



Discrimination of beef meat according to the level of farming intensification through visible and near infrared spectroscopy data and chemometrics

Annalaura Lopez^{a,b}, Elea Caumette^a, Rennie C. Eppenstein^c, Nicolò Amalfitano^d, Isabelle Legrand^e, Lynda S. Perkins^f, Nigel Scollan^f, Pablo Guarnido-Lopez^{g,1}, Isabelle Constant^a, Fabien Ferlay-Demiaz^a, Carmen L. Manuelian^h, Bruno Martin^a, Jean-François Hocquette^a, Donato Andueza^{a,*}

^a INRAE, Université Clermont Auvergne, VetAgro Sup, UMR1213 Herbivores, 63122 St-Genès-Champanelle, France

^b Department of Veterinary Medicine and Animal Science (DIVAS), University of Milan, Via dell'Università, 26900 Lodi, Italy

^c Research Institute of Organic Agriculture (FiBL), Ackerstrasse 113, 5070 Frick, Switzerland

^d Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy

^e Institut de l'Élevage (IDELE), Maison Régionale de l'Agriculture Nouvelle-Aquitaine, Boulevard des Arcades, Cedex 2, 87060 Limoges, France

^f Institute for Global Food Security (IGFS), School of Biological Sciences, Queen's University Belfast, 19 Chlorine Gardens, BT9 5DL Belfast, United Kingdom

^g Institute for Meat and Animal Product Quality (IPROCAR), University of Extremadura, Avenida de la Universidad s/n, 10003 Cáceres, Spain

^h Ruminant Research Group (G2R), Department of Animal and Food Sciences, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Spain

ARTICLE INFO

Keywords:

Meat
NIR
Authentication
Farming system
Quality
Extensification

ABSTRACT

The European livestock sector is highly diverse, comprising beef production systems which differ significantly in their levels of intensification, from grass-fed cull cows to intensive feedlot systems finishing young cattle. Different production practices are directly correlated to different quality grades of beef products, creating opportunities for differentiation in the market but also increase the risk of fraud. In this paper, we evaluated the performances of three spectral devices (two portable visible and NIR and one benchtop visible/NIR) and linear discriminant analysis in discriminating beef samples based on the level of intensification of farming systems. Discrimination performances were influenced by the characteristics of the device (such as the spectral range), as well as the sample preparation protocol and the dataset split approach chosen for the discriminant model. Overall, the best classification rates were obtained for samples belonging to the classes associated to the most distinct characteristics, including animal age and type (cull sucklers cows vs young beef) and production system (concentrates supplementation vs exclusively grass-fed systems). Misclassification was observed for classes characterized by high similarity, including animals fattened indoors in feedlots until 14–15 months of age with similar proportions of concentrates and maize silage. This study showed that visible and NIR spectroscopy combined with appropriate chemometric tools allow to detect differences in beef composition associated to different farming practices and degrees of intensification. Portable and lower-cost spectral devices achieved averagely lower classification accuracy than the benchtop instrument, providing new insights in innovation of authentication-driven technologies in the beef sector.

1. Introduction

In 2024, Europe produced 42.2 million tons of meat, of which bovine meat, including veal (animals <8 months) and beef (>8 months), accounted for 15.6% (6.6 million tons). This represented a 3.1% increase

over 2023 production, marking a notable upturn after five consecutive years of decline (Eurostat, 2025). The diversity of production systems and supply chains in the European beef sector has evolved over decades in response to varied geographical, climatic, economic, and societal conditions across different regions of the continent. While certain

* Corresponding author.

E-mail address: donato.andueza@inrae.fr (D. Andueza).

¹ Current address: Department of Animal Science, Texas A&M University, College Station, TX 77843, USA.

<https://doi.org/10.1016/j.meatsci.2026.110169>

Received 27 February 2026; Received in revised form 26 June 2026; Accepted 29 June 2026

Available online 30 June 2026

0309-1740/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

systems demonstrate high efficiency in terms of environmental impact or economic performance, others may be less efficient but fulfill critical roles in land management and rural vitality (Hocquette et al., 2018). This diversity offers consumers a range of opportunities, but it adds considerable complexity to the sector (Hocquette et al., 2018). In recent decades, livestock intensification has resulted in increased productivity per animal or per unit of land, potentially freeing up agricultural areas for alternative purposes. However, this production model faces growing criticism for its ethical and environmental impacts, as the expanding feed requirements of intensive systems drive land use expansion across multiple continents (van der Sluis et al., 2016).

Overall, European beef cattle systems can be broadly grouped into three categories based on management intensity and housing (Dumont et al., 2017). Intensive feedlot-type systems, representing about one-third of production, rely on high stocking densities and minimal grassland, aiming for uniform products at low cost. Grassland-based systems dominate in pastoral regions, often linked to quality labels and making limited use of land suitable for human-edible crops. Mixed crop–livestock systems show the greatest variability, facing competition between crops and livestock but often improving sustainability through diversified rotations and increased feed self-sufficiency. In reality, many intermediate systems blur these boundaries, combining indoor and outdoor practices according to local conditions and farmer strategies. The grazing period, for instance, varies considerably by region and can range from a few months to year-round pasture access in milder climates (Caumette et al., 2026). The degree of supplementary feeding, housing infrastructure, and finishing strategies creates a continuum of production models rather than discrete categories, reflecting the adaptive nature of European beef farming to diverse territorial contexts (Cozzi et al., 2025).

The growing interest in meat quality among consumers extends beyond compositional attributes to include factors belonging to both the intrinsic and the extrinsic quality traits (Prache et al., 2022; Prieto et al., 2017). These characteristics are often associated with premium and/or labeled products, creating opportunities for market differentiation. However, this also raises concerns about increased risk of fraud. Consequently, reliable analytical tools are increasingly needed to ensure authenticity and maintain consumer trust (Dimitrakopoulou & Vantarakis, 2023). Infrared (IR) spectroscopy has emerged as a promising technique for meat authentication and classification. Previous studies have demonstrated its effectiveness as a discriminative tool to assess meat and meat products based on species, breed, diet, geographical origin, quality grades, and technological treatments (Prieto et al., 2017; Sentandreu & Sentandreu, 2014). For beef, the spectral information relevant for discrimination is captured both in the visible (Vis) and the near infrared (NIR) regions and generally reflects differences in colour and chemical composition, including intramuscular fat, marbling, fatty acids, protein and moisture content (Barragán et al., 2021; Kombolo-Ngah et al., 2023; León-Ecay et al., 2025). Recent technological advances have introduced portable and consumer-grade NIR devices, which offer affordability and convenience, including reduced analysis time, reduction in number of chemical tests, reduction of hazards and overall cost (Dixit et al., 2017). Despite their narrower wavelength range (typically 900–1700 nm), these devices can show promising results. For beef, it has been demonstrated that portable NIR devices can be efficient in differentiating feeding systems (grass-fed vs grain-fed or barley vs corn), with high accuracy for classification models like linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA) (Barragán et al., 2021; Coombs et al., 2021).

Despite the growing body of literature on beef authentication through IR spectroscopy based on feeding systems (Barragán et al., 2021; Coombs et al., 2021; León-Ecay et al., 2025; León-Ecay, Insausti, et al., 2024), the application of this technique for discriminating beef samples by farm intensification levels remains unexplored. This study aimed to determine whether production systems with

different farming practices and degrees of intensification lead to detectable differences in beef quality when analysed using Vis, NIR, and combined Vis/NIR spectroscopy coupled with appropriate classification algorithms. It also sought to assess whether portable, lower-cost spectrometers can achieve classification performance comparable to conventional full-range laboratory instruments, with only limited reductions in accuracy. Accordingly, we evaluated the ability of Vis, NIR, and combined Vis/NIR technology to classify beef samples according to farm intensification level and compared the classification accuracy of portable devices with that of laboratory-based spectrometers.

2. Materials and methods

2.1. Samples

Samples analysed in this study were collected within the EU H2020 INTAQT Project (<https://h2020-intaqt.eu/>). A total of 497 beef samples from the *Longissimus thoracis* and *lumborum* muscles, collected between the 8th and the 9th rib, were included in the study. Samples were collected from $N = 55$ fattening beef farms set in five European countries: Switzerland ($N = 20$), Spain ($N = 5$), France ($N = 10$), Italy ($N = 10$) and the UK ($N = 10$). Five to ten animals from each farm were included in the study. Farms were categorised into six different classes (referred to as classes, Class0 to Class5 along the manuscript), based on the level of intensification of the studied farming systems as reported by Caumette et al. (2026). The main features of the six classes are presented in Table 1. Briefly, Class0 was represented by suckler farms in France (FR) fattening cows before culling at 90 months of age, either on pasture or indoor, in grass-based systems with medium-high levels of feed supplementation (median: 7.3 kg). Class1 was represented by Swiss (CH) suckler beef farms, fattening weaners from 4 to 11 months of age (young animals) in grass-based systems with few or none concentrates supplementation in the diet (median: 0 kg). Class2 was characterized by young dairy-beef crosses that arrived in CH farms at 6 months of age and were fattened until 23 months of age on grasslands with no or minimum supplementation (median: 0 kg). Class3 included medium-input farms from the UK, keeping fattening cattle on grasslands from 12 to 27 months of age (yearlings) for most of the year, and supplying a medium level of concentrate (median: 2.5 kg). Class4 was represented by medium-input CH farms fattening mixed genetic-type cattle indoors from 9 to 16 months of age using medium levels of concentrates (median: 5.8 kg). Class5 included high-input indoor fattening farms from Italy (IT), Spain (ES) and CH, relying on high levels of concentrates and maize silage in the diet (median: 8.2 kg) and fattening animals until 14 months. More detailed information is provided in Caumette et al. (2026).

