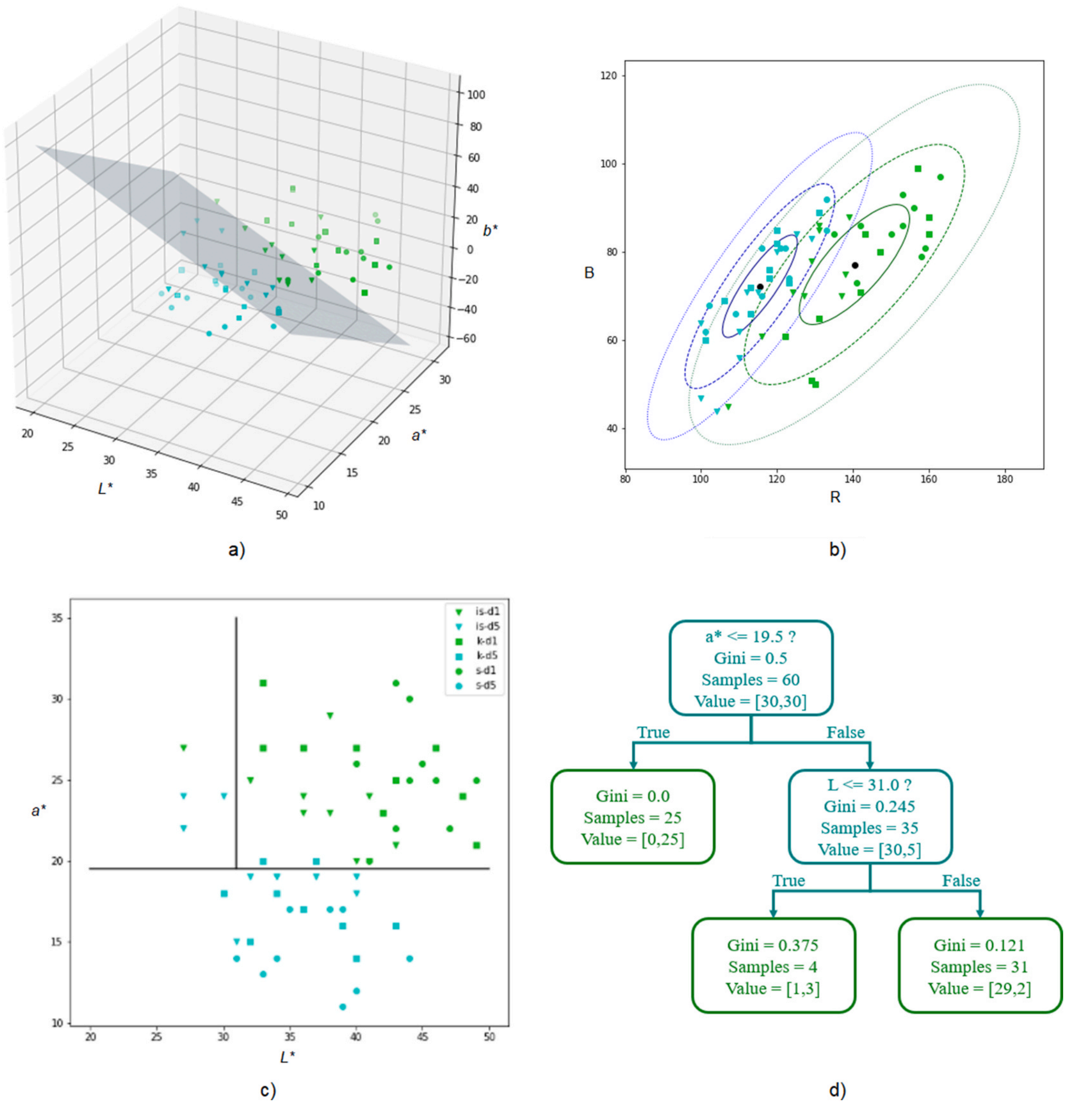

Fig. 6. White-box Machine Learning models for beef quality prediction based on color. a) Logistic Regression classifier. b) Multivariate Normal Distribution classifier. c) Decision Tree classifier. d) Space divisions of the Decision Tree classifier. Green and blue points represent the beef colors on the first and the fifth day after purchase, respectively. Triangles, squares, and circles were used to represent the inside skirt, knuckle, and sirloin samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

