Food Control 166 (2024) 110753

Contents lists available at ScienceDirect

Food Control

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

Machine learning-enabled fatty acid quantification and classification of pork from autochthonous breeds using low-field ^{1}H NMR spectroscopic data

José Luis Ramiro [a,b], Ana G. Neo [b], Trinidad Pérez-Palacios [a], Teresa Antequera [a], Carlos F. Marcos [b,*]

[a] Institute of Meat and Meat Products, Universidad de Extremadura, 10003, Cáceres, Spain
[b] Laboratory of Bioorganic Chemistry & Membrane Biophysics (L.O.B.O.), Universidad de Extremadura, 10003, Cáceres, Spain

ARTICLE INFO

Keywords:
Lipidomics
Fatty acids
Regression models
Nuclear magnetic resonance spectroscopy
Iberian pig
Celta pig

ABSTRACT

Traditional pig breeds, known for their sustainability and superior meat quality, are experiencing growing consumer preference. The lipid fraction composition of these meats plays a fundamental role in their health benefits and excellent organoleptic properties. Accordingly, accurate characterisation of intramuscular fat is crucial for maintaining quality standards and combating fraudulent practices. This study employs benchtop nuclear magnetic resonance (NMR) spectroscopy to delineate the lipidic profiles of various cuts from two emblematic Spanish autochthonous pig breeds. The implementation of chemometric and machine learning models enabled the classification of pork samples based on cut and breed of origin. Moreover, this investigation pioneers the coupling of benchtop NMR with machine learning models for quantitative purposes, achieving precise quantification of polyunsaturated, monounsaturated and saturated fatty acids in intramuscular fat. This novel approach holds promise for enhancing the traceability and authentication of traditional pig products, fostering consumer confidence and promoting sustainable livestock practices.

1. Introduction

Over the past few years, there has been an increasing demand for meat products that are both healthy and environmentally friendly. Consumers are seeking high-quality products that align with the values of animal welfare and sustainability (Vitale et al., 2020). Locally produced autochthonous breeds often fulfil these demands by relying on ancient techniques passed down through generations (Thompson et al., 2023). These breeds, raised in extensive production systems, contribute to local ecosystem maintenance, reduced emissions, and cultural heritage preservation (García-Gudiño, N. T. R. Monteiro, Espagnol, Blanco-Penedo, & Garcia-Launay, 2020; Plieninger et al., 2021). Additionally, their unique genetic makeup contributes to exceptional product quality (Pugliese & Sirtori, 2012). This combination fosters high-value products with a positive local economic impact.

Within the Iberian Peninsula, indigenous pig breeds (*Sus scrofa domestica*) can be classified into two distinct genetic lineages: Mediterranean and Celtic (Gama et al., 2013). The Iberian pig exemplifies the Mediterranean lineage, traditionally raised in the *dehesa* ecosystem of the southwest of the Iberian Peninsula (Ortiz et al., 2021) and known for its high-quality products, especially dry-cured specialties, which have achieved wider commercial recognition. In contrast, the Celta pig, a highly resilient breed perfectly adapted to the forests of the northwest Iberian Peninsula, represents the Celtic lineage (Temperan, Lorenzo, Castiñeiras, Franco, & Carballo, 2014). While both breeds are valued for their unique sensory and flavour characteristics, the Celta pig's commercial presence may be less extensive, although it is steadily growing. A characteristic that distinguishes these traditional breeds is their high intramuscular fat content, which enhances meat juiciness, flavour, and tenderness. (Scollan, Price, Morgan, Huws, & Shingfield, 2017). We have recently reported that intramuscular fat (IMF) in Iberian and Celta pigs ranges between 19 and 26% of the dry weight, with no statistical differences between breeds (Ramiro et al., 2024). These values are consistent with those previously reported for other southern European pig breeds (Pugliese et al., 2012).

A comparative characterisation of these two prized Iberian Peninsula breeds, the Iberian pig and Celta pig, focusing on the meat quality, could elucidate the similarities and differences arising from environmental

* Corresponding author.
E-mail address: cfernan@unex.es (C.F. Marcos).

https://doi.org/10.1016/j.foodcont.2024.110753
Received 22 May 2024; Received in revised form 10 July 2024; Accepted 19 July 2024
Available online 20 July 2024

conditions, genetic determinants and production practices. This knowledge would be particularly valuable for understanding the role of these factors in shaping the lipid profile of intramuscular fat (IMF), a critical determinant of meat flavour, texture and oxidative stability, all of which significantly impact consumer acceptance (Wood et al., 2008).

Here, Nuclear Magnetic Resonance (NMR) spectroscopy (Cao et al., 2021; Pajuelo et al., 2022) emerges as a powerful alternative to the traditional Gas Chromatography with Flame Ionisation Detection (GC-FID) method (Ramiro et al., 2024). Unlike GC-FID, which requires tedious derivatisation (Sandler & Karo, 1992, pp. 83–91), NMR provides a comprehensive characterisation of the lipid fraction in a single experiment, offering valuable qualitative and quantitative data. Furthermore, NMR requires minimal sample preparation and eliminates the need for calibration curves (Zaukuu, Benes, Bázár, Kovács, & Fodor, 2022). The combination of NMR analysis with chemometrics (Wang et al., 2022) or machine learning techniques (Corsaro et al., 2022) offers an exceptional approach to face modern challenges in food science, particularly for complex food matrices like meat.

Recent advancements in affordable and user-friendly benchtop NMR spectrometers open new avenues for food analysis (Giberson, Scicluna, Legge, & Longstaffe, 2021). Combined with chemometrics, benchtop NMR has been successfully applied to differentiate food products and assess quality (Galvan, de Aguiar, Bona, Marini, & Killner, 2023; Galvan et al., 2021; Migues, Rivas, Moyna, Kelly, & Heinzen, 2022; Soyler et al., 2021). However, its application in meat quality analysis remains largely unexplored beyond initial studies (Jakes et al., 2015; Pajuelo et al., 2023).

Multivariate statistical analysis, such as Principal Component Analysis (PCA) and Partial Least Square Discriminant Analysis (PLS-DA), offer powerful tools for extracting meaningful information from complex datasets like NMR data (Wang et al., 2022). On the other hand, diverse machine learning algorithms have emerged to resolve classification and regression problems on food samples. These techniques have been widely applied over the last few years for discrimination, classification, fraud detection and quality control of foods using data obtained by different analytical methods (Jiménez-Carvelo, González-Casado, Bagur-González, & Cuadros-Rodríguez, 2019). While less common, successful applications of NMR spectral data with machine learning have been reported for predicting the geographical origin of various products, such as hazelnuts (Bachmann, Klockmann, Haerdter, Fischer, & Hackl, 2018), black tea (Cui et al., 2023), asparagus (Klare et al., 2020) and rice (Saeed et al., 2022). Surprisingly, to our knowledge, machine learning regression algorithms have never been used on NMR data for quantification, despite NMR being an inherent quantitative technique.

This study aims to develop chemometrics and machine learning models using ^{1}H NMR spectral data of pork lipid fractions. Specifically, these models will be tailored to achieve two key objectives: 1) classifying commercially important raw pork cuts (loin -*Longissimus thoracis*-, presa -*Serratus ventralis*-, cheek, and liver) from Iberian and Celta pigs, and 2) accurately predicting the fatty acid profiles of these same pork cuts based on the same ^{1}H NMR data.

2. Materials and methods

2.1. Reagents and standards

All chemicals were of analytical grade and were used as received without any further purification. Chloroform, methanol and sodium sulphate from Scharlau were used for the lipid extraction. *n*-Hexane, methanol, sodium methoxide, sulphuric acid from Scharlau, sodium chloride from Panreac, Supelco 37-component FAME mix and tridecanoic acid from Sigma were used in the transesterification procedure. Deuterochloroform ($CDCl_3$) containing 1% tetramethylsilane (TMS) from Sigma was used in the NMR analysis.

2.2. Experimental design

This study used meat from five Iberian and five Celta pigs (*Sus scrofa domestica*) sourced from two separate meat industries. The animals, between 12 and 16 months old and weighing approximately 140–160 kg, were humanely slaughtered at an industrial abattoir in Guijuelo (Salamanca, Spain) under approved procedures. Portions (50 g each) taken from the central part of fresh liver, cheek, *presa* (*Serratus ventralis*) and loin (*Longissimus thoracis*) were individually minced, yielding a total of 40 samples. To guarantee triplicate analysis, three 5 g subsamples were taken from each sample. The resulting 120 samples were vacuum-sealed and frozen at −80 °C until analysis.

2.3. Lipid extraction

Intramuscular fat (IMF) was extracted following the method of Folch, Lees, and Stanley (1957) modified by Perez-Palacios, Ruiz, Martin, Muriel, and Antequera (2008).

2.4. Determination of fatty acids by gas chromatography

The lipid extracts were transesterified with methanol and the resulting fatty acid methyl esters (FAME) were analysed in a Hewlett–Packard HP-5890A gas chromatograph coupled to flame ionisation detector (GC-FID) as previously described (Ramiro et al., 2024). Quantification of fatty acids (FA) was performed using the external calibration curve method and the results, expressed in % mol, were calculated using the exact weight of the sample and the molecular weight of each FAME.