2.2. Vis and NIR spectra measurements

Spectral data from beef samples were collected using three different devices. The first device was the FlameNIR spectrometer by Ocean Optics, Inc. (Orlando, FL, USA). This is a portable spectrometer reading reflectance in the NIR range 935 to 1665 nm and controlled via the OceanView software. The spectral resolution of the FlameNIR was set to 5–6 nm, and the raw spectra collected with this device consisted of a total of 128 data points. The second device was a Vis spectrometer (IndiGo from Goyalab, Talence, France). This is a portable device that reads reflectance in the Vis range from 380 to 780 nm and is controlled via Spectrolab software. The spectral resolution of this tool was 0.5 nm. Only the range between 460.5 and 700 nm was used, thus the raw spectra collected with this device consisted of a total of 480 data points. The third device was a NIRSTM DS2500 scanning spectrometer (for the sake of clarity, DS2500 along the paper), a benchtop laboratory spectrometer reading reflectance values in the Vis/NIR range 400 to 2500 nm, run using the ISIScan Nova software package (FOSS Analytical,

Table 1

Description of intensification classes used in this study, as defined by input-derived variables.

	Classes					
	Class0	Class1	Class2	Class3	Class4	Class5
Genetic type	Beef	Dairy	Mixed	Mixed	Mixed	Mixed
Farms (number and Countries)	10 FR	5 CH	5 CH	10 UK	8 CH	17 (10 IT, 5 ES, 2 CH)
Slaughter age (months)						
median	90	11	23	28	16	14
IQR	79–99	10–11	23–24	24–29	14–16	13–18
Slaughter weight (kg)						
median	827	400	585	700	545	550
IQR	800–901	375–420	555–600	650–780	522–562	530–598
Grazing or indoor	Grazing	Grazing	Grazing	Grazing	Indoor	Indoor
HEFI (kg/day/head)						
median	7.3	0	0	2.5	5.8	8.2
IQR	5.5–10.9	0–0	0–0	2.0–8.3	2.9–7.1	6.5–8.5
Total fertilization (N/ha UAA)						
median	80	89	238	371	213	264
IQR	55–140	82–204	158–260	174–432	166–232	205–327
N meat samples	97	50	50	98	80	122

Countries: CH – Switzerland, ES – Spain, FR – France, IT – Italy, UK – The United Kingdom. IQR = Inter Quartile Range; HEFI = High Energy Feed Inputs, including concentrates and maize silage; UAA = Utilised Agricultural Area. For further details on farm characteristics and intensification input/output variables, read [Caumette et al. \(2026\)](#).

Hillerød, Denmark). In this case, the instrumental resolution was set to 2 nm and the raw spectra collected with the DS2500 consisted of a total of 1050 data points.

For the FlameNIR and the IndiGo Vis devices, no previous preparation of the samples was performed, but the readings were made directly on the surface of the entire steak (approximately 5 cm²). Samples were kept at –20 °C until the day of the analysis, when they were thawed at room temperature. The probes of the devices were placed in contact with the surface of the steak after covering it with a plastic film purposely made for contact with food. Ten replicates were made for each sample, covering different points of the steak surface. Spectra were inspected using the WinISI 4 software (FOSS Analytical, Hillerød, Denmark). The root mean square statistic was used to assess the agreement between spectral replicates for each sample. These were consequently filtered choosing a standard value equal to 4374 as a cut-off, to develop a representative library spectral file, free of errors due to intra-sample variation, as suggested by [Martínez et al. \(1998\)](#). Then, filtered replicate scans for each sample were averaged prior data analysis. Samples analysed with the DS2500 were finely ground in liquid nitrogen and stored at –80 °C from collection until the moment of the analysis. On the day of the analysis, samples were defrosted at room temperature (about one hour) and in the dark. Approximately 5 g of sample was used to fill the entire surface of a specific small ring cup with a quartz-glass window (50 mm diameter and 10 mm thickness) used for the readings. Each spectrum obtained with the DS2500 tool represented the average of 32 sub-spectra acquired during the automatic rotation of the cup. Two replicates were made on each sample, and these were averaged before data analysis. All spectral data were collected as reflectance and converted into absorbance expressed as log₁₀(1/Reflectance).

2.3. Chemometric procedure and statistical analysis

Data pre-processing techniques were applied to the absorbance datasets, to reduce noise effect and variability incurred by the light scatter effect. The effect of the different data pre-treatments, including mathematical treatments (derivatives), smoothing and scatter correction techniques (standard normal variate and multiplicative scatter correction) and their combinations was evaluated in terms of overall classification errors on the models (data not reported). The best choice for all the spectral data analysed in this study was determined as first Norris derivative - 1,4,4,1 ([Hopkins, 2001](#)) followed by Savitzky-Golay polynomial smoothing ([Savitzky & Golay, 1964](#)) of a 3-order polynomial and 11-point (FlameNIR and DS2500 datasets) or 31-point

(IndiGo dataset) window. In the mathematical calculation of first derivative 1,4,4,1, the first digit indicates the derivative order, the second digit indicates the gap size, the third digit indicates the segment size and the fourth digit indicates the repetition factor. After data pre-treatment, a down-sampling step using a factor of 3 or 4 was performed on the datasets obtained from the IndiGo device (from 0.5 nm to 2 nm, 481 to 120 data points) and the DS2500 device (from 2 nm to 6 nm, 1050 to 259 data points), in order to reduce dimensionality of dataset, decrease collinearity between adjacent wavelength and speed up the following computations without losing signal information.

Samples were labeled using Class0 to Class5 as determined in [Section 2.1](#). For each dataset, the algorithm developed by [Kennard & Stone \(1969\)](#) was used to split the samples into the calibration (70%) and validation (30%) subsets. The data were split balancing the proportion of samples from each class in both in the calibration and in the validation sets, ensuring that all classes were represented in the same proportions as the original dataset. Two dataset split approaches were considered in this study. In the first case, the farm of origin of the sample was not considered, to focus specifically on the ability of the model to classify individual samples in the corresponding class. In this case (Animal-split), animals coming from the same farm were split in both the calibration and the validation set, and the experimental unit was represented by the individual ($N = 497$). In the second case, the effect of the farm of origin was taken into account, to check whether samples coming from a hypothetical external farm with respect to the ones used for training the model would be correctly classified by the same algorithm as the individual samples. In this case (Farm-split), the experimental unit was the farm ($N = 55$), and samples collected in 41 farms were combined into the calibration set, while samples collected in the remaining 14 farms were combined into the validation set. First, an exploratory analysis was performed to observe the existence of natural groupings. For this purpose, principal component analysis was applied to the overall spectral data set. After that, LDA was used as a classification method. A first round of LDA was performed on the whole dataset, including all the spectral variables. Afterward, the Covariate Selection (CovSel) algorithm was tested for relevant feature selection before LDA ([Roger et al., 2011](#)). Briefly, the CovSel algorithm is an iterative process based on the calculation of the covariance matrix between the original variables: it selects the most informative feature among the group of correlated variables and removes others, to avoid redundancy. In the case of discrimination analysis, the covariance maximized by the CovSel algorithm is a compromise between X variance (spectral variables in the dataset), Y variance (response variable, in our case the class) and their

correlation, identifying the variables of X which explain Y the most (Roger et al., 2011). The calibration subset from each dataset was used in the variable selection step. In each subset, the CovSel algorithm was routinely applied to filter 1 to 100 original variables, in order to retain the maximum number of original variables representative of the original spectra, avoiding collinearity in the X matrix. The best CovSel variables threshold was selected on the basis of a 10-fold cross-validation (CV) performed on the calibration set, selecting the number of CovSel variables which maximized the classification accuracy in CV at the knee point (Supplementary Materials, Fig. S1). Consequently, the validation set was used for the predictions and model validation. Confusion matrices were used to evaluate the performance of the individual models, along with the sensitivity, specificity, precision, F1 score, false discovery rate (FDR) and balanced accuracy (BAC) (Opateye et al., 2025). These metrics have been calculated as it follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN};$$

$$\text{Specificity} = \frac{TN}{FP + TN};$$

$$\text{Precision} = \frac{TP}{TP + FP};$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}};$$