$$H(X) = \sum_{x} P_x * (1 - P_x) \tag{2}$$

2.4.2. Logistic regression

Logistic Regression is a linear model for solving a binary classification problem [37,39]. It divides the data using a plane (see Fig. 6(a)). The model is defined as Equation (3). The input is defined as the vector $X = [x_1, x_2, \ldots, x_d], x_i \in \mathbb{R}$. Each variable is multiplied by a coefficient $\beta_i \in \mathbb{R}$ β. In this model, when a new sample is evaluated, the class is assigned according to the sign obtained from Equation (3).

$$LR(X) = \beta_0 + \sum_{i} \beta_i x_i \tag{3}$$

2.4.3. Multivariate normal distribution

A multivariate normal distribution, based on the Gaussian function, is used to describe the distribution of numerical data that contains dependent features. Its density function is defined by Equation (4) [40], where the input is defined as the vector $X = [x_1, x_2, \ldots, x_d], x_i \in \mathbb{R}$, μ represents the mean of the data, Σ the covariance matrix, $|\Sigma|$ the determinant of Σ, and p is the rank of Σ.

$$f(X) = \frac{e^{\left(\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)\right)}}{(2\pi)^{p/2} |\Sigma|^{1/2}} \tag{4}$$

We can create a classification model based on multivariate normal distribution by fitting a Gaussian for each class. The Gaussian is defined by its mean vector and the covariance matrix. In this research, we fit a Gaussian for the fresh meat class and another for the non-fresh meat class. In this sense, to predict the label to a new sample, the class corresponding to the model with the highest likelihood is assigned.

3. Experiments and results

This section presents a detailed analysis of the color of beef based on three color spaces: RGB, HSV, and CIELab*. We analyzed 30 samples of beef: 10 of inside skirt, 10 of knuckle, and 10 of sirloin (see an example of each type of meat in Fig. 1(a–f)). In total, we had 60 color samples, 30 of them obtaining the color just after purchase, and the remaining 30 samples were taken after the fifth day that meat was purchased.

First, we present the results of different methods for the color correction of CVS images. Then, we analyzed the median of lean meat colors by comparing different color spaces. Finally, we presented the results of the performance of three white-box classifiers to solve the beef quality prediction based on color.

We used the following colors to simplify the reading and presentation of the results. Green and blue points represent the beef colors on the first and the fifth day after purchase, respectively. In addition, we used the following abbreviations: inside skirt (is), knuckle (k), sirloin (s), day 1 (d1), and day 5 (d5).

Image processing and machine learning techniques was implemented in the Python programming language using the libraries Scikit-learn, Scipy, and OpenCV.

3.1. CVS color correction

Section 2.2 explains the color correction methodology that uses a color checker matrix. We captured an image of the color checker matrix (see Fig. 2(b)) inside the cabin. Then, we used the Mean Absolute Error among the colors of that image and the images used in the experiments. Table 1 presents the errors in colors before and after calibration. Color correction significantly reduces the error among the colors in the images, and the Vandermonde technique got the best results. Knowing the range of pixels colors is between 0 and 255, the errors are small enough. In addition, using the ANOVA test, we found that the difference among the errors before and after correction (using Vandermonde) is statistically significant in all the cases, with a p-value of $9.33e^{-66}$, $3.34e^{-40}$, and $3.29e^{-62}$ for inside skirt, knuckle, and sirloin, respectively. The remaining results presented in this section take as input the images after color correction by applying the Vandermonde technique.

Table 1
Mean Absolute Error of colors before (original images) and after color calibration (Cheung and Vandermonde). The values are calculated based on the RGB color space in a range between 0 and 255.