2.5. ^{1}H NMR spectroscopy

A portion of extracted lipids from each sample (25 mg) was dissolved in 500 μL of $CDCl_3$ containing 1% TMS as internal reference and placed in standard 5 mm NMR tubes. ^{1}H NMR spectra were recorded in a Magritek Spinsolve 80 MHz Carbon Ultra spectrometer, using a spectral width of 5000 Hz. Acquisition was carried out at 298 K, collecting 32768 data points after 256 scans, with a relaxion delay of 38.79 s and pulse width of 11.6.

The spectra were processed using MestreNova software (Mestrelab Research, 2019). A gaussian apodisation of 0.4 MHz was applied to free induction decay data prior to Fourier transformation. Automatic phase correction employing Metabonomics and Regions Analysis algorithms in conjunction with Whittaker Smoother baseline correction gave the best results and were routinely applied to all spectra. The spectra were then aligned using TMS signal as a reference, cut between 5.5 ppm and 0.6 ppm and normalised to the sum of all the area. To ensure a homogeneous treatment of the spectra, a process template was used to carry out the spectra processing. ^{1}H NMR signals were assigned according to previous studies (Pajuelo et al., 2022, 2023) and integrated individually.

2.6. Chemometrics

The processed ^{1}H NMR spectra were divided into bins of 0.1 ppm width, which were integrated yielding 49 variables per sample for the entire dataset (n = 120). This data was then subjected to sparse Partial Least Squares Discriminant Analysis (sPLS-DA) using the mixOmics package in R v.4.3.1 (R Core Team, 2023). sPLS-DA models were constructed for each of the four cut type subsets to classify the samples based on their breed of origin.

2.7. Machine learning modelling

2.7.1. Classification

The entire binned spectral dataset was also analysed using Machine Learning (ML) classification models (Sarker, 2021) implemented in the

caret, random forest and gbml packages in R v.4.3.1 (R Core Team, 2023). A comparative evaluation of various machine learning models was conducted to identify the one that yielded the most accurate classification performance. The models examined included Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Random Forest (RF), Linear Discriminant Analysis (LDA), Gradient Boosting Regression Trees (GBRT) and Naïve-Bayes (NB).

SVM models employ hyperplanes to delineate class boundaries, achieving optimal separation by identifying the hyperplane that maximises the margin between classes and minimises the distance to training data points. KNN, one of the simplest machine learning algorithms, classifies observations by identifying observations in the training set that are similar to the test observation (nearest neighbours) and assigning the predominant class among those neighbours as the predicted class. Random Forest models consist of a set of individual predictive models formed by binary rules (decision trees), each trained with a slightly different sample of the training data generated through bootstrapping. LDA is designed to identify a linear combination of features that effectively discriminates between distinct categories. GBRT models, also an ensemble learning approach, construct a series of decision trees, each iteratively improving upon the errors of the previous trees. NB models employ Bayes' theorem to estimate the probability that an observation belongs to each class and assign it to the class with the highest predicted probability.

2.7.2. Quantification

For fatty acid quantification purposes, eight peaks **A**, **B**, **D**, **E**, **F**, **P**, **G** and **I**, identified in the ^{1}H NMR spectra based on chemical shifts and corresponding to specific lipid protons, were integrated using MestreNova software (Mestrelab Research, 2019). The integration intervals for these peaks are summarised in Table S1 (Supplementary Material). The resulting integrals for each of the 120 samples served as independent variables in a dataset used to construct ML regression models (Sarker, 2021) implemented using R software (R Core Team, 2023). These models were then used to predict the concentrations of PUFA, MUFA and SFA obtained from the separate GC analysis.

The ML regression models used for quantification included Linear Support Vector Machines (SVML), Support Vector Machine with kernel radial basis function (SMV-RBF), K-Nearest Neighbours (KNN), Random Forest (RF), Generalised Linear Model (GLM), which is a multiple regression algorithm, and Gradient Boosting Regression Trees (GBRT).

2.8. Evaluation of the chemometric and machine learning models

To evaluate the performance of chemometric and ML models, 10-fold cross-validation was carried out, and relevant statistical metrics were collected for each case. For the ML models, the complete spectral integrals dataset was randomly split into a training set, which represents the 80% of samples, and a test group consisting of the remaining samples (20%), ensuring a representative distribution of pig breeds and cuts across both subsets. The training set was used to construct and validate the models, while the test set was employed to assess their generalizability on unseen data.

Chemometric classification models were evaluated by four key metrics: accuracy, kappa, sensitivity, and specificity. ML classification models were evaluated by accuracy and kappa. Accuracy is a measure of how close a classification model's predictions are to the actual values. It is calculated as the ratio of the number of correct classifications to the total number of samples predicted (equation (1)). Sensitivity is a statistical measure that describes the proportion of samples belonging to a particular class A, which is defined as positive, that are correctly classified as positive (equation (2)). Specificity is a statistical measure that describes the proportion of samples belonging to the negative class B that are correctly classified as negative (equation (3)). In this study, the positive class A was arbitrarily assigned to Iberian pork samples, while the negative class B was assigned to Celta pork samples. Kappa (equation (4)), a more stringent measure than accuracy, evaluates the proportion of agreements beyond what would be expected by chance alone (Cuadros-Rodríguez, Pérez-Castaño, & Ruiz-Samblás, 2016).

$$\text{Accuracy} = (TP + TN) / (P + N) \quad (1)$$

$$\text{Sensitivity} = TP / P \quad (2)$$

$$\text{Specifity} = TN / N \quad (3)$$

$$\text{Kappa} = (P_a - P_{ch}) / (1\text{-}P_{ch}) \quad (4)$$

P = the number of real positive cases in the data, N = the number of real negative cases in the data, TP = true positives, TN = true negatives, P_a = the agreement probability and P_{ch} = the chance-agreement probability.

A confusion matrix was generated using the cvms package in R v.4.3.1 to assess the discriminative ability of the optimal machine learning model (R Core Team, 2023).

To assess the performance of quantitative methods, we computed Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2) in both the training and test datasets. These metrics provide a measure of the difference between predicted and actual values. Mean Absolute Error (MAE) is the average of the absolute differences between the predicted and actual values (equation (5)). Root Mean Square Error (RMSE) is the square root of the average of the squared differences between the predicted and actual values (equation (6)). The coefficient of determination (R^2) is a measure of how well the data fit the regression model (equation (7)).

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} \left(y_i - \frac{1}{n} \sum_{i=1}^{n} y_i \right)^2} \quad (7)$$

$\hat{y}_i$ = predicted value for sample i, y_i = true value for sample i, n = sample size.

FA quantities obtained by GC and ^{1}H NMR derived ML algorithms, were statistically compared employing the parametric *t*-test for independent samples or the non-parametric U-Mann Whitney test, basing of normality distribution criteria of samples estimated with Shapiro-Wilk test, using a significance level of 95% (*p-value* 0.05).

3. Results and discussion

3.1. ^{1}H NMR spectra assignment

The whole set of samples was analysed by GC-FID and benchtop ^{1}H NMR spectroscopy. Fig. 1 shows representative 80 MHz ^{1}H NMR spectra corresponding to the lipid extract of each Iberian and Celta pig cut. The characteristic bands of different protons of triacyl glycerides and phospholipids are easily identified and have been tagged following an alphanumerical nomenclature. The signals were assigned according to reported assignations in benchtop and high field ^{1}H NMR spectra (Supplementary Material, Table S1) (Pajuelo et al., 2022, 2023).