$$\text{FDR} = 1 - \text{Precision};$$

$$\text{BAC}_k = \frac{\text{Sensitivity}_k + \text{Specificity}_k}{2};$$

$$\text{BAC} = \frac{1}{K} \sum_{k=1}^K \text{BAC}_k$$

where TP = true positives, FN = false negatives, FP = false positives; TN

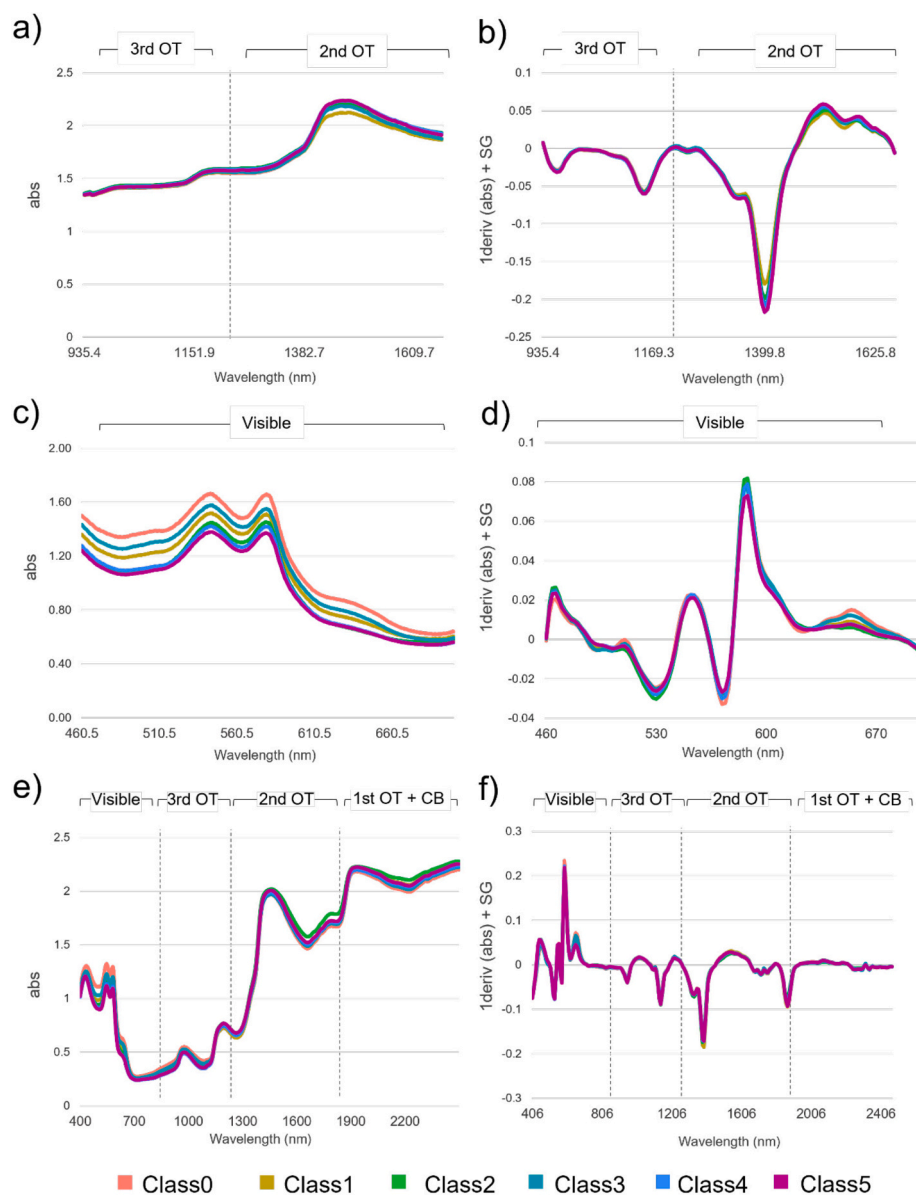


Fig. 1. Average spectrum for the FlameNIR (a: raw absorbance; b: first derivative and smoothing), IndiGo (c: raw absorbance; d: first derivative and smoothing) and DS2500 (e: raw absorbance; f: first derivative and smoothing) devices. Abs = absorbance; 1deriv = first derivative; SG = Savitzky-Golay smoothing; OT = overtones; CB = combination bands. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

= true negatives; k = k class; K = number of classes. The accuracy for each class, along with individual metrics and the confusion matrix, was also assessed by analysing the receiver operating characteristic (ROC) curves and area under the curve (AUC) for each class (Flach et al., 2003). Permutation tests on the final models (validation) for the balanced accuracy, misclassification and AUC metrics were performed on the validation sets with thousand permutations for each model, in order to reduce the risk of overfitting. Statistical analysis was performed using the following packages of R (v 4.4.1) run in RStudio (v 2024.09.0): *caret*, *MASS*, *mixOmics*, *prospectr* and *signal*.

3. Results

3.1. Vis/NIR spectra

Fig. 1 shows the overlapped spectra for each class obtained by the FlameNIR, IndiGo and DS2500 devices, raw (expressed as absorbance) (Fig. 1a, c, e) and after application of the first derivative and Savitzky-Golay smoothing (Fig. 1b, d, f). Although spectra for the different levels of intensification showed a similar trend, differences may be

observed among spectra obtained through the three devices. Spectra from FlameNIR, IndiGo, and DS2500 revealed characteristic absorption bands associated with the main components of meat. In the first derivative spectra, the FlameNIR (Fig. 1b) showed the maximum peaks at 1017 nm, 1216 nm, 1256 nm, 1354 nm, 1519 nm, 1587 nm and 1626 nm, and the minimum peaks at 965 nm, 1152 nm, 1239 nm, 1348 nm, 1400 nm, 1564 nm and 1620 nm. The IndiGo showed maximum peaks at 466 nm, 496 nm, 510 nm, 554 nm, 588 nm and 654 nm, and minimum peaks at 494 nm, 502 nm, 530 nm, 572 nm, and 626 nm (Fig. 1d). The DS2500 showed maximum peaks at 454 nm, 558 nm, 590 nm, 654 nm, 1022 nm, 1110 nm, 1222 nm, 1350 nm, 1542 nm, 1694 nm, 1734 nm, 1806 nm, and minimum peaks at 534, 574, 522, 950, 1142, 1334, 1390, 1870 nm, plus several narrower shifting in the region ranging from 2000 to 2500 nm (Fig. 1f).

3.2. Discrimination of the different intensification levels

Initial classification analysis was performed using LDA on the complete spectral datasets for the three devices, and the results are reported in Supplementary Table S1. Fig. 2 shows the CovSel variables selected

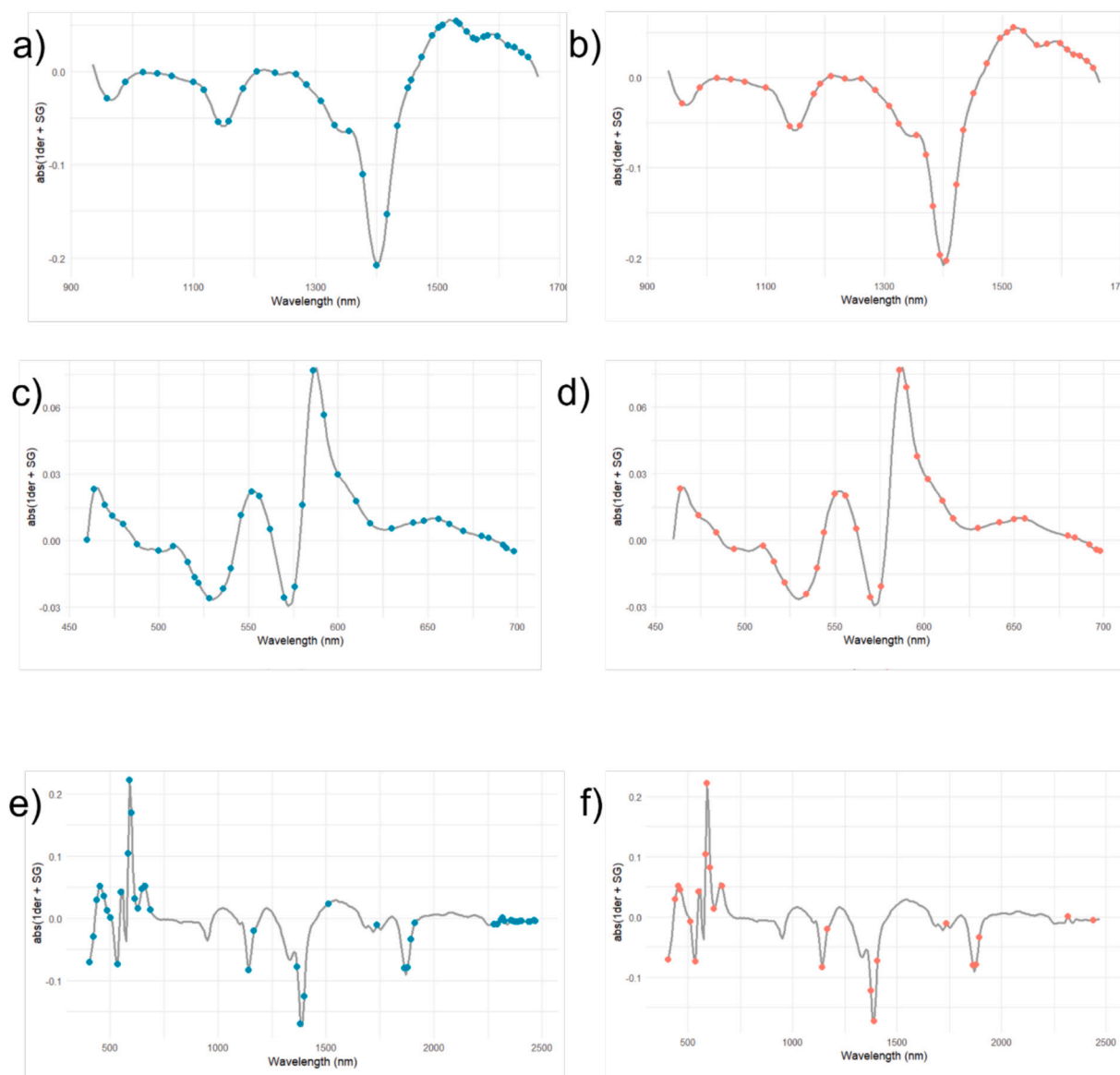


Fig. 2. CovSel variables highlighted for the FlameNIR (a: Animal-split; b: Farm-split), IndiGo (c: Animal-split; d: Farm-split) and DS2500 (e: Animal-split, f: Farm-split) devices. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for each of the three devices, highlighted on the average spectra (raw and after first-derivative calculation and Savitzky-Golay smoothing) and calculated using both Animal- and Farm-split approaches. Results from the 10-fold CV performed on the calibration set during the tuning of the model led to the choice of 39 and 37 features from FlameNIR (Animal- and Farm-split, respectively), 37 and 30 features for IndiGo, 44 and 23 features for DS2500. The variables selected by the CovSel algorithm were distributed along the entire spectra for the FlameNIR (Fig. 2 a and b) and IndiGo (Fig. 2 c and d) devices, while they were more gathered in the first (400–700 nm) and last (1700–2500 nm) segments of the spectrum for the DS2500 (Fig. 2 e and f), with fewer wavelengths selected in the intermediate region (700–2300 nm).