	Inside skirt	Knuckle	Sirloin
Original images	8.72	11.49	11.79
Cheung	6.49	5.02	7.93
Vandermonde	4.37	2.90	5.05

3.2. Analysis of beef color

As we explained in the previous section, we calculated the median of the lean meat color, without fat and connective tissue, obtained by the CVS to analyze the color of fresh and non-fresh meat samples. The values of colors in the color spaces: RGB, HSV, and CIELab*, are presented in Table 2 and Fig. 4(a–i). It is important to note that the R and V channels have the same values in this experiment. It is because $V = \max(R,G,B)$, and in these images, the bigger value is R. From Fig. 4(a–i) and Table 2, we can observe that the pair of channels that better differentiate between fresh and non-fresh beef colors are L* and a*, and, R and G. This is in accordance with Quevedo et al. that affirmed that the channels that differentiate fresh and non-fresh meat better are R, L*, and a* [24]. Some documents define ranges of colors' values to consider an acceptable piece of beef. For example, Holman et al. established that beef color was considered acceptable when a*≥14.5 [28]. Realini et al. defined a fresh beef steak when the color values are L*>39.5, a*>16.8, and b*>6.3 [25]. Zhang et al. defined the range of fresh beef in the ranges of L*>31.4, a*>16.4, and b*>6.5 [26]. Wang et al. established that acceptable frozen beef rolls have a*>16.4 [27]. They all establish that a fresh piece of beef must have an a* value greater than 14.5 or 16.8. However, as seen in Fig. 4(a–i), drawing a horizontal or vertical line that divides the quality attributes according to their color between fresh and non-fresh samples is impossible. It means that beef quality prediction must be by multivariate techniques instead of using the channel values separately.

Fig. 5(a–d) presents a color analysis but uses principal component analysis. The explained variance in the maps is 99.1%, 99.5%, 99.5%, and 96.3%, respectively. In all these figures, we can observe that fresh and non-fresh beef colors are separated. Consequently, fitting a model for beef quality prediction based on color is possible.

Fig. 1(a–f) shows the median color obtained from each image. It can be observed that differences between fresh and non-fresh groups are more notable in knuckle and sirloin than in inside skirt. Also, in Fig. 5(a–d), the inside skirt samples are close to the boundary of fresh and non-fresh samples.

3.3. Prediction of color parameters in fresh meat

This research's main objective is to identify fresh meat's color parameters. Once the lean meat median color is identified, we must determine if it corresponds to a fresh or a non-fresh piece of meat. We compared the performance of three white-box classifiers: Decision Tree, Logistic Regression, and the use of Multivariate Normal Distributions. In addition, we analyzed the combination of all color spaces and each space separately. We used accuracy as the performance metric representing the percentage of correct predictions. Its value goes from zero to one, being one a perfect performance. For validating the results, we used cross-validation with five folds.

Table 3 shows the classification results. It can be observed that the classifier with better results is Logistic Regression, followed using Multivariate Normal Distributions. Logistic Regression was the only one with perfect prediction in two cases: using the three-color spaces and using only CIELab*. In terms of color spaces, the one with better results was CIELab*, and its results are generally better even when all three spaces are used. However, the best color space is RGB when the classifier is based on Multivariate Normal Distributions.

We aim to analyze the classification models and understand their equations or graphs. Below, we present the model with the best performance of each type of classifier.

Logistic Regression classifier: As we said in Section 2.7, Logistic Regression uses a plane that divides the samples of different categories. In this experiment, it is the model with the best performance. Equation (5) presents the plane that divides the colors of fresh and non-fresh pieces of beef using the CIELab* color space. A positive result indicates non-fresh meat. Otherwise, a negative result indicates fresh meat. Fig. 6(a) shows the plane and the color samples.

$$LR(L^*, a^*, b^*) = 67.845 - 0.877L^* - 1.533a^* - 0.353b^* \quad (5)$$

Multivariate Normal Distribution: It is the model with the second-best results. As we said previously, it fitted a multivariate normal distribution to each class. On the one hand, the distribution for the colors of fresh beef, in terms of RGB, is defined by the mean

Table 2
Mean beef meat color values in the color spaces: RGB, HSV, and CIELab*.