Peaks **A** though F and **I** are due to the different protons of FA chains. The triplet **A** between 0.72 ppm and 1.03 ppm corresponds to the terminal methyl group of the FA chains. Due to the limited resolution of low-field NMR spectra, it is not possible to distinguish between the signal associated with ω3 FA and that of other FA, as they both merge

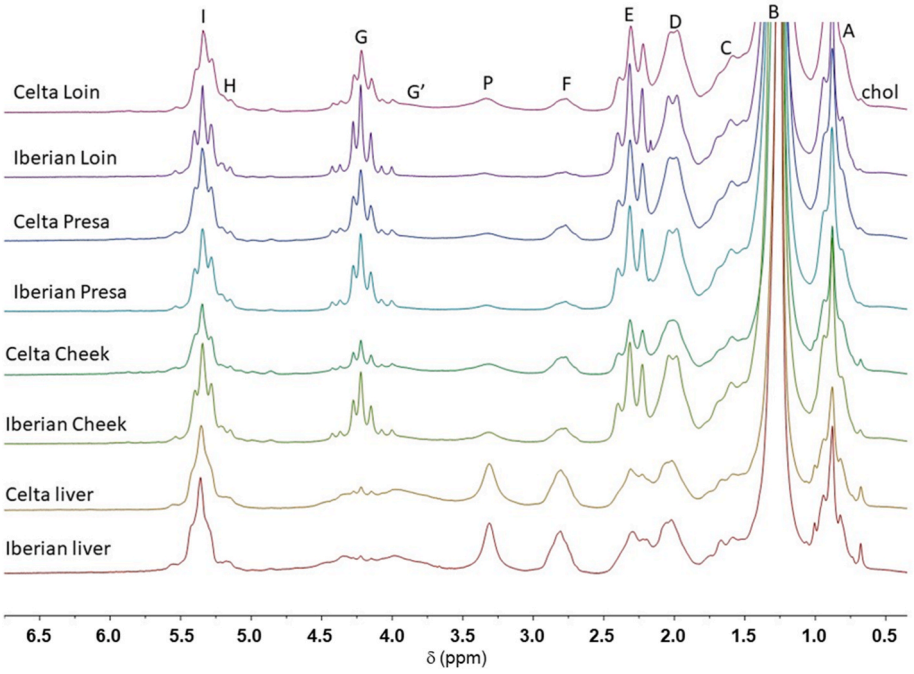

Fig. 1. Benchtop ^{1}H NMR spectra assignment of lipid extract from different tissues derived from Iberian and Celta pigs. **Chol.** CH_3, cholesterol C-18. **A** FA terminal CH_3, **B** FA $(CH_2)_n$, **C** β-CH_2, **D** MUFA and PUFA allyl CH_2, **E** FA α-CH_2, **F** PUFA bis-allyl CH_2, **P** phosphatidylcholine $N(CH_3)_3$ and phosphoethanolamine $C\underline{H}_2NH_3$, **G** TAG glyceril CH_2, **G′** PL glyceril CH_2, **H** TAG glyceril CH and PL glyceril CH, **I** MUFA and PUFA vinyl CH.

into peak A. The large peak **B** corresponds to all the CH_2 protons of FA that are not next to a double bond or in α or β to the carboxylic group. Peaks **C** and **E** correspond to FA CH_2 groups respectively in β and α to the carboxylic group. Band **D** results from allylic protons, and signal **F** corresponds to bis-allylic protons of polyunsaturated fatty acids (PUFA) and is significantly larger in tissues that contain a larger proportion of phospholipids (PL), such as cheek and especially liver. Finally, peak **I** is due to the vinylic protons of unsaturated FA (UFA). Furthermore, peaks **G** and **G'** result from the CH_2 protons of the glyceryl skeleton of triacylglycerides (TAG) and PL, respectively, while peak **H**, which corresponds to the CH signal for glycerol, shows quite different chemical shifts when belongs to TAG or PL. On the other hand, peak **P** comprises

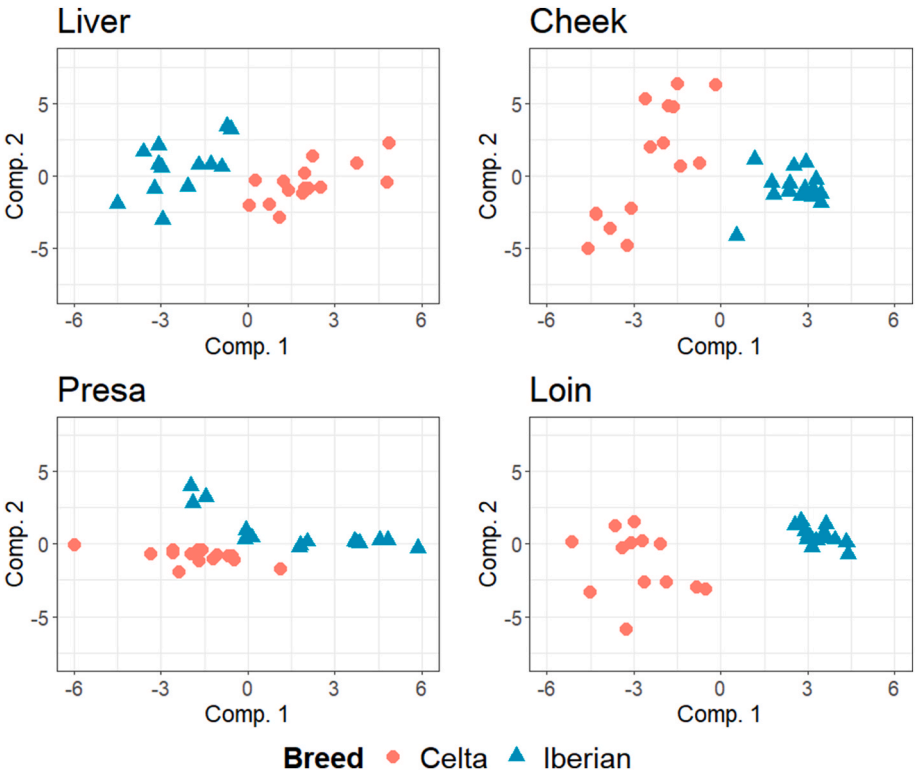

Fig. 2. sPLS-DA two-dimensional scores plots for each pork cut. Samples cluster according to breed.

the signals of phosphatidylcholine $N(CH_3)_3$ and phosphatidylethanolamine CH_2 next to the amine group.

3.2. Benchtop ^1H NMR, chemometrics and machine learning analysis as classification tools for Iberian and Celta pig cuts

Although the NMR spectra of the different cuts reveal clear differences in composition, discerning variations between spectra of the same cut from distinct breeds can be challenging at first inspection. This observation aligns with previous findings on FA composition disparities within the Iberian pig cheek, loin, and presa cuts reported by Morcuende, Estevez, Ruiz, and Cava (2003) and variations between loin and liver observed by Benitez et al. (2015). Similarly, R. Domínguez, Martínez, Carballo, and Franco (2014) described compositional heterogeneity across several Celta pig muscles. We reasoned that chemometric and machine learning techniques could help to determine the breed of origin of the different cuts.

To achieve this goal, the four subsets of ^1H NMR bin data corresponding to each pork cut were analysed using sparse Partial Least Square Discriminant Analysis (sPLS-DA). Fig. 2 shows the two-dimensional score plots representing the two first components for each cut. A clear breed-based clustering pattern is observed in all meat cuts, although this distinction is less evident in the case of *presa* samples. Accuracy, Kappa, Sensitivity, and Specificity were computed to assess the performance of the models. Most of these metrics showed values between 0.9 and 1.0, except in *presa* cuts, which presented slightly lower values, although always over 0.8 (See Supplementary Material Table S2 for more details).

Multivariate analysis has been used before to breed discrimination of pork using lipidomic data obtained by GC and HPLC-MS analysis (Li et al., 2021).

The Variable Importance in Projection (VIP) score provides a measure of the importance of single variables in the sPLS-DA model (See Supplementary Material Fig. S1) (Chong & JunC, 2005). Interestingly, the variables with higher VIP scores indicate the main ^1H NMR spectral regions that allow for discrimination of breeds in the different pork cuts.

Key differentiation variables for breed classification in presa and loin samples are bis-allylic protons, derived from both ω-3 and ω-6 PUFAs. This difference likely reflects variations in the ω-6 to ω-3 FA ratio, which has been previously shown to distinguish loin meat from Pulawska and Polish Landrace pigs (Kasprzyk, Tyra, & Babicz, 2015). Additionally, phospholipid glyceride protons and FA α-CH_2 protons are important variables, potentially indicating a distinction in the TAG to phospholipid ratio between Iberian and Celta pigs. This characteristic might be related to IMF deposition (Gkarane et al., 2018). However, in a previous study, we did not find significant differences in fat content between Iberian and Celta pig muscles (Ramiro et al., 2024). Therefore, we cannot definitively conclude that this is a reliable marker for breed differentiation based solely on NMR data. Still, the prominence of phosphatidylcholine and phosphatidylethanolamine signals in the VIP of loin samples further strengthens the hypothesis of a TAG/phospholipid ratio difference. In cheek samples, the key differentiation markers are peaks corresponding to $(CH_2)_n$ and terminal CH_3, common to all FA. Notably, the cholesterol peak appears as a crucial discriminator between Celta and Iberian cheek, further supporting the influence of breed-related genetic factors on cholesterol content in pork, as demonstrated by Liu et al. (2009). In contrast to other cuts, the primary differentiation markers in liver samples appear to be spectral regions associated with mono- and diacylglycerides and caproleic acid. The presence of mono and diacylglycerides in liver is not surprising, as these molecules can result from TAG lipolysis or serve as direct TAG biosynthetic precursors in this tissue (Yi, Huang, Wang, & Shan, 2023). On the other hand, caproleic acid has not been often detected in pork products. However, it has been reported as a marker to differentiate fresh from cured Toscano PDO hams (Sirtori et al., 2020). These signals are readily identifiable in the 500 MHz ^1H NMR spectra of liver samples (see Supplementary Materials, Fig. S2) (Boccia, Cusano, Scano, & Consonni, 2020), although they are not evident in low-field spectra. Despite their invisibility in low-field spectra, these signals still contribute to the overall spectral shape. This demonstrates the ability of multivariate analysis to uncover spectral information that is undetectable by the naked eye.