The overall performance of the LDA models calculated using the CovSel-selected wavelengths is shown in Table 2. For all three devices, the model built on the dataset using the Animal-split demonstrated comparable metrics in calibration but higher in validation than the model built using the Farm-split. With the Animal-split, the validation BAC was 0.87 for FlameNIR, and 0.89 for both IndiGo and DS2500. With the Farm-split, the validation BAC was 0.74 for FlameNIR, 0.79 for IndiGo and 0.85 for DS2500. In this scenario, the three devices showed comparable performances when using the Animal-split. However, the models built using spectral data collected by the DS2500 showed the lowest gap in performances between the two dataset split approaches, and reaching the highest sensitivity, specificity, precision, F1 score and BAC (and the lowest FDR) when using the Farm-split. The FlameNIR showed the lowest performances with both the dataset split approaches tested and with the highest gap observed between the Animal- and the Farm-split. Finally, the IndiGo showed equal performances than the DS2500 for the Animal-split, while, when switching to the Farm-split, intermediate performances between DS2500 and FlameNIR were observed for this device.

Tables 3 and 4 show the metrics per group and the confusion matrices obtained by the CovSel-LDA model for the three devices on the Animal- and Farm-split, respectively, and Fig. 3 shows the AUC-ROC curves for each class. Metrics and confusion matrices obtained in calibration for the models are presented in Supplementary Material Tables S2 and S3. Regarding the Animal-split, excellent performance across all devices was observed with AUC values ranging from 0.90 to 0.99. When switching to the Farm-split, a moderate to substantial performance decrease was observed, with the AUC values falling up to 0.72–0.79 for Class4 and Class3 with the FlameNIR (Fig. 3 b) and the IndiGo (Fig. 3 d). Some classes were more predictable than others. Overall, Class0 showed the highest number of correctly classified samples, reaching 83.3% (IndiGo), 86.7% (FlameNIR) and 100.0% (DS2500) precision with the Animal-split (Table 3), while less correct classifications were observed with the Farm-split (Table 4). Particularly, for the FlameNIR, these

values showed a significant decrease, from 86.7% to 46.7%, from the Animal- to the Farm-split. Class5 showed good and consistent precision, ranging from 83.3 to 85.7% in the Animal-split (Table 3) to 60.5–89.7% in the Farm-split (Table 4). This was the class that showed the less variability among split approaches and devices. Class3 was relatively stable, reaching 82.8% to 89.7% precision for the DS2500 and IndiGo devices respectively with the Animal-split (Table 3), and decreasing to 71.4% and 77.8% with the Farm-split (Table 4). Higher gap between the two split approaches was observed for the precision of Class3 using FlameNIR, decreasing from 89.7% (Animal-split; Table 3) to 70.0% (Farm-split; Table 4). On the other hand, Class1 and Class2 showed unstable behavior depending on the spectral tool and the dataset split approach. For Class1, classification precision ranged from 60.0% (DS2500) to 85.7% (FlameNIR and IndiGo) in Animal-split (Table 3), while from 70.0% (DS2500) to 80.0% (FlameNIR) in Farm-split (Table 4). For Class2, classification precision ranged from 53.3% (FlameNIR) to 93.3% (DS2500) with Animal-split (Table 3) and increased from 80.0% (FlameNIR and IndiGo) to 100.0% (DS2500) in Farm-split (Table 4). Finally, Class4 showed moderate to good precision in the Animal-split approach (58.3% to 79.2% depending on the device; Table 3), but a drastic decrease was observed for all the devices when switching to the Farm-split, where 20.0% to 47.4% precision was observed (Table 4).

Supplementary Fig. S2-S4 show the results of the permutation tests for both dataset split approaches of the final models. An empirical *p*-value of 0.009 was observed for all the models (devices, split approaches and metric considered).

4. Discussion

4.1. Characteristics of the three devices and data analysis

The absorption bands observed in the spectra from the three devices tested in this study reflected contributions from water, protein, and lipid components, confirming trends widely documented in the literature (Cozzolino & Murray, 2004). In the Vis range, absorption bands between 480 and 580 nm could correspond to myoglobin- and oxymyoglobin-related regions (León-Ecay, López-Campos, et al., 2024). These absorption bands were observed in both IndiGo and DS2500 devices. In the NIR range, the spectral bands observed around 1450–1468 nm for both FlameNIR and DS2500 could be associated with O—H overtones associated with moisture (Andrés et al., 2008). Additionally, DS2500 revealed combination bands related to water absorption near 1935 nm (Alomar et al., 2003). Protein-related N—H overtones were evident in the NIR region between 1190 and 1510 nm for both FlameNIR and DS2500 (Dixit et al., 2017). For these devices, fat bands were detected in

Table 2

Metrics in calibration (Cal) and validation (Val) for the three devices (FlameNIR, 935–1665 nm; IndiGo, 460.5–700 nm and DS, 400–2500 nm) and the two dataset split approaches (Animal-split and Farm-split).

Dataset split approach	Metric	Device					
		FlameNIR		IndiGo		DS2500	
		Cal	Val	Cal	Val	Cal	Val
Animal	Sensitivity	0.81	0.78	0.85	0.82	0.89	0.82
Farm		0.86	0.56	0.84	0.66	0.83	0.73
Animal	Specificity	0.96	0.96	0.96	0.96	0.98	0.96
Farm		0.97	0.91	0.97	0.93	0.97	0.96
Animal	Precision	0.82	0.77	0.83	0.78	0.90	0.81
Farm		0.86	0.63	0.84	0.67	0.83	0.78
Animal	F1 score	0.81	0.77	0.84	0.79	0.89	0.81
Farm		0.86	0.56	0.84	0.65	0.83	0.72
Animal	FDR	0.18	0.23	0.17	0.22	0.10	0.19
Farm		0.14	0.37	0.16	0.33	0.17	0.22
Animal	BAC	0.89	0.87	0.91	0.89	0.93	0.89
Farm		0.91	0.74	0.90	0.79	0.90	0.85

BAC = Balanced Accuracy; FDR = False Discovery Rate.

Table 3

Metrics and confusion matrix for the validation of the CovSel-LDA model obtained using spectral data from the three devices using the Animal-split approach.

Device: FlameNIR (39 CovSel variables)													
Validation metrics	Class						Confusion matrix	Predicted Class in Validation (%)					
	Class0	Class1	Class2	Class3	Class4	Class5		True Class	Class0	Class1	Class2	Class3	Class4
Sensitivity	0.72	0.75	0.73	0.84	0.82	0.83	Class0	86.7	7.1	13.3	3.4	8.3	11.4
Specificity	0.96	0.98	0.95	0.97	0.92	0.95	Class1	3.3	85.7	6.7	0	8.3	0
Precision	0.87	0.86	0.53	0.90	0.58	0.86	Class2	3.3	0	53.3	3.4	4.2	0
F1 score	0.79	0.80	0.62	0.87	0.68	0.85	Class3	0	7.1	0	89.7	16.7	0
FDR	0.13	0.14	0.47	0.10	0.42	0.14	Class4	0	0	13.3	0	58.3	2.9
Sample size (N)	30	14	15	29	24	35	Class5	6.7	0	13.3	3.4	4.2	85.7
Device: IndiGo (37 CovSel variables)													
Validation metrics	Class						Confusion matrix	Predicted Class in Validation (%)					
	Class0	Class1	Class2	Class3	Class4	Class5		True Class	Class0	Class1	Class2	Class3	Class4
Sensitivity	0.93	0.80	0.99	0.80	0.76	0.65	Class0	83.3	0	0	6.9	0	0
Specificity	0.96	0.98	0.96	0.93	0.96	0.94	Class1	0	85.7	0	6.9	0	2.8
Precision	0.83	0.86	0.67	0.69	0.79	0.83	Class2	0	0	66.7	0	0	0
F1 score	0.88	0.83	0.80	0.74	0.78	0.73	Class3	10	0	0	69	0	5.6
FDR	0.17	0.14	0.33	0.31	0.21	0.17	Class4	0	7.1	0	6.9	79.2	8.3
Sample size (N)	30	14	15	29	24	36	Class5	6.7	7.1	33.3	10.3	20.8	83.3
Device: DS2500 (44 CovSel variables)													
Validation metrics	Class						Confusion matrix	Predicted Class in Validation (%)					
	Class0	Class1	Class2	Class3	Class4	Class5		True Class	Class0	Class1	Class2	Class3	Class4
Sensitivity	0.99	0.82	0.82	0.86	0.64	0.79	Class0	100	0	0	0	0	0
Specificity	0.99	0.96	0.99	0.96	0.94	0.95	Class1	0	60	0	6.9	0	0
Precision	0.99	0.60	0.93	0.83	0.67	0.84	Class2	0	13.3	93.3	0	0	2.7
F1 score	0.99	0.69	0.88	0.84	0.65	0.82	Class3	0	20	0	82.8	4.2	0
FDR	0.00	0.40	0.07	0.17	0.33	0.16	Class4	0	6.7	6.7	6.9	66.7	13.5
Sample size (N)	29	15	15	29	24	37	Class5	0	0	0	3.4	29.2	83.8

FDR: False Discovery Rate.