	1st day			5th day		
	R	G	B	R	G	B
Inside skirt	127.9	72.5	72.9	112.5	65.3	66.2
Knuckle	142.1	77.2	73.3	116.3	74.5	74.6
Sirloin	151.0	88.0	85.3	117.6	80.3	76.0
	H	**S**	**V**	**H**	**S**	**V**
Inside skirt	49.5	115.0	127.9	48.9	109.4	112.5
Knuckle	51.3	125.5	142.1	49.8	93.9	116.3
Sirloin	51.1	110.9	151.0	52.9	89.4	117.6
	L*	***a****	***b****	***L****	***a****	***b****
Inside skirt	32.7	23.6	10.5	33.2	19.8	7.9
Knuckle	40.3	26.3	14.1	36.0	17.1	6.9
Sirloin	44.2	25.2	12.8	37.6	14.5	8.5

Table 3
Beef meat freshness prediction results using the three machine learning models and different color spaces. The results are presented using accuracy. Cross-validation with five folds was used. The last five columns present the results of each fold, and the column mean shows the mean result. Bold numbers enhance the best results of each model.

	Mean	1	2	3	4	5
Decision Tree						
All channel	0.87	0.75	1.00	0.83	0.92	0.83
RGB	0.78	0.75	0.83	0.83	0.83	0.67
HSV	0.85	0.92	0.83	0.83	0.83	0.83
CIELab*	**0.90**	0.83	0.92	0.92	0.92	0.92
Logistic Regression						
All channels	1.00	1.00	1.00	1.00	1.00	1.00
RGB	0.97	1.00	1.00	0.83	1.00	1.00
HSV	0.98	1.00	1.00	1.00	1.00	0.92
CIELab*	**1.00**	1.00	1.00	1.00	1.00	1.00
Multivariate normal distribution						
All channels	0.93	0.92	0.83	1.00	0.92	1.00
RGB	**0.98**	1.00	0.92	1.00	1.00	1.00
HSV	0.92	0.92	0.75	1.00	1.00	0.92
CIELab*	0.97	0.92	0.92	1.00	1.00	1.00

vector μ_{d1} and the covariance matrix Σ_{d1} presented in Equation (6). On the other hand, the distribution of the colors of non-fresh beef, in terms of RGB, is defined by the mean vector μ_{d5} and the covariance matrix Σ_{d5} presented in Equation (7). Besides, Fig. 6(b) presents the graphical representation of these models. It can be observed that the mean difference is the R-value.

$$\mu_{d1} = \begin{bmatrix} 140.33 \\ 79.23 \\ 77.17 \end{bmatrix} \quad \Sigma_{d1} = \begin{bmatrix} 203.29 & 171.82 & 144.84 \\ 171.82 & 204.31 & 187.39 \\ 144.84 & 187.39 & 178.61 \end{bmatrix} \tag{6}$$

$$\mu_{d5} = \begin{bmatrix} 115.47 \\ 73.37 \\ 72.27 \end{bmatrix} \quad \Sigma_{d5} = \begin{bmatrix} 93.92 & 100.83 & 94.28 \\ 100.83 & 148.57 & 134.80 \\ 94.28 & 134.80 & 129.93 \end{bmatrix} \tag{7}$$

Decision Tree classifier: In this experiment, the classifier with the lowest results is the decision tree, which can be explained because the main idea of decision trees is to divide the space by the feature values. Fig. 6(c–d) shows the decision tree fitted to solve the beef quality prediction using the CIELab* color space. The first division is based on channel a*. The model classifies all the samples with a*≤19.5 as non-fresh. For samples with a*>19.5, another question is presented, the samples with L*≤31.0 are labeled as non-fresh, and the ones with L*>31.0 are labeled as fresh. In other words, fresh samples have a*>19.5 and L*>31.0 values. Interestingly, the samples presented in the rectangle that appears in the up-left corner are inside skirt. It means that the first node of the tree (a*≤19.5) could be enough if we only classified samples of knuckle and sirloin.