In order to simultaneously classify the samples from the whole dataset according both to breed and cut, we have then trained and tested various ML models. These include Supported Vector Machines (SVM), K-Nearest Neighbour (KNN), Naïve-Bayes (NB), Linear Discriminant Analysis (LDA), Random Forest (RF) and Gradient Boosting Regression Trees (GBRT). Hyperparameters of each model were tuned for optimal performance, which was evaluated by the prediction accuracy on the test set (Table 1).

Evaluating the performance metrics revealed a progressive improvement in performance. NB often struggles with complex datasets like ours (Wickramasinghe & Kalutarage, 2020), although exhibited the lowest overall accuracy and other metrics, suggesting it may not be well-suited for this classification task. KNN and RF models yielded slight improvements, potentially due to their ability to capture some non-linear relationships in the data (Vigneau, Courcoux, Symoneaux, Guérin, & Villière, 2018). However, LDA and GBRT demonstrated a significant leap in performance compared to NB, KNN, and RF. This could be attributed to LDA's strength in handling well-separated classes and GBRT's ability to learn complex decision boundaries (Sarker, 2021). Finally, the SVM model surpassed all other models in terms of training accuracy, kappa statistic, and even achieved a 0% error rate on the test set. The corresponding confusion matrix (Supplementary Information, Fig. S3) offers further insights into the model's generalizability by illustrating its performance on unseen data. Examining the confusion matrix shows that all predictions for each cut matched the true class of the samples, with an accuracy of 100% for each class. Overall, the exceptional performance on both training and test data, coupled with SVM's ability to handle non-linear relationships potentially present in our data, suggests its suitability for accurate classification of meat cut and breed in new samples. This success aligns with previous studies demonstrating the effectiveness of SVM for classification tasks using ^1H NMR spectroscopy. Notably, SVM was successfully applied to classify wines (Nyitrainé Sárdy, Ladányi, Varga, Szövényi, & Matolcsi, 2022), asparagus (Klare et al., 2020), rice (Saeed et al., 2022) and eggs (Bischof, Januschewski, & Juadjur, 2024) according to various characteristics. However, to our knowledge, this is the first application of machine learning models based on ^1H NMR spectroscopic data to classify meat products.

3.3. Machine learning models for the fatty acid quantification in Iberian and Celta pig meat

The quantitative nature of NMR techniques relies on the values of the peak integrals, which are proportional to the number of protons contributing to each peak. This allows for calculating the proportion of different types of FA from the integrals of ^1H NMR peaks that correspond to different functional groups. In previous studies, we have used this approach successfully to quantify the proportion of SFA, MUFA and PUFA from both high field (Pajuelo et al., 2022) and low field (Pajuelo

Table 1
Performance metrics of machine learning algorithms for classification of traditional pig tissues using benchtop ^1H NMR spectral data.

ML model	Accuracy (training)	Kappa (training)	Accuracy (test)	Kappa (test)	Prediction error rate
NB	0.71	0.66	0.71	0.67	29.17%
KNN	0.78	0.74	0.83	0.81	16.67%
RF	0.81	0.78	0.88	0.86	12.5%
GBRT	0.83	0.81	0.92	0.91	8.33%
LDA	0.90	0.88	0.92	0.91	8.33%
SVM	**0.90**	**0.88**	**1.00**	**1.00**	**0%**

et al., 2023) NMR spectra of Iberian cured hams. Other groups have also used NMR spectra in quantitative lipid analysis of different foods (Castejón, Mateos-Aparicio, Molero, Cambero, & Herrera, 2014; Hajjar, Haddad, Rizk, Akoka, & Bejjani, 2021; Mannina et al., 2008; Nieva-Echevarría, Goicoechea, Manzanos, & Guillén, 2014).

For instance, PUFA content in pork meat could be directly estimated from the integral of ^{1}H NMR peak **F**, which corresponds to bis-allylic protons. The major PUFA in pork meat samples are arachidonic and linoleic acids, each possessing six and two bis-allylic protons, respectively; however, their relative proportions differ substantially across various pig tissues. Additionally, other PUFA with varying proton counts also contribute to peak **F** and their exact ratios remain unknown. Therefore, solely relying on peak **F** limits the determination of precise PUFA content.

Other NMR signals important for quantification are peaks **D** and **E**. Peak **D** corresponds to the allylic protons of MUFA and PUFA, while peak **E** accounts for the total content of FA. Unfortunately, limitations in resolution and peak overlap make it difficult to estimate FA concentration accurately, particularly in phospholipid-rich samples like liver or cheek. To overcome these challenges, more advanced methods are necessary.

3.3.1. *Machine learning regression algorithms to quantify FA from benchtop NMR*

ML techniques can extract hidden information from large amounts of data by using different algorithms (Sarker, 2021). Predictive models for FA quantification could be created using ML techniques based on the ^{1}H NMR spectra of lipid extracts from meat samples. For this, the integration intervals corresponding to the different NMR peaks (**A**, **B**, **D**, **E**, **F**, **P**, **G**, **I**) were carefully delimited to reflect as closely as possible the proportion of each type of protons in the samples. The resulting integrals were used as independent variables to feed ML models with the aim of correlating them with the concentrations of PUFA, MUFA and SFA determined by GC-FID. Independent models were trained to estimate the percentages of PUFA, MUFA and SFA, respectively. Interestingly, while this paper was under revision, Hernández-Jiménez, Revilla, Hernández-Ramos, and Vivar-Quintana (2024) reported a conceptually similar approach for the prediction of the FA profiles of Iberian pig products applying Artificial Neural Networks with Near Infrared Spectroscopy (NIR). The exploration of ML combined with diverse spectroscopic techniques underscores the potential for rapid and reliable quality assessment methods in this important food sector.

Different ML algorithms were explored, including K-Nearest Neighbours (KNN), Random Forest (RF), Gradient Boosting (GBRT), Generalised Linear Model (GLM), Supported Vector Machines (SVML) and SVM with radial basis function (SVM-RBF). Hyperparameters were tuned for each model using a systematic trial-and-error procedure to optimise their performance metrics (Table 2). Finally, the best model was selected for FA quantification.

The K-Nearest Neighbours (KNN) models developed in this study exhibited moderate ability to approximate the ratio of PUFA, MUFA, and SFA in pork meat samples. However, the models' performance was hindered by high root-mean-square error (RMSE) and mean absolute error (MAE) values, particularly for the test dataset (Table 2). These elevated error metrics indicate that the KNN models often made inaccurate predictions, leading to substantial discrepancies between predicted and actual fatty acid ratios. The low R^2 values for both training and test datasets further underscore the models' shortcomings.

Methods based on trees, such as RF and GBRT (Table 2), performed much better, but R^2 values for SFA were still relatively low. GLM showed better results than previous algorithms, especially for SFA, where it shows an R^2 value of 0.82 in the test database, with also fairly good MAE and RMSE both in the training and test subsets. (Table 2).

SVM models show high capability to treat high dimensionality data, often encountered in spectroscopic analyses like NMR. This characteristic makes them particularly advantageous for fatty acid quantification in pork meat samples. Additionally, SVM's inherent ability to handle non-linear relationships is crucial (Brereton & Lloyd, 2010). Fatty acid ratios might not have a simple linear relationship with the complex spectral data from NMR. Here, the use of the Radial Basis Function (RBF) kernel in SVM-RBF plays a key role. This kernel function allows SVM to capture these non-linear relationships, leading to more accurate predictions (Scholkopf et al., 1997). An interesting property of SVM is its low tendency to overfitting, even when a large number of samples are not available for training (Deris, Zain, & Sallehuddin, 2011). The developed SVM models present a predictive capacity and experimental correlation quite similar to that found with GLM for all the classes (Table 2). SVM applying kernel radial basis function (SVM-RBF) usually performs better than other kernel functions and was also tested (Table 2). SVM with RBF kernel demonstrated strong predictability for FA quantification, achieving low MAE and RMSE in training and test datasets for PUFA and MUFA, with high values of R^2 (0.95–0.97). Besides, it shows low error metrics and much better R^2 values (0.84–0.89) than other models for SFA datasets. These results highlight the effectiveness of SVM-RBF in capturing the complex relationship between NMR data and fatty acid ratios, leading to superior quantification accuracy for all fatty acid classes. Support vector machine regression has been previously successfully used for quantification of different food parameters, such as some edible oil FA from Terahertz spectroscopic data (Jiusheng, 2010), physicochemical characteristics of Bísaro pig loin using NIR (Vasconcelos et al., 2023) and sweetness of sugars and artificial sweeteners from molecular descriptors (Zhong, Chong, Nie, Yan, & Yuan, 2013).