Table 4

Metrics and confusion matrix for the validation of the CovSel-LDA model obtained using spectral data from the three devices using the Farm-split approach.

Device: FlameNIR (37 CovSel variables)													
Validation metrics	Class						Confusion matrix	Predicted Class in Validation (%)					
	Class0	Class1	Class2	Class3	Class4	Class5		True Class	Class0	Class1	Class2	Class3	Class4
Sensitivity	0.78	0.47	0.36	0.53	0.47	0.77	Class0	46.7	10	0	10	0	0
Specificity	0.86	0.98	0.99	0.90	0.90	0.85	Class1	13.3	80	0	6.7	15	0
Precision	0.47	0.80	0.80	0.70	0.40	0.61	Class2	6.7	0	80	6.7	0	7.9
F1 score	0.58	0.59	0.50	0.60	0.43	0.68	Class3	10	10	20	70	40	15.8
FDR	0.53	0.20	0.20	0.30	0.60	0.39	Class4	10	0	0	0	40	15.8
Sample size (N)	30	10	5	30	20	38	Class5	13.3	0	0	6.7	5	60.5
Device: IndiGo (30 CovSel variables)													
Validation metrics	Class						Confusion matrix	Predicted Class in Validation (%)					
	Class0	Class1	Class2	Class3	Class4	Class5		True Class	Class0	Class1	Class2	Class3	Class4
Sensitivity	0.95	0.78	0.53	0.70	0.36	0.61	Class0	70	0	0	3.7	0	0
Specificity	0.92	0.98	0.98	0.94	0.87	0.89	Class1	0	77.8	0	3.7	0	2.5
Precision	0.70	0.78	0.80	0.78	0.20	0.75	Class2	3.3	0	80	0	10	10
F1 score	0.81	0.78	0.64	0.74	0.26	0.67	Class3	20	11.1	0	77.8	0	5
FDR	0.30	0.22	0.20	0.22	0.80	0.25	Class4	3.3	0	0	11.1	20	7.5
Sample size (N)	30	9	10	27	20	40	Class5	3.3	11.1	20	3.7	70	75
Device: DS2500 (23 CovSel variables)													
Validation metrics	Class						Confusion matrix	Predicted Class in Validation (%)					
	Class0	Class1	Class2	Class3	Class4	Class5		True Class	Class0	Class1	Class2	Class3	Class4
Sensitivity	0.83	0.54	0.42	0.89	0.90	0.83	Class0	92.3	0	0	11.4	0	2.6
Specificity	0.98	0.98	1.00	0.91	0.92	0.96	Class1	0	70	0	14.3	5.3	0
Precision	0.92	0.70	1.00	0.71	0.47	0.90	Class2	0	20	100	2.9	10.5	5.1
F1 score	0.87	0.61	0.59	0.79	0.62	0.86	Class3	7.7	10	0	71.4	0	0
FDR	0.08	0.30	0.00	0.29	0.53	0.10	Class4	0	0	0	0	47.4	2.6
Sample size (N)	26	10	5	35	19	39	Class5	0	0	0	0	36.8	89.7

FDR: False Discovery Rate.

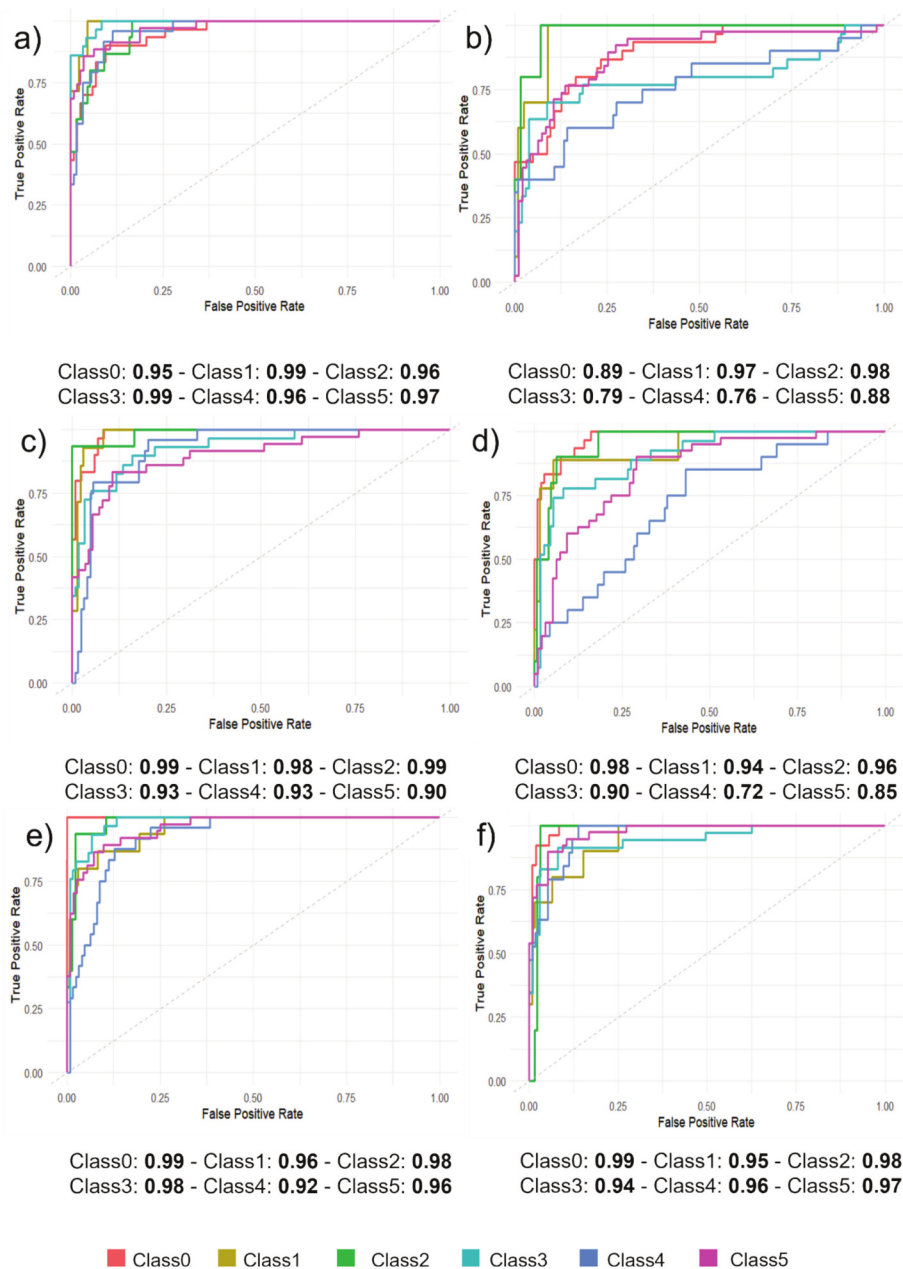


Fig. 3. ROC Curves and AUC values (under each curve) of the five intensification classes (Class0 to Class5) for the FlameNIR (a: Animal-split; b: Farm-split), IndiGo (c: Animal-split; d: Farm-split) and DS2500 (e: Animal-split, f: Farm-split) devices. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the NIR region around 1200 nm (Sierra et al., 2008). Further, the bands beyond 2300 nm detected by DS2500, could align with C—H and C=C groups of fatty acids combination tones (Cozzolino & Murray, 2004).

When LDA was first applied to the raw dataset without prior variable selection, the model showed apparently excellent calibration performance (calibration BAC > 0.95), yet the marked difference between calibration and validation metrics indicated overfitting (validation BAC 0.63–0.94). Moreover, the intrinsic nature of NIR data poses its own challenge because collinearity among the variables is very high, leading to stability problems when some multivariate models, such as LDA, are applied (Næs & Mevik, 2001; Roger et al., 2011). Thus, the CovSel algorithm proposed by Roger et al. (2011) was used to identify the most informative and least redundant wavelengths. This variable reduction technique specifically targets the collinearity problem inherent in spectral data by selecting a subset of wavelengths that maximizes the

covariance between the spectral matrix and the response variable while minimizing redundancy among selected variables. This approach not only mitigates overfitting when the LDA technique is used but also improves model interpretability by identifying the specific spectral regions most relevant for discriminating between classes. The sets of CovSel-selected wavelengths for the three devices tested in this study covered different spectral ranges, but in all cases, very few differences were observable between the two dataset split approaches for the specific wavelengths selected. Interestingly, in the DS2500 datasets the variables selected by the CovSel algorithm included only 6 over 44 (Animal-split) wavelengths and 5 over 23 (Farm-split) wavelengths in the FlameNIR range (Fig. 2). At the same time, 13 (Animal-split) and 8 (Farm-split) wavelengths among the CovSel variables selected in the DS2500 were in the IndiGo range. This finding may be explained by differences in the instrumental characteristics of the spectral devices. Since the DS2500

operates across a substantially broader spectral range individual wavelengths in the reduced NIR region 1100–1600 nm may have provided redundant information relative to the FlameNIR dataset, potentially introducing multicollinearity that prompted their exclusion from the CovSel algorithm and the selection of the optimal variables number step. Consequently, the extended spectral coverage of the DS2500 likely enabled the classification algorithm to identify discriminant features distributed throughout the spectrum. At the same time, the selection of several wavelengths in the Vis (400–700 nm) and extended NIR region (1600–2500 nm) in the DS2500 datasets suggested that such spectral regions had a significant contribution in driving the discrimination obtained through the CovSel-LDA models, confirmed by the comparable classifications obtained using the DS2500 and the IndiGo devices.