In this experiment, Logistic Regression is the model with the best performance. However, the ones that are simpler to understand are Multivariate Normal Distribution and Decision Tree. These models confirm that the channels that best differentiate between fresh and non-fresh beef colors are R, $L*$, and $a*$. We can say that color of fresh and non-fresh beef has an R-value of around 140.33 and 115.47, with a standard deviation of 14.26 and 9.69, respectively. If we use channel $a*$, pieces with values lower than 19.5 are considered non-fresh. However, in this analysis, we determine that for a correct prediction, a multivariate classifier is needed; the use of only one channel is not enough.

4. Discussion

Color is the main attribute that the consumer values when buying fresh meat, and it is one factor that determines the product's value at the time of its commercialization. Also, it is a parameter used to measure the quality of meat. The consumer relates the color of the meat with the sensory and microbial quality. Therefore, it is important to evaluate the color of the meat to determine the parameters presented by fresh or non-fresh meat.

Like [24], our results show that the channels that differentiate fresh and non-fresh meat better are R, L*, and a*. Some documents define ranges or values of colors to consider an acceptable piece of beef. For example, Holman et al. established that beef color was considered acceptable when a*≥14.5 [28]. Realini et al. defined a fresh beef steak when the color values are L*>39.5, a*>16.8, and b*>6.3 [25]. Zhang et al. defined the range of fresh beef in the ranges of L*>31.4, a*>16.4, and b*>6.5 [26]. Wang et al. established that acceptable frozen beef rolls have a*>16.4 [27]. They all establish that a fresh piece of beef must have an a* value greater than 14.5 or 16.8. Most documents analyze the color channels separately, and our proposal performs a multivariate analysis. From our results, we conclude that it is not possible to divide the fresh and non-fresh samples using only one channel (see Fig. 4(a–i)), It is necessary to use two or more channels. Our decision tree (see Fig. 6(c–d)) shows that fresh beef pieces have color values of a*> 19.5 and L*> 31.0. In our case, the best classifier was Logistic Regression, which uses a plane that divides the fresh and non-fresh samples (see Fig. 6(a)) based on the CIELab* color space.

Some documents have proposed machine learning algorithms based on computer vision and machine learning techniques for meat quality prediction. For example, Taheri-Garavand et al. proposed a nondestructive intelligent approach to chicken meat freshness based on a genetic algorithm and an artificial neuronal network [22]. Medeiros et al. proposed AutoML models to classify the freshness levels of tuna and salmon [23]. Arsalane et al. used principal component analysis and support vector machines for rapid beef freshness prediction and identification [29]. All of them have results that are high accuracy. Moreover, the analysis of how the prediction is made is so complex. In contrast, our proposal also has high results and allows us to analyze and understand the prediction models.

5. Conclusion

Color in beef is the most critical parameter for purchasing this food product because the consumer relates it to meat quality. We have discussed in this research that computing vision and white-box machine learning techniques can lead us to measure beef quality through color analysis adequately. Our proposal was divided into three stages: (1) obtaining the color of beef samples through CVS; (2) using CIELab* or RGB as color space; and (3) using white-box classifiers for estimating beef quality according to color. This method could be used as a quality parameter in stores and food factories.

Our results show that the channels better differentiate fresh and non-fresh beef pieces are R, L*, and a*. However, separate channels cannot adequately differentiate among samples. There is a need for a multivariate methodology. We use white-box machine learning techniques (Logistic Regression, Decision Trees, and Multivariate Normal Distributions) that not only allow us to obtain high accuracy in predictions but also to understand and analyze how the predictions are made. From the decision tree model, we can establish that fresh beef pieces have color values of $a^* > 19.5$ and $L^* > 31.0$. However, the best prediction was performed by using CIELab* and Logistic Regression, which uses a plane to separate the fresh and non-fresh beef samples in the 3d color space (L*, a*, and b*).