As shown in Fig. 3A there is a high correlation between the values of PUFA, MUFA and SFA predicted by the SVM with RBF kernel for the test data set and the values measured by GC-FID, with R^2 values of 0.97, 0.95 and 0.84, respectively. Furthermore, statistic *t*-test U-Mann Whitney test showed there are no significant differences between the predicted and GC-measured values for each type of FA as evidenced by Fig. 3B. The calculated *p*-values for PUFA, MUFA and SFA are 0.983, 0.893 and 0.698, respectively. Fig. 3A and B confirm the SVM model's ability to accurately predict FA content, particularly for PUFA and MUFA. While the correlation for SFA is slightly lower, the lack of significant difference between predicted and measured values suggests acceptable prediction in practice.

Our SVM-RBF model accurately quantified PUFA, MUFA, and SFA in

Table 2
Performance metrics of ML algorithms for FA quantification using benchtop ^{1}H NMR spectral integrals.

	PUFA						MUFA						SFA					
	Training			Test			Training			Test			Training			Test		
ML model	*MAE*	*RMSE*	R^2	*MAE*	*RMSE*	R^2	*MAE*	*RMSE*	R^2	*MAE*	*RMSE*	R^2	*MAE*	*RMSE*	R^2	*MAE*	*RMSE*	R^2
KNN	2.60	3.57	0.84	3.41	5.62	0.67	3.40	4.62	0.82	4.32	6.80	0.71	1.61	2.03	0.59	1.91	2.78	0.38
RF	1.70	2.49	0.91	1.74	2.38	0.95	2.35	3.27	0.90	2.07	2.94	0.94	1.25	1.60	0.74	1.44	2.31	0.52
GBRT	1.67	2.39	0.92	1.38	1.78	0.97	2.27	3.06	0.93	2.49	3.34	0.93	1.15	1.46	0.78	1.87	2.59	0.43
GLM	1.63	2.26	0.94	1.64	2.03	0.96	2.20	2.82	0.94	2.24	2.97	0.95	1.32	1.79	0.71	1.13	1.40	0.82
SVM	1.51	2.157	0.95	1.50	2.18	0.97	2.27	2.88	0.94	2.61	3.27	0.94	1.292	1.70	0.78	1.20	1.45	0.81
SVM-RBF	**1.36**	**1.83**	**0.97**	**1.21**	**2.05**	**0.97**	**1.70**	**2.15**	**0.97**	**1.92**	**2.66**	**0.95**	**1.06**	**0.82**	**0.89**	**1.12**	**1.45**	**0.84**

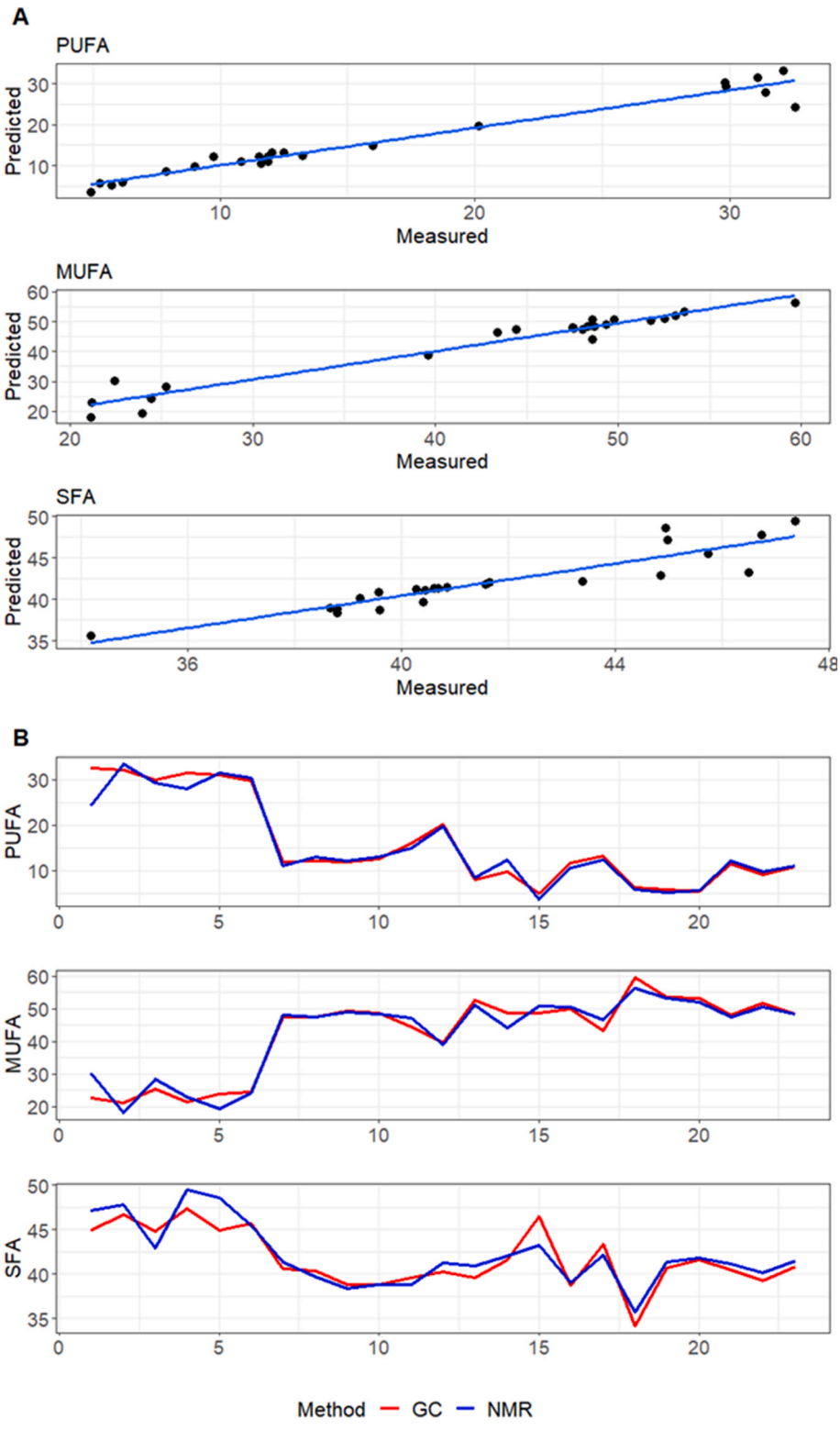

Fig. 3. A) Measured FA against predicted FA percentages based on the SVM-RBF model on the test set. B) Comparison between test set values predicted by SVM-RBF and experimental GC-FID values.

all samples, demonstrating high concordance with GC-FID measurements (Table 3). Complete quantitative FA molar percentages obtained from the SMV-RBF model are provided in Table S3 (Supplementary Material).

Fig. 4 presents the corresponding PUFA, MUFA and SFA profiles for different cuts and breeds. Two-way analyses of variance with a non-parametric aligned rank transformed data test on the SVM-RBF model outputs revealed significant differences (p-values <0.05) in the PUFA, MUFA and SFA content due to breed, cut and their interaction.

Profile graphs (Fig. 4) evidenced pronounced variation across cuts, while breed-specific differences within cuts were less obvious. The most notable disparity was in PUFA content, with multiple comparisons

Table 3
Mean values of PUFA, MUFA and SFA in different Iberian and Celta pork cuts, referred to the whole dataset. Values are expressed in molar percentages. GC and NMR results are shown together with p-values derived from statistical treatment comparing both methods.

	PUFA			MUFA			SFA		
	GC	*SVM-RBF*	*P*	*GC*	*SVM-RBF*	*P*	*GC*	*SVM-RBF*	*P*
Iberian Liver	30.56 ± 1,66	30.17 ± 2.35	0.605	23.23 ± 2.04	23.37 ± 1.68	0.885	46.21 ± 1.13	46.52 ± 1.47	0.516
Celta Liver	29.82 ± 2.61	29.63 ± 2.23	0.821	24.38 ± 4.47	24.46 ± 4.68	0.958	45.79 ± 2.76	45.71 ± 2.56	0.932
Iberian Cheek	12.52 ± 0.70	12.56 ± 0.85	0.897	47.75 ± 1.13	47.75 ± 1.29	0.997	39.51 ± 0.87	39.59 ± 0.94	0.805
Celta Cheek	15.94 ± 4.01	15.79 ± 3.80	0.775	43.55 ± 5.46	44.01 ± 8.31	0.806	40.51 ± 1.63	40.19 ± 5.10	0.624
Iberian Presa	7.36 ± 1.94	7.75 ± 2.33	0.617	49.71 ± 2.19	49.34 ± 2.15	0.650	42.94 ± 2.16	42.76 ± 1.55	0.800
Celta Presa	11.34 ± 1.31	10.79 ± 1.07	0.210	47.39 ± 2.40	48.36 ± 2.44	0.114	41.26 ± 1.30	40.97 ± 1.09	0.527
Iberian Loin	5.41 ± 0.59	5.41 ± 0.65	0.969	54.13 ± 2.99	53.53 ± 2.33	0.539	40.46 ± 3.28	40.50 ± 2.75	0.870
Celta Loin	10.32 ± 1.13	11.10 ± 1.85	0.250	49.49 ± 1.89	48.21 ± 2.57	0.174	40.17 ± 0.91	40.74 ± 0.99	0.111