4.2. Performance of the different protocols and devices

The application of CovSel-LDA protocol for discriminating beef samples according to farm intensification level revealed differences in the classification performances based on spectral device, sample preparation protocols, and the dataset split approach considered. Notably, the LDA model applied under the Animal-split yielded comparable performances across the three devices, with the IndiGo and DS2500 achieving identical accuracy value and the FlameNIR scoring only marginally lower. However, when switching to the Farm-split, the DS2500 maintained consistent accuracy with that obtained through the Animal-split, whereas both the IndiGo and the FlameNIR showed a notable decline, reaching lower performances if compared to the DS2500. These differences are likely attributable to multiple methodological factors along with instrumental specifications. Each tool is defined by a series of technical characteristics that directly influence spectral performances and field applicability, including wavelength range, data point interval, wavelength accuracy, spectral resolution, and so on. The three devices tested in this study differed substantially in their spectral coverage. In this scenario, only the DS2500 captured vibrational signals across the first, second, and third vibrational overtone regions, as well as combination tones and signals within the Vis range, whereas the FlameNIR covered only the second and third overtone regions, and the IndiGo operated exclusively in the Vis range (Fig. 1). Moreover, the shorter data-point interval of DS2500 and IndiGo compared to the FlameNIR may have contributed to their superior classification performance. Indeed, a denser spectral sampling allows capturing signals from beef traits not necessarily linked to specific vibrational properties and generally benefit of a larger number of wavelengths collected even to develop accurate calibration equations, as reported by Patel et al. (2021). Consistent with this, in a parallel study (Lopez et al., n.d.), we observed superior performances of the DS2500 in predicting the biochemical composition of beef samples relative to both the FlameNIR and the IndiGo, while the latter showed a strong predictive ability (comparable to the DS2500's one) for molecules associated with respiratory pigments, whose vibrational signals are expressed in the Vis range.

Another key differential factor was the variation in the sample preparation protocol between spectral recordings with different devices. While measurements with the FlameNIR and the IndiGo were performed on intact meat surfaces, DS2500 analyses required prior to grinding and homogenization of samples. This distinction is highly significant, as spectral absorbance intensity is known to differ between intact and homogenized matrices due to differences in light-scattering behavior. Literature findings regarding the effects of meat sample preparation (intact vs homogenized) are conflicting, with an average higher efficiency recorded when working on intact samples. Some authors reported superior prediction of meat quality traits and detection ability of dark-cutting beef using intact steaks rather than ground meat samples (De Marchi et al., 2013; Prieto et al., 2014), attributing this advantage to the fact that homogenization severely disrupts muscle architecture (through fiber destruction and randomization), eliminating structural

attributes (such as colour, texture, juiciness) and classification-relevant information. However, the same authors reported that the structural complexity of intact meat, generating substantial light scattering, makes the full-range Vis/NIR instrumentation needed to achieve adequate classification results, rather than instruments covering only specific Vis or NIR ranges. At the same time, León-Ecay, Insausti, et al. (2024) obtained better classification of beef from grass-, barley- and corn-fed cattle using ground rather than intact samples, attributing this result to the reduction of light scattering resulting from the homogenization step. Further and last distinguishing factor in the analysis protocols of the devices tested in this study was the differing approaches to measurement replicates. Namely, being applied on the intact meat samples, the FlameNIR and IndiGo protocols incorporated ten replicate scans per sample, to cover as many surface locations as possible. Consequently, replicate filtering followed by detection and removal of uneven spectra before averaging in each sample was performed, assuring that representative mean spectra were obtained from the replicate readings (Martínez et al., 1998). Conversely, the DS2500 protocol, which was applied to finely ground meat samples (higher homogeneity of the matrix), included only two replicates for ground meat samples, which were averaged before data analysis. The variations in spectral analysis protocols, including sample treatment (intact vs ground meat) and measurement replicates (ten vs two), were fundamentally different factors that may have influenced the performance of the chemometrics models along with the infrared technology (different devices and spectral range).

Regardless of the device used, classification accuracy decreased when transitioning from the Animal- to the Farm-split approach, with the only exception of FlameNIR in calibration. Higher gap between calibration and validation performances was observed with the Farm-split compared to the Animal-split. This pattern indicated in general, that while the models performed well on the collected dataset, their generalizability to samples from farms not represented in the calibration set was reduced. This decrease in performance when using the Farm-split could be attributable to two key factors. First, this approach resulted in a data partition that consisted of only $N = 15$ farms included in the validation set. This relatively small number of validation farms may have limited the accuracy of performance estimates. Second, the inherent challenge of the external validation strategy, in which samples from the validation farms were completely absent from the calibration data, prevented the model from learning any farm-specific characteristics that might influence spectral profiles. This represented a more realistic assessment of the model performance in practical applications, where predictions must be made on samples from previously unseen farms. However, the Farm-split, while yielding lower accuracy, provides a more rigorous assessment of model generalization to samples from previously unseen production systems. This dataset split approach ensures that farm-specific signatures, which may reflect factors such as management practices, local feed sources, or environmental conditions, do not influence performance estimates. In contrast, the Animal-split, though achieving higher accuracy, included samples from the same farms in both calibration and validation sets. This allows the model to potentially leverage farm-specific patterns, which can increase performance estimates because the validation samples resemble those used for calibration. Actually, this effect was expected, since when calibration and validation datasets share samples from the same origin (batch, farm, country), model statistics may be overoptimistic as the results may reflect relationships specific to that dataset rather than genuine generalizability to new samples or conditions (Patel et al., 2021). In this context, the Farm-split approach represents a more conservative yet more realistic benchmark for practical application, reflecting scenarios in which the farms present in the validation set are absent from the calibration set.

4.3. Discrimination of beef samples according to intensification levels

The experimental design presented in this study was characterized by notable classification challenges: the six-class structure represents a relatively complex discrimination strategy, particularly given that several classes shared substantial similarities in their underlying characteristics. Some farming systems under investigation exhibited overlapping features that naturally resulted in less distinct spectral boundaries between some specific classes, reflecting the real-world continuum of agricultural practices. Results from the LDA models showed differences in the classification performance metrics for the different classes included in this study (Class0 to Class5). Overall, the best results were obtained for Class0 and Class5 across different devices. Class0 represented the most distinct class in the dataset in terms of characteristics of slaughtered animals, being characterized by FR suckler cows fattened before culling (90 months, 827 kg), supplied medium-high concentrate levels (median: 7.3 kg/day). Age, breed and genotype/breed, are known to be major factor of variations in several quality traits for meat, encompassing both sensory properties (including colour) and chemical composition, including water holding capacity, intramuscular fat content, collagen solubility and meat structure, with meat from younger animals being lighter, tender and leaner, with higher moisture content (Clinquart et al., 2022; Maltin et al., 1998; Moloney et al., 2022; Scollan et al., 2014). In addition to intrinsic factors, environmental and management factors, including animal feeding strategies, play an uncontroversial role in determining meat quality, with particular influence on fatty acids composition, with grass-fed cattle (with no concentrates) showing higher polyunsaturated fatty acid content, especially α -linolenic acid content, and, consequently, lower n-6/n-3 ratios (Nogoy et al., 2022; Prache et al., 2020; Varre et al., 2025). It was previously reported that differences in spectral features are observable between different cattle diets, with meat from grass-fed cattle vs grain- or cereal-fed cattle being correctly discriminated by Vis, NIR and combined Vis/NIR spectral information obtained through portable devices and discriminant models (Coombs et al., 2021; León-Ecay, Insausti, et al., 2024). The combination of advanced age and weight at slaughter and high concentrate intake have probably generated a cumulative spectral signature that placed Class0 sufficiently far from the other classes in the spectral features space. Most misclassified samples from Class0 were classified as Class3. Such misclassification patterns observed for these classes could be partially explained by similarities in their chemical composition, given that these two classes were represented by grazing sucklers and beef >2 years supplemented by medium to high amounts of concentrates, sometimes overlapping, which could have led to a comparable chemical composition (fat content, lipid profile) detected through spectral analysis. Actually, Class3 was characterized by UK yearlings fattened until 28 months on grasslands with a moderate level of concentrate input (median: 2.5 kg). Class5 represented the more consistent class when switching from the Animal- to the Farm-split approach. Class5 (IT, ES and CH farms) was represented by animals fattened in indoor feedlots until 14 months of age (550 kg), and with the highest proportion of high-energy feed inputs (median: 8.2 kg/day), including concentrates and maize silage. Due to this characteristic, Class5 was probably distinguished by a specific spectral signature associated to the elevated high energy feed inputs. Moreover, highly standardized management practices typical of feedlot system (including the high-energy diet and the uniform slaughter age) could have contributed to a relatively homogeneous sample, nevertheless the multi-country origin of samples from this class (IT, ES and CH). This could suggest that feedlot practices converged toward similar spectral profiles regardless of geographic origin, and that this diversity may reinforce intra-class consistency (even more than Class0) and therefore inter-class separability. Even Class3 was relatively stable to the switching dataset split approach, showing similar classification in both the Animal- and the Farm-split, and without showing a misclassification pattern toward any specific class. This overlaps with the other classes suggested that,

although the model captured differences between this grass-based system from UK, similarities with other systems (considering both animal and management factors) introduced confusion.