Our study's limitation is that until now, we have considered only the median value of the lean meat color. In future work, it could be interesting to analyze also the fat proportion and color. In addition, we can analyze more cuts to expand the knowledge of more beef parts.

Author contribution statement

Claudia Nallely Sanchez: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper. María Teresa Orvañanos-Guerrero: Performed the experiments; Wrote the paper. Julieta Domínguez-Soberanes: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper. Yenizey M. Álvarez-Cisneros: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Data availability statement

Data associated with this study has been deposited at https://data.mendeley.comunder the link https://data.mendeley.com/datasets/wvhkpppddp/1.

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

Authors want to thank Maximiliano Lara, Esteban García, Juan Pablo Cisneros, Luis Enrique Orozco, Ernesto Rosales-Tavera, and Esthela Fernandez for helping us with the images caption, labelling and processing. This research was funded by Universidad Panamericana through the grant "Fondo Fomento a la Investigación 2020", under Project Code UP-CI-2020-AGS-07-ING.

References

[1] S. Rajic, S. Simunovic, V. Djordjevic, M. Raseta, I. Tomasevic, I. Djekic, Quality multiverse of beef and pork meat in a single score, Foods 11 (8) (2022) 1154.
[2] I. Tomasevic, I. Djekic, M. Font-i-Furnols, N. Terjung, J.M. Lorenzo, Recent advances in meat color research, Curr. Opin. Food Sci. 41 (2021) 81–87.
[3] B.A. Altmann, J. Gertheiss, I. Tomasevic, C. Engelkes, T. Glaesener, J. Meyer, D. Mörlein, Human perception of color differences using computer vision system measurements of raw pork loin, Meat Sci. 188 (2022), 108766.
[4] A.M.S. Association, et al., AMSA Meat Color Measurement Guidelines: AMSA, American Meat Science Association, 2012.
[5] B.H. Salueña, C.S. Gamasa, J.M.D. Rubial, C.A. Odriozola, CIELAB color paths during meat shelf life, Meat Sci. 157 (2019), 107889.
[6] I. Tomasević, V. Tomović, B. Milovanović, J. Lorenzo, V. Đorđević, N. Karabasil, I. Đekić, Comparison of a computer vision system vs. traditional colorimeter for color evaluation of meat products with various physical properties, Meat Sci. 148 (2019) 5–12.
[7] S.P. Suman, P. Joseph, Myoglobin chemistry and meat color, Annu. Rev. Food Sci. Technol. 4 (1) (2013) 79–99.
[8] R. Ramanathan, M.C. Hunt, T. Price, G.G. Mafi, Strategies to limit meat wastage: focus on meat discoloration, Adv. Food Nutr. Res. 95 (2021) 183–205 (Academic Press).
[9] I. Tomasevic, V. Tomovic, P. Ikonic, J.M.L. Rodriguez, F.J. Barba, I. Djekic, D. Zivkovic, Evaluation of poultry meat colour using computer vision system and colourimeter: is there a difference? Br. Food J. 121 (5) (2019) 1078–1087.
[10] I.B. Tomasevic, Computer vision system for color measurements of meat and meat products: a review, Theo. Prac. Meat Process. 3 (4) (2018) 4–15.