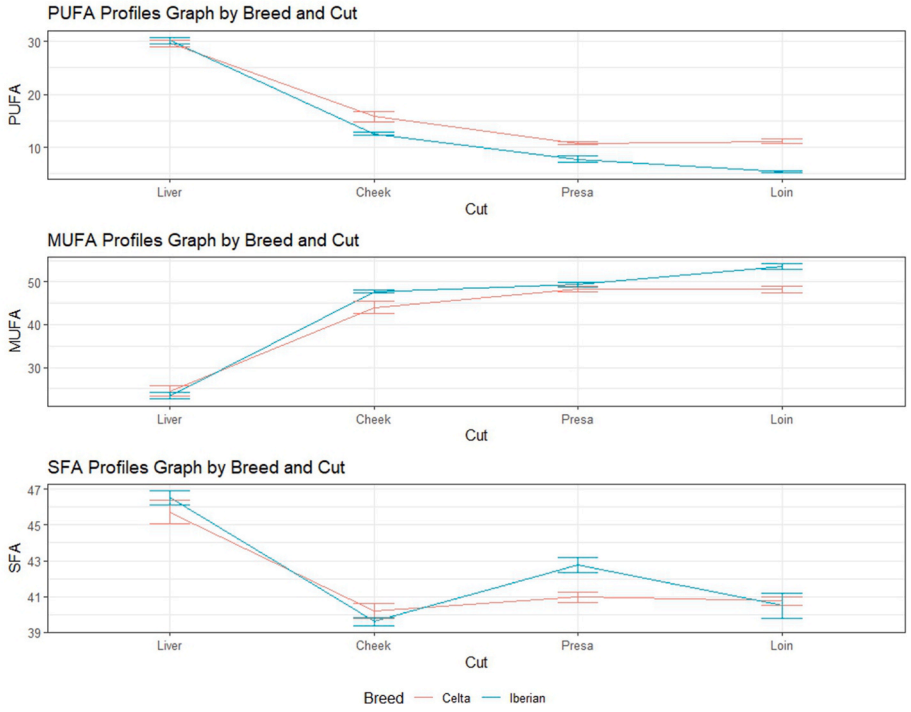

Fig. 4. Profile graphics for the PUFA, MUFA and SFA mean predicted values across cuts of both pig breeds.

revealing significant differences between specific breed-cut combinations, particularly PUFA and MUFA (Table S4, Supplementary Material). This variation can likely be attributed to a combination of genetic factors (Corominas et al., 2013; Valdés-Hernández et al., 2023) and feeding practices (Yi et al., 2023). Iberian pigs are characterised by a high MUFA content, potentially due to both the exceptional desaturation capacity of the breed (Benitez et al., 2015) and the consumption of acorns (González-Domínguez, Sayago, & Fernández-Recamales, 2020; Pajuelo et al., 2023; Pérez-Palacios, Ruiz, Tejeda, & Antequera, 2009). Conversely, a chestnut-based diet is likely responsible for the elevated concentration of both MUFA and PUFA in Celta pigs (Bermudez, Franco, Franco, Carballo, & Lorenzo, 2012; Ruben Domínguez, Martínez, Gómez, Carballo, & Franco, 2015). Consistent with this, Celta presa and loin displayed a higher PUFA content compared to Iberians. However, no significant differences in PUFA, MUFA and SFA were observed between Celta and Iberian cheeks and between Celta and Iberian livers, nor in SFA content between Celta and Iberian loins.

As expected, both breeds exhibited elevated PUFA levels in liver tissue, reaching approximately 30%. MUFA content also displayed notable breed-related variations across cuts, generally favouring Iberians over Celta pigs. MUFA levels ranged from 23% to 24% in livers to 54% in Iberian loin, with all other cuts containing levels of nearly 50%. SFA content remained consistent across breeds for the different cuts. Liver demonstrated the highest SFA content (46%), constituting the primary lipid component in this cut. Notably, despite its high SFA content, liver maintained a favourable PUFA/SFA ratio of 0.65, exceeding the recommended minimum of 0.4 for optimal health (Wood et al., 2008). This ratio dipped to just below 0.4 in Celta cheeks and just over 0.3 in Iberian cheeks, and was considerably lower in presa and loin, despite their reduced SFA content. Interestingly, no significant differences were observed in PUFA, MUFA and SFA content between Celta presa and loin. In contrast, these FA contents displayed significant variations between the same cuts of Iberian pigs. These FA profiles align with previous reports on Iberian and Celta breeds, but differ from commercially reared pork, exhibiting higher MUFA and lower PUFA content (Ruben Domínguez et al., 2015; Ramiro et al., 2024; Rey, Daza, López-Carrasco, & López-Bote, 2006).

4. Conclusion

This study pioneers the application of benchtop NMR, coupled with chemometrics and machine learning, for the complete characterisation of the lipid fraction in raw meat from Iberian and Celta pigs. Notably, the implemented benchtop NMR technique successfully resolved spectral signatures corresponding to PUFA, MUFA, SFA, TAG, cholesterol and phospholipids. ^{1}H NMR spectral data provided a powerful tool for

precise discrimination between Iberian and Celta breeds across individual meat cuts, facilitated by sPLS-DA models. Furthermore, machine learning classification models, particularly Support Vector Machines (SVM), allowed seamless differentiation between individual meat cuts regardless of breed, demonstrating robustness against the inherent complexity of meat samples and eliminating the need for prior meat cut separation.

Fatty acid quantification was achieved by developing regression models based on benchtop NMR peak integrals. Support Vector Machines with a Radial Basis Function kernel (SVM-RBF) emerged as the most effective approach for this purpose.

In summary, this work establishes the feasibility of chemometrics and machine learning for analysing benchtop spectroscopic NMR data for meat classification and FA content quantification. This approach offers a simple and efficient method requiring minimal sample preparation, potentially applicable to a wide range of food matrices.

CRediT authorship contribution statement

José Luis Ramiro: Writing – review & editing, Methodology, Investigation, Formal analysis. **Ana G. Neo:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Trinidad Pérez-Palacios:** Writing – review & editing, Supervision, Resources, Formal analysis, Conceptualization. **Teresa Antequera:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Carlos F. Marcos:** Writing – original draft, Supervision, Resources, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jose Luis Ramiro reports financial support was provided by Extremadura Employment Public Service (TE-0007-21). Other 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.

Data availability

Data will be made available on request.

Acknowledgments

Authors thank financial support from Junta de Extremadura and FEDER (GR21123). Authors also thank the companies Jamón y Salud and Cárnicas Teijeiro for technical support and sample supply, and the Association of Breeders of Porco Celta (ASOPORCEL) for their kind permission to use one of their photographs in the Supporting Material associated with this article. J. L. R. thanks Extremadura Employment Public Service and the European Social Fund for a research contract.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2024.110753.

References

Bachmann, R., Klockmann, S., Haerdter, J., Fischer, M., & Hackl, T. (2018). ^{1}H NMR spectroscopy for determination of the geographical origin of hazelnuts. *Journal of Agricultural and Food Chemistry, 66*(44), 11873–11879. https://doi.org/10.1021/acs.jafc.8b03724

Benitez, R., Nunez, Y., Fernandez, A., Isabel, B., Fernandez, A. I., Rodriguez, C., … Ovilo, C. (2015). Effects of dietary fat saturation on fatty acid composition and gene transcription in different tissues of Iberian pigs. *Meat Science, 102*, 59–68. https://doi.org/10.1016/j.meatsci.2014.12.005

Bermudez, R., Franco, I., Franco, D., Carballo, J., & Lorenzo, J. M. (2012). Influence of inclusion of chestnut in the finishing diet on fatty acid profile of dry-cured ham from Celta pig breed. *Meat Science, 92*(4), 394–399. https://doi.org/10.1016/j.meatsci.2012.05.001

Bischof, G., Januschewski, E., & Juadjur, A. (2024). Authentication of laying hen housing systems based on egg yolk using ^{1}H NMR spectroscopy and machine learning. *Foods, 13*(7), 1098–1109. https://doi.org/10.3390/foods13071098

Boccia, A. C., Cusano, E., Scano, P., & Consonni, R. (2020). NMR lipid profile of milk from alpine goats with supplemented hempseed and linseed diets. *Molecules, 25*(7), 1491–1503. https://doi.org/10.3390/molecules25071491

Brereton, R. G., & Lloyd, G. R. (2010). Support vector machines for classification and regression. *The Analyst, 135*(2), 230–267. https://doi.org/10.1039/b918972f

Cao, R., Liu, X., Liu, Y., Zhai, X., Cao, T., Wang, A., et al. (2021). Applications of nuclear magnetic resonance spectroscopy to the evaluation of complex food constituents. *Food Chemistry, 342*, 128258–128267. https://doi.org/10.1016/j.foodchem.2020.128258