Across all models, a consistent proportion of samples from Class4 were misclassified as Class5. Class4 comprised animals from CH farms where animals were fattened indoor until 16 months of age (545 kg) with a median high-energy feed input of 5.8 kg/day. Thus, this class differed from Class5 for the lower amount of concentrates furnished to animals and for its mono-country origin. Under the Animal-split approach, the models performed reasonably well in recognizing specific patterns of medium vs high intensification levels. However, when switching to the Farm-split approach, the model could not generalize such differences across farms, which were characterized by different feed formulations and animal genetic types (beef, dairy and their crosses). This class was characterized as the one more sensitive to the farm effect and it could be supposed that, at the farm-level, spectral variations related to the farm effect may have overwhelmed the difference in spectral signal related to the grade of intensification. The presence of some differences within the same class led to a continuum rather than discrete separation between the two classes, and consequently to an overlap for some farms from these classes. In discrimination methods based on NIR spectroscopy, high variability in specific chemical entities among the classes of interest leads to better model accuracy (Alomar et al., 2003), which was not the case for these two closely related intensive systems. Chemical similarity between the two classes could have likely contributed to the classification confusion. As evidenced above, meat from cattle fed high levels of concentrates, especially corn, is chemically distinguishable from meat from grass-fed cattle and such a difference can be captured by spectral analysis. Grain-fed beef meat from indoor systems is usually characterized by higher concentrations of intramuscular fat content, higher percentages of saturated fats and higher proportions of n-6 fatty acids (Varre et al., 2025). Moreover, grain-fed beef meat was frequently observed to be lighter in colour than beef from grazing cattle, which most likely reflects changes in muscle characteristics (increase of myoglobin concentration) because of the physical activity of grazing cattle (Moloney et al., 2022) and increased marbling (Klopatek et al., 2022).

Class1 and Class2 represented the most challenging classes in the model, sharing a common biological background. Both comprised animals raised exclusively at pasture in CH grasslands, without concentrates, and differing exclusively in slaughter age and weight (11 months, 400 kg for Class 1 and 23 months, 585 kg for Class2) and genetic type (dairy for Class1 and dairy/beef crosses for Class2). This proximity was somewhat reflected in the classification performances: Class1 was the class with the highest degradation when switching from the Animal- to the Farm-split approach and showing the highest misclassification spread over the adjacent classes. Class2, in contrast, displayed a pattern completely different from all the other classes. Under the Animal-split, it showed the greatest inter-device variability in the entire dataset, and when switching to the Farm-split the performance improved across all the devices. Taken together, these patterns suggested that the spectral signature distinguishing these classes was intrinsically weak and that separability of these samples from the other classes depended critically on the sampling structure. Moreover, the reduced sample size for these two classes, ranging from 34 to 45 samples in calibration and 5 to 15 in validation, cannot be overlooked. This reduced sample size could have substantially impacted model performance, as so few samples in calibration may be insufficient to adequately capture the spectral signature within each class and it could have been overwhelmed by intra-class variability.

5. Conclusions

The findings of this study showed that the employment of CovSel feature selection algorithm followed by Linear Discriminant Analysis can be an effective strategy for classifying beef samples according to the

intensification level of the farming system through Vis, NIR, and combined Vis/NIR spectral data. When the farm is considered as the experimental unit of the model (Farm-split approach), the classification performance decrease, suggesting the driving influence of specific farm characteristics. To mitigate this effect, future studies should refine the experimental design by increasing the number of farms included in each intensification class and Country: a more balanced and representative distribution of farms across calibration and validation sets would improve the robustness of the model and provide a more reliable assessment of classification performance. Accordingly, future studies would benefit from including farms representing different intensification systems within more restricted geographical areas other than from different Countries, ensuring a balanced representation in both dimensions. The portable device operating in the Vis range (IndiGo) exhibited overall performances comparable to the benchtop one (DS2500) under the Animal-split approach, while comparable results for the two portable devices (the IndiGo and the FlameNIR, the latter working in the NIR range) were observed under the Farm-split approach. Although in the present study all three instruments were used under controlled laboratory conditions, fast and low-cost portable devices could offer the potential to acquire spectral measurements directly on-site, for example, on carcasses during grading or at processing facilities. Such an approach could substantially reduce time, costs, and analytical effort, while introducing innovation into authentication systems for the beef industry. Nonetheless, further research is required to evaluate the performance and robustness of these technologies under real industrial conditions.

Declaration of generative AI technologies

During the preparation of this work the author used the AI tool Claude.ai to refine language and improve clarity. All AI-generated content was critically reviewed and edited as needed. The authors remain fully responsible for the content of the publication.

Statement for studies in humans/animals

No ethical approval was needed for this study, as beef samples were collected in commercial abattoirs.

CRediT authorship contribution statement

Annalaura Lopez: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Elea Caumette:** Writing – review & editing, Validation, Resources, Investigation, Data curation. **Rennie C. Eppenstein:** Writing – review & editing, Resources, Conceptualization. **Nicolò Amalfitano:** Writing – review & editing, Resources. **Isabelle Legrand:** Writing – review & editing, Resources. **Lynda S. Perkins:** Writing – review & editing, Resources. **Nigel Scollan:** Writing – review & editing, Resources. **Pablo Guarnido-Lopez:** Writing – review & editing, Resources. **Isabelle Constant:** Writing – review & editing, Investigation, Data curation. **Fabien Ferlay-Demiaz:** Writing – review & editing, Investigation, Data curation. **Carmen L. Manuelian:** Writing – review & editing, Validation, Conceptualization. **Bruno Martin:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Jean-François Hocquette:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Donato Andueza:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Funding acquisition, Data curation, Conceptualization.

Ethical statement

No ethical approval was needed for this study, as beef samples were collected in commercial abattoirs.

Declaration of competing interest

We declare no competing financial interest or personal relationship that could have appeared to influence the work reported in this manuscript.

Acknowledgements

This work was supported by INTAQT project, funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 101000250, and the Watson project, funded by the European Union's Horizon 2020 Europe research and innovation program, under grant agreement No. 101084265. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. Annalaura Lopez's salary was supported by a research contract funded by University of Milan. Carmen L. Manuelian received funding from the Ramon y Cajal program of the Spanish Ministry of Science (Spain; Grant Ref. RYC2023-042902-I). Authors would like to thank the following people for their help with collecting samples in the abattoirs and farm survey data: Shannon McLaughlin (Queen's University, The United Kingdom) and David Morcuende Sanchez (University of Extremadura, Spain). We would like to thank also Verena Bühl and Maria Lozano Jaramillo (FiBL, Switzerland) for survey handling and finalisation, Anthony Barret (FiBL, Switzerland) for support in data management, Cyril Labonne and Paul Laissus (INRAE, France) for handling, preparing and scanning the samples, and all the abattoirs and farmers involved in the farm network who consented to answer the survey. Annalaura Lopez would like to thank Prof. Vittorio Maria Moretti (University of Milan, Italy) for his support in fostering the network that enabled her to establish this collaboration and the realization of this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meatsci.2026.110169>.

Data availability

Data used in this paper are available at the following link: <https://doi.org/10.57745/S6T3P8>.

References

- Alomar, D., Gallo, C., Castañeda, M., & Fuchslocher, R. (2003). Chemical and discriminant analysis of bovine meat by near infrared reflectance spectroscopy (NIRS). www.elsevier.com/locate/meatsci.
- Andrés, S., Silva, A., Soares-Pereira, A. L., Martins, C., Bruno-Soares, A. M., & Murray, I. (2008). The use of visible and near infrared reflectance spectroscopy to predict beef *M. longissimus thoracis et lumborum* quality attributes. *Meat Science*, 78(3), 217–224. <https://doi.org/10.1016/j.meatsci.2007.06.019>
- Barragan, W., Aalhus, J. L., Penner, G., Dugan, M. E. R., Juárez, M., López-Campos, V., ... Prieto. (2021). Authentication of barley-finished beef using visible and near infrared spectroscopy (vis-NIRS) and different discrimination approaches. *Meat Science*, 172. <https://doi.org/10.1016/j.meatsci.2020.108342>
- Caumette, E., Berton, M., Coppa, M., Eppenstein, R., Rey-Cadilhac, L., De Smet, S., ... Leiber, F. (2026). Livestock farming intensification: Mitigating or exacerbating feed-food competition? Contrasting results in broiler and dairy and beef productions. *Agricultural Systems*, 238, 104871. <https://doi.org/10.1016/j.agsy.2026.104871>
- Clinquart, A., Ellies-Oury, M. P., Hocquette, J. F., Guillier, L., Santé-Lhoutellier, V., & Prache, S. (2022). Review: On-farm and processing factors affecting bovine carcass and meat quality. *Animal*, 16. <https://doi.org/10.1016/j.animal.2021.100426>. Elsevier B.V.
- Coombs, C. E. O., Liddle, R. R., & González, L. A. (2021). Portable vibrational spectroscopic methods can discriminate between grass-fed and grain-fed beef. *Journal of Near Infrared Spectroscopy*, 29(6), 321–329. <https://doi.org/10.1177/09670335211049506>
- Cozzi, G., Knierim, U., Marti, S., Mullan, S., Ashe, S., Lima, E., ... Winckler, C. (2025). Common husbandry systems and practices for keeping beef cattle. *EFSA Supporting Publications*, 22(7). <https://doi.org/10.2903/sp.efsa.2025.EN-9565>