[11] M. Chmiel, M. Słowiński, K. Dasiewicz, T. Florowski, Use of computer vision system (CVS) for detection of PSE pork meat obtained from m. semimembranosus, LWT–Food Sci. Technol. 65 (2016) 532–536.
[12] M. Modzelewska-Kapituła, S. Jun, The Application of Computer Vision Systems in Meat Science and Industry–A Review, Meat Science, 2022, 108904.
[13] K. Chen, X. Sun, C. Qin, X. Tang, Color grading of beef fat by using computer vision and support vector machine, Comput. Electron. Agric. 70 (1) (2010) 27–32.
[14] R.E. Larraín, D.M. Schaefer, J.D. Reed, Use of digital images to estimate CIE color coordinates of beef, Food Res. Int. 41 (4) (2008) 380–385.
[15] M. Chmiel, M. Słowiński, K. Dasiewicz, Lightness of the color measured by computer image analysis as a factor for assessing the quality of pork meat, Meat Sci. 88 (3) (2011) 566–570.
[16] M. Chmiel, M. Słowiński, The use of computer vision system to detect pork defects, LWT (Lebensm.-Wiss. & Technol.) 73 (2016) 473–480.
[17] M. Chmiel, M. Słowiński, K. Dasiewicz, T. Florowski, Application of a computer vision system to classify beef as normal or dark, firm, and dry, J. Anim. Sci. 90 (11) (2012) 4126–4130.
[18] M.F. Rahman, A. Iqbal, M.A. Hashem, A.A. Adedeji, Quality assessment of beef using computer vision technology, Food Sci. Ani. Resour. 40 (6) (2020) 896.
[19] A. Girolami, F. Napolitano, D. Faraone, A. Braghieri, Measurement of meat color using a computer vision system, Meat Sci. 93 (1) (2013) 111–118.
[20] A. Issac, M.K. Dutta, B. Sarkar, Computer vision based method for quality and freshness check for fish from segmented gills, Comput. Electron. Agric. 139 (2017) 10–21.
[21] B. Milovanovic, I. Đekić, B. Solowiej, S. Novaković, V. Đorđević, I. Tomašević, Computer Vision System: a better tool for assessing pork and beef colour than a standard colourimeter, Scien. J. "Meat Technology" 61 (2) (2020) 153–160.
[22] A. Taheri-Garavand, S. Fatahi, F. Shahbazi, M. de la Guardia, A nondestructive intelligent approach to real-time evaluation of chicken meat freshness based on computer vision technique, J. Food Process. Eng. 42 (4) (2019), e13039.
[23] E.C. Medeiros, L.M. Almeida, J.G.D.A.T. Filho, Computer vision and machine learning for tuna and salmon meat classification, Informatics 8 (No. 4) (2021) 70 (MDPI).*Key reference: They proposed using computer vision and machine learning as a nondestructive method for external quality detection of tuna and salmon meat products getting good prediction results.
[24] R. Quevedo, E. Valencia, G. Cuevas, B. Ronceros, F. Pedreschi, J.M. Bastías, Color changes in the surface of fresh cut meat: a fractal kinetic application, Food Res. Int. 54 (2) (2013) 1430–1436.
[25] C.E. Realini, M. Staincliffe, K. Taukiri, C. Craigie, M.M. Farouk, Is colour preference for beef changing for New Zealand consumers?, in: Paper Presented at the 64th International Congress of Meat Science and Technology, 2018 (Melbourne, AUS).*Key reference: They determined that seventy two percent of the consumers preferred colour steaks corresponding to 39.5–40.9 L*, 16.8–21.6 a* and 6.3–9.1 b* colour values. However, they do not perform a multivariate analysis. They analyze the color of dark-cutting beef based on Asian consumers. They use Linear Regression to determine acceptance thresholds $L^* \geq 31.4, a^* \geq 16.4, b^* \geq 6.5$. However, these models are fitted by each channel separately instead of a multivariate analysis such as our proposal.