Castejón, D., Mateos-Aparicio, I., Molero, M. D., Cambero, M. I., & Herrera, A. (2014). Evaluation and optimization of the analysis of fatty acid types in edible oils by ^{1}H-NMR. *Food Analytical Methods, 7*(6), 1285–1297. https://doi.org/10.1007/s12161-013-9747-9

Chong, I.-G., & Jun C, H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems, 78*(1–2), 103–112. https://doi.org/10.1016/j.chemolab.2004.12.011

Corominas, J., Ramayo-Caldas, Y., Puig-Oliveras, A., Estellé, J., Castelló, A., Alves, E., … Folch, J. M. (2013). Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition. *BMC Genomics, 14*(1), 843–857. https://doi.org/10.1186/1471-2164-14-843

Corsaro, C., Vasi, S., Neri, F., Mezzasalma, A. M., Neri, G., & Fazio, E. (2022). NMR in metabolomics: From conventional statistics to machine learning and neural network approaches. *Applied Sciences, 12*(6), 2824–2862. https://doi.org/10.3390/app12062824

Cuadros-Rodríguez, L., Pérez-Castaño, E., & Ruiz-Samblás, C. (2016). Quality performance metrics in multivariate classification methods for qualitative analysis. *TRAC Trends in Analytical Chemistry, 80*, 612–624. https://doi.org/10.1016/j.trac.2016.04.021

Cui, C., Xu, Y., Jin, G., Zong, J., Peng, C., Cai, H., et al. (2023). Machine learning applications for identify the geographical origin, variety and processing of black tea using ^{1}H NMR chemical fingerprinting. *Food Control, 148*, 109686–109697. https://doi.org/10.1016/j.foodcont.2023.109686

Deris, A. M., Zain, A. M., & Sallehuddin, R. (2011). Overview of support vector machine in modeling machining performances. *Procedia Engineering, 24*, 308–312. https://doi.org/10.1016/j.proeng.2011.11.2647

Domínguez, R., Martínez, S., Carballo, J., & Franco, I. (2014). Fatty acid profile and cholesterol and retinol contents in different locations of Celta pig breed. *Grasas Y Aceites, 65*(3), e036. https://doi.org/10.3989/gya.0115141

Domínguez, R., Martínez, S., Gómez, M., Carballo, J., & Franco, I. (2015). Fatty acids, retinol and cholesterol composition in various fatty tissues of Celta pig breed: Effect of the use of chestnuts in the finishing diet. *Journal of Food Composition and Analysis, 37*, 104–111. https://doi.org/10.1016/j.jfca.2014.08.003

Folch, J., Lees, M., & Stanley, G. H. S. (1957). A simple method for the isolation and purification of total lipids from animal tissues. *Journal of Biological Chemistry, 226*(1), 497–509. https://doi.org/10.1016/S0021-9258(18)64849-5

Galvan, D., de Aguiar, L. M., Bona, E., Marini, F., & Killner, M. H. M. (2023). Successful combination of benchtop nuclear magnetic resonance spectroscopy and chemometric tools: A review. *Analytica Chimica Acta, 1273*, 341495–341521. https://doi.org/10.1016/j.aca.2023.341495

Galvan, D., Tanamati, A. A. C., Casanova, F., Danieli, E., Bona, E., & Killner, M. H. M. (2021). Compact low-field NMR spectroscopy and chemometrics applied to the analysis of edible oils. *Food Chemistry, 365*, 130476. https://doi.org/10.1016/j.foodchem.2021.130476, 130476.

Gama, L. T., Martínez, A. M., Carolino, I., Landi, V., Delgado, J. V., Vicente, A. A., … Sousa, C. O. (2013). Genetic structure, relationships and admixture with wild relatives in native pig breeds from Iberia and its islands. *Genetics Selection Evolution, 45*(1), 18–32. https://doi.org/10.1186/1297-9686-45-18

García-Gudiño, J., Monteiro, N. T. R. A., Espagnol, S., Blanco-Penedo, I., & Garcia-Launay, F. (2020). Life cycle assessment of iberian traditional pig production system in Spain. *Sustainability, 12*(2), 627–645. https://doi.org/10.3390/su12020627

Giberson, J., Scicluna, J., Legge, N., & Longstaffe, J. (2021). Developments in benchtop NMR spectroscopy 2015–2020. In G. A. Webb (Ed.), *Annual reports on NMR spectroscopy* (pp. 153–246). Academic Press. https://doi.org/10.1016/bs.arnmr.2020.10.006.

Gkarane, V., Brunton, N. P., Harrison, S. M., Gravador, R. S., Allen, P., Claffey, N. A., … Monahan, F. J. (2018). Volatile profile of grilled lamb as affected by castration and age at slaughter in two breeds. *Journal of Food Science, 83*(10), 2466–2477. https://doi.org/10.1111/1750-3841.14337

González-Domínguez, R., Sayago, A., & Fernández-Recamales, Á. (2020). Fatty acid profiling for the authentication of iberian hams according to the feeding regime. *Foods, 9*(2), 149–158. https://doi.org/10.3390/foods9020149

Hajjar, G., Haddad, L., Rizk, T., Akoka, S., & Bejjani, J. (2021). High-resolution ^{1}H NMR profiling of triacylglycerols as a tool for authentication of food from animal origin: Application to hen egg matrix. *Food Chemistry, 360*, 130056–130062. https://doi.org/10.1016/j.foodchem.2021.130056

Hernández-Jiménez, M., Revilla, I., Hernández-Ramos, P., & Vivar-Quintana, A. M. (2024). Prediction of the fatty acid profiles of iberian pig products by near infrared spectroscopy: A comparison between multiple regression tools and artificial neural

Networks. *Food and Bioprocess Technology*. https://doi.org/10.1007/s11947-024-03486-x

Jakes, W., Gerdova, A., Defernez, M., Watson, A. D., McCallum, C., Limer, E., … Kemsley, E. K. (2015). Authentication of beef versus horse meat using 60 MHz ^{1}H NMR spectroscopy. *Food Chemistry, 175*, 1–9. https://doi.org/10.1016/j.foodchem.2014.11.110

Jiménez-Carvelo, A. M., González-Casado, A., Bagur-González, M. G., & Cuadros-Rodríguez, L. (2019). Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – a review. *Food Research International, 122*, 25–39. https://doi.org/10.1016/j.foodres.2019.03.063

Jiusheng, L. (2010). Support vector regression for the determination of the nutritional components of edible oil by Terahertz spectroscopy. *IEEE Transactions on Instrumentation and Measurement, 59*(8), 2094–2098. https://doi.org/10.1109/TIM.2009.2030927

Kasprzyk, A., Tyra, M., & Babicz, M. (2015). Fatty acid profile of pork from a local and a commercial breed. *Archives of Animal Breeding, 58*(2), 379–385. https://doi.org/10.5194/aab-58-379-2015

Klare, J., Rurik, M., Rottmann, E., Bollen, A., Kohlbacher, O., Fischer, M., et al. (2020). Determination of the geographical origin of Asparagus officinalis L. By ^{1}H NMR spectroscopy. *Journal of Agricultural and Food Chemistry, 68*(49), 14353–14363. https://doi.org/10.1021/acs.jafc.0c05642

Li, J., Zhang, J., Yang, Y., Zhu, J., He, W., Zhao, Q., … Zhang, J. (2021). Comparative characterization of lipids and volatile compounds of Beijing Heiliu and Laiwu Chinese black pork as markers. *Food Research International, 146*, 110433. https://doi.org/10.1016/j.foodres.2021.110433, 110433.

Liu, G. F., Fang, W. N., Lin, H. C., Wang, X. F., Fu, J. L., & Wang, A. G. (2009). Expression of HMGR and corresponding cholesterol content in tissues of two pig breeds. *Animal, 3*(10), 1436–1441. https://doi.org/10.1017/S1751731109990255

Mannina, L., Sobolev, A. P., Capitani, D., Iaffaldano, N., Rosato, M. P., Ragni, P., … Coppola, R. (2008). NMR metabolic profiling of organic and aqueous sea bass extracts: Implications in the discrimination of wild and cultured sea bass. *Talanta, 77*(1), 433–444. https://doi.org/10.1016/j.talanta.2008.07.006

Mestrelab Research, S. L. (2019). *MestReNova (mnova)*. Spain: Santiago de Compostela, 14.0.0 ed.).

Migues, I., Rivas, F., Moyna, G., Kelly, S. D., & Heinzen, H. (2022). Predicting Mandarin fruit acceptability: From high-field to benchtop NMR spectroscopy. *Foods, 11*(16), 2384–2391. https://doi.org/10.3390/foods11162384

Morcuende, D., Estevez, M., Ruiz, J., & Cava, R. (2003). Oxidative and lipolytic deterioration of different muscles from free-range reared Iberian pigs under refrigerated storage. *Meat Science, 65*(3), 1157–1164. https://doi.org/10.1016/S0309-1740(02)00344-3

Nieva-Echevarría, B., Goicoechea, E., Manzanos, M. J., & Guillén, M. D. (2014). A method based on ^{1}H NMR spectral data useful to evaluate the hydrolysis level in complex lipid mixtures. *Food Research International, 66*, 379–387. https://doi.org/10.1016/j.foodres.2014.09.031

Nyitrainé Sárdy, Á. D., Ladányi, M., Varga, Z., Szövényi, Á. P., & Matolcsi, R. (2022). The effect of grapevine variety and wine region on the primer parameters of wine based on ^{1}H NMR-spectroscopy and machine learning methods. *Diversity, 14*(2), 74–90. https://doi.org/10.3390/d14020074

Ortiz, A., Díaz-Caro, C., Tejerina, D., Escribano, M., Crespo, E., & Gaspar, P. (2021). Consumption of fresh Iberian pork: Two-stage cluster for the identification of segments of consumers according to their habits and lifestyles. *Meat Science, 173*, 108373. https://doi.org/10.1016/j.meatsci.2020.108373, 108373.