- Cozzolino, D., & Murray, I. (2004). Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. *LWT*, 37(4), 447–452. <https://doi.org/10.1016/j.lwt.2003.10.013>
- De Marchi, M., Penasa, M., Cecchinato, A., & Bittante, G. (2013). The relevance of different near infrared technologies and sample treatments for predicting meat quality traits in commercial beef cuts. *Meat Science*, 93(2), 329–335. <https://doi.org/10.1016/j.meatsci.2012.09.013>
- Dimitrakopoulou, M. E., & Vantarakis, A. (2023). Does traceability lead to food authentication? A systematic review from a European perspective. *Food Reviews International*, 39(1), 537–559. <https://doi.org/10.1080/87559129.2021.1923028>
- Dixit, Y., Casado-Gavaldà, M. P., Cama-Moncunill, R., Cama-Moncunill, X., Markiewicz-Keszycska, M., Cullen, P. J., & 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. Blackwell Publishing Inc <https://doi.org/10.1111/1541-4337.12295>
- Dumont, B., Dupraz, P., Sabatier, R., & Donnars, C. (2017). A collective scientific assessment of the roles, impacts, and services associated with livestock production systems in Europe. *Fourrages*, 229, 63–76. <https://hal.science/hal-01604662v1>
- Eurostat. (2025). Agricultural production - livestock and meat. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Agricultural_production_-_livestock_and_meat
- Flach, P., Blockeel, H., Ferri, C., Hernandez-Orallo, J., & Struyf, J. (2003). Decision support for data mining: An introduction to ROC analysis and its application. In D. Mladenic, N. Lavrac, M. Bohanec, & S. Moyle (Eds.), *Data mining and decision support*. Springer Science+Business Media, LLC.
- Hocquette, J. F., Elies-Oury, M. P., Lherm, M., Pineau, C., Deblitz, C., & Farmer, L. (2018). Current situation and future prospects for beef production in Europe - A review. *Asian-Australasian Journal of Animal Sciences*, 31(7), 1017–1035. Asian-Australasian Association of Animal Production Societies <https://doi.org/10.5713/jas.18.0196>
- Hopkins, D. W. (2001). What is a Norris derivative. *NIR News*, 12(3). <https://doi.org/10.1255/nirn.611>
- Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics*, 11(1), 137–148. <https://doi.org/10.1080/00401706.1969.10490666>
- Klopatek, S. C., Marvinney, E., Duarte, T., Kendall, A., Yang, X.(C.), & Oltjen, J. W. (2022). Grass-fed vs. grain-fed beef systems: Performance, economic, and environmental trade-offs. *Journal of Animal Science*, 100(2). <https://doi.org/10.1093/jas/skab374>
- Kombolo-Ngah, M., Goi, A., Santinello, M., Rampado, N., Atanassova, S., Liu, J., ... Hocquette, J. F. (2023). Across countries implementation of handheld near-infrared spectrometer for the on-line prediction of beef marbling in slaughterhouse. *Meat Science*, 200. <https://doi.org/10.1016/j.meatsci.2023.109169>
- León-Ecay, S., Insausti, K., Arazuri, S., Goenaga, I., & López-Maestresalas, A. (2024). Combination of spectral and textural features of hyperspectral imaging for the authentication of the diet supplied to fattening cattle. *Food Control*, 159. <https://doi.org/10.1016/j.foodcont.2024.110284>
- León-Ecay, S., Insausti, K., López-Maestresalas, A., & Prieto, N. (2025). Non-destructive spectroscopy-based technologies for meat and meat product discrimination - A review. *Meat Science*, 228. <https://doi.org/10.1016/j.meatsci.2025.109893>. Elsevier Ltd.
- León-Ecay, S., López-Campos, Ó., López-Maestresalas, A., Insausti, K., Schmidt, B., & Prieto, N. (2024). Using portable visible and near-infrared spectroscopy to authenticate beef from grass, barley, and corn-fed cattle. *Food Research International*, 198. <https://doi.org/10.1016/j.foodres.2024.115327>
- Lopez, A., Hocquette, J.F., Graulet, B., Constant, I., Labonne, C., Santé-Lhoutellier, V., De Smet, S., Mohuanna, A., Soglia, F., Estevez, M., Manuelian, C.L., De Marchi, M., Martin, B., & Andueza, D. (Unpublished results). Visible/Near infrared spectroscopy in muscle beef analysis: How device choice and sampling affect prediction accuracy.
- Maltin, C. A., Sinclair, K. D., Warriss, P. D., Grant, C. M., Porter, A. D., Delday, M. I., & Warkup, C. C. (1998). The effects of age at slaughter, genotype and finishing system on the biochemical properties, muscle fibre type characteristics and eating quality of bull beef from suckled calves. *Animal Science*, 66(2), 341–348. <https://doi.org/10.1017/S135772980009462>
- Martínez, M. L., Garrido-Varo, A., De Pedro, E., & Sánchez, L. (1998). Effect of sample heterogeneity on near infrared meat analysis: The use of the RMS statistic. In *Journal of Near Infrared Spectroscopy*, 6.
- Moloney, A. P., O' Riordan, E. G., Monahan, F. J., & Richardson, R. I. (2022). The colour and sensory characteristics of longissimus muscle from beef cattle that grazed grass or consumed concentrates prior to slaughter. *Journal of the Science of Food and Agriculture*, 102(1), 113–120. <https://doi.org/10.1002/jsfa.11337>
- Næs, T., & Mevik, B. H. (2001). Understanding the collinearity problem in regression and discriminant analysis. *Journal of Chemometrics*, 15(4), 413–426. <https://doi.org/10.1002/cem.676>
- Nogoy, K. M. C., Sun, B., Shin, S., Lee, Y., Li, X. Z., Choi, S. H., & Park, S. (2022). Fatty acid composition of grain- and grass-fed beef and their nutritional value and health implication. *Food Science of Animal Resources*, 42(1), 18–33. Korean Society for Food Science of Animal Resources <https://doi.org/10.5851/ksosfa.2021.e73>
- Opatye, G. O., Johnson, C. A. G., Oyelahan, T. T., & Awe, O. O. (2025). Overweight prediction using weighted hard and soft voting ensemble machine learning classifiers. In O. O. Awe, & E. A. Vance (Eds.), *Practical statistical learning and data science methods*. Springer Nature Switzerland.
- Patel, N., Toledo-Alvarado, H., & Bittante, G. (2021). Performance of different portable and hand-held near-infrared spectrometers for predicting beef composition and quality characteristics in the abattoir without meat sampling. *Meat Science*, 178. <https://doi.org/10.1016/j.meatsci.2021.108518>
- Prache, S., Adamiec, C., Astruc, T., Baéza-Campane, E., Bouillot, P. E., Clinquant, A., ... Santé-Lhoutellier, V. (2022). Review: Quality of animal-source foods. *Animal*, 16. <https://doi.org/10.1016/j.animal.2021.100376>. Elsevier B.V.
- Prache, S., Martin, B., & Coppa, M. (2020). Review: Authentication of grass-fed meat and dairy products from cattle and sheep. *Animal*, 14(4), 854–863. Cambridge University Press <https://doi.org/10.1017/S1751731119002568>
- Prieto, N., López-Campos, O., Zijlstra, R. T., Uttaro, B., & Aalhus, J. L. (2014). Discrimination of beef dark cutters using visible and near infrared reflectance spectroscopy. *Canadian Journal of Animal Science*, 94(3), 445–454. <https://doi.org/10.4141/CJAS-2014-024>
- Prieto, N., Pawluczyk, O., Dugan, M. E. R., & Aalhus, J. L. (2017). A review of the principles and applications of near-infrared spectroscopy to characterize meat, fat, and meat products. *Applied Spectroscopy*, 71(7), 1403–1426. SAGE Publications Inc <https://doi.org/10.1177/0003702817709299>
- Roger, J. M., Palagos, B., Bertrand, D., & Fernandez-Ahumada, E. (2011). CovSel: Variable selection for highly multivariate and multi-response calibration. Application to IR spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 106(2), 216–223. <https://doi.org/10.1016/j.chemolab.2010.10.003>
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8).
- Scollan, N. D., Dannenberger, D., Nuernberg, K., Richardson, I., MacKintosh, S., Hocquette, J. F., & Moloney, A. P. (2014). Enhancing the nutritional and health value of beef lipids and their relationship with meat quality. *Meat Science*, 97(3), 384–394. Elsevier BV <https://doi.org/10.1016/j.meatsci.2014.02.015>
- Sentandreu, M.A., & Sentandreu, E. (2014). Authenticity of meat products: Tools against fraud. *Food Research International*, 60, 19–29. Elsevier Ltd <https://doi.org/10.1016/j.foodres.2014.03.030>
- Sierra, V., Aldai, N., Castro, P., Osoro, K., Coto-Montes, A., & Oliván, M. (2008). Prediction of the fatty acid composition of beef by near infrared transmittance spectroscopy. *Meat Science*, 78(3), 248–255. <https://doi.org/10.1016/j.meatsci.2007.06.006>
- van der Sluis, T., Pedrolí, B., Kristensen, S. B. P., Lavinia Cosor, G., & Pavlis, E. (2016). Changing land use intensity in Europe – Recent processes in selected case studies. *Land Use Policy*, 57, 777–785. <https://doi.org/10.1016/j.landusepol.2014.12.005>
- Varre, J. V., Statham, T., Smith, A. F., Ahsin, M., Cloward, J., Herrera, M. C., ... Van Vliet, S. (2025). *Nutritional composition of beef: A comparison of commercial North American grass-and grain-finishing systems*. <https://doi.org/10.1093/jas/skaf436/8383911>