[26] Y. Zhang, B.W. Holman, Y. Mao, X. Chen, X. Luo, D.L. Hopkins, Y. Zhang, Determination of a pH threshold for dark cutting beef based on visual evaluation by Asian consumers, Meat Sci. 172 (2021), 108347.*Key reference: They analyze the color of dark-cutting beef based on Asian consumers. They use Linear Regression to determine acceptance thresholds $L^* \geq 31.4, a^* \geq 16.4, b^* \geq 6.5$. However, these models are fitted by each channel separately instead of a multivariate analysis such as our proposal.
[27] F. Wang, B.W. Holman, Y. Zhang, X. Luo, Y. Mao, D.L. Hopkins, Investigation of colour requirements of frozen beef rolls by Chinese consumers for hot pot, Meat Sci. 162 (2020), 108038.*Key reference: They found that to satisfy consumers' requirements of the color of frozen beef rolls the value a* is the most suitable indicator with an acceptance threshold of $a^* \geq 16.4$. They calculate this value by performing linear regression where the models were fitted by each channel separately instead of a multivariate analysis such as our proposal.
[28] B.W. Holman, R.J. van de Ven, Y. Mao, C.E. Coombs, D.L. Hopkins, Using instrumental (CIE and reflectance) measures to predict consumers' acceptance of beef colour, Meat Sci. 127 (2017) 57–62.
[29] A. Arsalane, N. El Barbri, A. Tabyaoui, A. Klilou, K. Rhofir, A. Halimi, An embedded system based on dsp platform and PCA-SVM algorithms for rapid beef meat freshness prediction and identification, Comput. Electron. Agric. 152 (2018) 385–392.*Key reference: They propose a rapid beef freshness prediction methodology. They use hue, saturation, and intensity color space and PCA and SVM for the prediction. They obtained a 100% success rate, like ours. However, our proposal allows us to analyze and understand the prediction models.
[30] G.D. Finlayson, M. Mohammadzadeh Darrodi, M. Mackiewicz, The alternating least squares technique for nonuniform intensity color correction, Color Res. Appl. 40 (3) (2015) 232–242.
[31] T. Mansencal, M. Mauderer, M. Parsons, N. Shaw, K. Wheatley, S. Cooper, J.D. Vandenberg, L. Canavan, K. Crowson, O. Lev, K. Leinweber, S. Sharma, T. J. Sobotka, D. Moritz, M. Pppp, C. Rane, P. Eswaramoorthy, J. Mertic, B. Pearlstine, M. Leonhardt, O. Niemitalo, M. Szymanski, M. Schambach, S. Huang, M. Wei, N. Joywardhan, O. Wagih, P. Redman, J. Goldstone, S. Hill, J. Smith, F. Savoir, G. Saxena, S. Chopra, I. Sibiryakov, T. Gates, G. Pal, N. Tessore, A. Pierre, Colour 0.4.1, 2022, https://doi.org/10.5281/zenodo.6288658.
[32] V. Cheung, S. Westland, D. Connah, C. Ripamonti, A comparative study of the characterisation of colour cameras by means of neural networks and polynomial transforms, Color. Technol. 120 (1) (2004) 19–25.
[33] S. Westland, C. Ripamonti, V. Cheung, Computational Colour Science Using MATLAB, John Wiley & Sons, 2012.
[34] C.N. Sánchez, M. Rivera, R. Velázquez, Robust multiband image segmentation method based on user clues, 2017, in: IEEE 37th Central America and Panama Convention, IEEE, 2017, pp. 1–6 *(CONCAPAN XXXVII)*.
[35] I. Steinwart, A. Christmann, Support Vector Machines, Springer Science & Business Media, 2008.
[36] G. Bradski, The openCV library, Dr. Dobb's J. Softw. Tools Prof. Program. 25 (11) (2000) 120–123.
[37] C.M. Bishop, N.M. Nasrabadi, Pattern Recognition And Machine Learning (Vol. 4, No. 4, p. 738), Springer, New York, 2006.
[38] W.Y. Loh, Classification and regression trees, Wiley Interdis. Rev.: Data Min. Knowl. Discov. 1 (1) (2011) 14–23.
[39] H.F. Yu, F.L. Huang, C.J. Lin, Dual coordinate descent methods for logistic regression and maximum entropy models, Mach. Learn. 85 (1) (2011) 41–75.
[40] T. Hastie, R. Tibshirani, J.H. Friedman, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2, springer, New York, 2009, pp. 1–758.