Pajuelo, A., Ramiro, J. L., Fernández-Marcos, M. L., Sánchez, S., Pérez-Palacios, T., Antequera, T., et al. (2023). Lipidomic analysis and classification of Iberian dry-cured hams with low field NMR. *Food Frontiers, 4*(4), 1810–1818. https://doi.org/10.1002/fft2.318

Pajuelo, A., Sánchez, S., Pérez-Palacios, T., Caballero, D., Díaz, J., Antequera, T., et al. (2022). ^{1}H NMR to analyse the lipid profile in the glyceride fraction of different categories of Iberian dry-cured hams. *Food Chemistry, 383*, 132371–132379. https://doi.org/10.1016/j.foodchem.2022.132371

Perez-Palacios, T., Ruiz, J., Martin, D., Muriel, E., & Antequera, T. (2008). Comparison of different methods for total lipid quantification in meat and meat products. *Food Chemistry, 110*(4), 1025–1029. https://doi.org/10.1016/j.foodchem.2008.03.026

Pérez-Palacios, T., Ruiz, J., Tejeda, J. F., & Antequera, T. (2009). Subcutaneous and intramuscular lipid traits as tools for classifying Iberian pigs as a function of their feeding background. *Meat Science, 81*(4), 632–640. https://doi.org/10.1016/j.meatsci.2008.10.022

Plieninger, T., Flinzberger, L., Hetman, M., Horstmannshoff, I., Reinhard-Kolempas, M., Topp, E., … Huntsinger, L. (2021). Dehesas as high nature value farming systems: A social-ecological synthesis of drivers, pressures, state, impacts, and responses. *Ecology and Society, 26*(3), 23–44. https://doi.org/10.5751/es-12647-260323

Pugliese, C., & Sirtori, F. (2012). Quality of meat and meat products produced from southern European pig breeds. *Meat Science, 90*(3), 511–518. https://doi.org/10.1016/j.meatsci.2011.09.019

R Core Team. (2023). *R: A language and environment for statistical computing*, 4.3.1 (16/06/2023.

Ramiro, J. L., Pérez-Palacios, T., Marcos, C. F., Neo, A. G., Pajuelo, A., & Antequera, T. (2024). Classification of raw cuts from Iberian and Celta pigs based on lipid analysis and chemometrics. *Journal of Food Composition and Analysis, 130*, 106173–106182. https://doi.org/10.1016/j.jfca.2024.106173

Rey, A. I., Daza, A., López-Carrasco, C., & López-Bote, C. J. (2006). Feeding Iberian pigs with acorns and grass in either free-range or confinement affects the carcass characteristics and fatty acids and tocopherols accumulation in Longissimus dorsi muscle and backfat. *Meat Science, 73*(1), 66–74. https://doi.org/10.1016/j.meatsci.2005.10.018

Saeed, M., Kim, J. S., Kim, S. Y., Ryu, J. E., Ko, J., Zaidi, S. F. A., … Choi, H. K. (2022). Differentiation of geographical origin of white and brown rice samples using NMR spectroscopy coupled with machine learning techniques. *Metabolites, 12*(11), 1012–1026. https://doi.org/10.3390/metabo12111012

Sandler, S. R., & Karo, W. (1992). *Sourcebook of advance organic laboratory preparations*. Academic Press.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science, 2*(3), 160–180. https://doi.org/10.1007/s42979-021-00592-x

Scholkopf, B., Kah-Kay, S., Burges, C. J. C., Girosi, F., Niyogi, P., Poggio, T., et al. (1997). Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing, 45*(11), 2758–2765. https://doi.org/10.1109/78.650102

Scollan, N. D., Price, E. M., Morgan, S. A., Huws, S. A., & Shingfield, K. J. (2017). Can we improve the nutritional quality of meat? *Proceedings of the Nutrition Society, 76*(4), 603–618. https://doi.org/10.1017/S0029665117001112

Sirtori, F., Dimauro, C., Bozzi, R., Aquilani, C., Franci, O., Calamai, L., et al. (2020). Evolution of volatile compounds and physical, chemical and sensory characteristics of Toscano PDO ham from fresh to dry-cured product. *European Food Research and Technology, 246*(3), 409–424. https://doi.org/10.1007/s00217-019-03410-0

Soyler, A., Cikrikci, S., Cavdaroglu, C., Bouillaud, D., Farjon, J., Giraudeau, P., et al. (2021). Multi-scale benchtop ^{1}H NMR spectroscopy for milk analysis. *LWT-Food Science & Technology, 139*, 110557–110566. https://doi.org/10.1016/j.lwt.2020.110557

Temperan, S., Lorenzo, J. M., Castiñeiras, B. D., Franco, I., & Carballo, J. (2014). Carcass and meat quality traits of Celta heavy pigs. Effect of the inclusion of chestnuts in the finishing diet. *Spanish Journal of Agricultural Research, 12*(3), 694–707. https://doi.org/10.5424/sjar/2014123-5057

Thompson, L., Rowntree, J., Windisch, W., Waters, S. M., Shalloo, L., & Manzano, P. (2023). Ecosystem management using livestock: Embracing diversity and respecting ecological principles. *Animal Frontiers, 13*(2), 28–34. https://doi.org/10.1093/af/vfac094

Valdés-Hernández, J., Ramayo-Caldas, Y., Passols, M., Sebastià, C., Criado-Mesas, L., Crespo-Piazuelo, D., … Folch, J. M. (2023). Global analysis of the association between pig muscle fatty acid composition and gene expression using RNA-Seq. *Scientific Reports, 13*(1), 535–547. https://doi.org/10.1038/s41598-022-27016-x

Vasconcelos, L., Dias, L. G., Leite, A., Ferreira, I., Pereira, E., Silva, S., … Teixeira, A. (2023). SVM regression to assess meat characteristics of Bísaro pig loins using NIRS methodology. *Foods, 12*(3), 470–485. https://doi.org/10.3390/foods12030470

Vigneau, E., Courcoux, P., Symoneaux, R., Guérin, L., & Villière, A. (2018). Random forests: A machine learning methodology to highlight the volatile organic compounds involved in olfactory perception. *Food Quality and Preference, 68*, 135–145. https://doi.org/10.1016/j.foodqual.2018.02.008

Vitale, M., Kallas, Z., Rivera-Toapanta, E., Karolyi, D., Cerjak, M., Lebret, B., … Oliver, M. A. (2020). Consumers' expectations and liking of traditional and innovative pork products from European autochthonous pig breeds. *Meat Science, 168*, 108179–108186. https://doi.org/10.1016/j.meatsci.2020.108179

Wang, H. P., Chen, P., Dai, J. W., Liu, D., Li, J. Y., Xu, Y. P., et al. (2022). Recent advances of chemometric calibration methods in modern spectroscopy: Algorithms, strategy, and related issues. *Trends in Analytical Chemistry, 153*, 116648–116663. https://doi.org/10.1016/j.trac.2022.116648

Wickramasinghe, I., & Kalutarage, H. (2020). Naive Bayes: Applications, variations and vulnerabilities: A review of literature with code snippets for implementation. *Soft Computing, 25*(3), 2277–2293. https://doi.org/10.1007/s00500-020-05297-6

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., … Whittington, F. M. (2008). Fat deposition, fatty acid composition and meat quality: A review. *Meat Science, 78*(4), 343–358. https://doi.org/10.1016/j.meatsci.2007.07.019

Yi, W., Huang, Q., Wang, Y., & Shan, T. (2023). Lipo-nutritional quality of pork: The lipid composition, regulation, and molecular mechanisms of fatty acid deposition. *Animal Nutrition, 13*, 373–385. https://doi.org/10.1016/j.aninu.2023.03.001

Zaukuu, J. L. Z., Benes, E., Bázár, G., Kovács, Z., & Fodor, M. (2022). Agricultural potentials of molecular spectroscopy and advances for food authentication: An overview. *Processes, 10*, 214–255. https://doi.org/10.3390/pr10020214

Zhong, M., Chong, Y., Nie, X., Yan, A., & Yuan, Q. (2013). Prediction of sweetness by multilinear regression analysis and support vector machine. *Journal of Food Science, 78*(9), S1445–S1450. https://doi.org/10.1111/1750-3841.12199