This document contains the **post-print pdf-version** of the referred paper:

"**Demystifying food flavor: Flavor data interpretation through machine learning**"

by **Huabin Luo**, **Simen Akkermans, Jan F.M. Van Impe**

which has been archived on the university repository Lirias (https://lirias.kuleuven.be/) of the Katholieke Universiteit Leuven.

The content is identical to the content of the published paper, but without the final typesetting by the publisher.

When referring to this work, please cite the full bibliographic info:

Luo, H., Akkermans, S., & Van Impe, J. F. M. (2025). Demystifying food flavor: Flavor data interpretation through machine learning. *Food Chemistry, 483*, Article 144000.

The journal and the original published paper can be found at:
https://www.sciencedirect.com/science/article/pii/S0308814625012518

The corresponding author can be contacted for additional info.

Conditions for open access are available at: http://www.sherpa.ac.uk/romeo/

Post print version of paper published in Food Chemsitry, 2025, Volume 483, Article 144000.

The content is identical to the published paper, but without the final typesetting by the publisher.

Demystifying food flavor:

Flavor data interpretation through machine learning

Huabin Luo[a], Simen Akkermans[a], Jan F.M. Van Impe[a,*]

[a] BioTeC+, Chemical and Biochemical Process Technology and Control, Department of Chemical Engineering, KU Leuven, Ghent, Belgium

[*] Corresponding author

E-mail address: jan.vanimpe@kuleuven.be (J.F.M. Van Impe)

Journal homepage: http://www.sciencedirect.com/journal/food-chemistry

Original file available at: https://www.sciencedirect.com/science/article/pii/S0308814625012518

Post print version of paper published in Food Chemsitry, 2025, Volume 483, Article 144000.

The content is identical to the published paper, but without the final typesetting by the publisher.

ABSTRACT

Flavor data obtained from analytical techniques are vast and complex, which increases the difficulty of multi-factorial analysis. This study aims to provide a machine learning (ML)-based framework to interpret flavor data, exploiting four widely used techniques, i.e., Principal Component Analysis (PCA), Redundancy Analysis (RDA), Partial Least Squares (PLS), and Random Forest (RF). To demonstrate the potential of these ML techniques, two case studies, one with semi-quantitative data and the other with quantitative data, were discussed. Results indicate that PCA is useful for data exploration; RDA can quantify the statistical significance of factors; combining feature importance analysis results from PLS and RF offer a comprehensive understanding of marker compounds. Regarding classification performance, PLS excels in handling collinear data, whereas RF captures complex patterns if sufficient data are available. However, overfitting is a risk for datasets with small sample sizes. Overall, carefully selecting and integrating those ML techniques could demystify food flavor.

Keywords: Flavor omics; Data mining; Flavor interpretation; Feature importance; Artificial Intelligence

Journal homepage: http://www.sciencedirect.com/journal/food-chemistry

Original file available at: https://www.sciencedirect.com/science/article/pii/S0308814625012518

1. Introduction

Flavor is one of the major sensory attributes affecting the acceptance and preference of foods by consumers. In general, food flavor is a combination of taste (non-volatile compounds), aroma (volatile compounds) and texture, among which aroma plays the most crucial role (Song & Liu, 2018). In this study, food flavor refers specifically to aroma. Flavor development in foods involves a complex interplay of (bio)chemical reactions and physical processes. There is a plethora of influencing factors, including but not limited to raw materials, processing conditions, packaging, and storage. Besides, microbial composition and fermentation time matter for fermented foods such as cheese, sausage, and beer. Together, these factors form the basis for food scientists to modulate flavor.

Sensory evaluation is a traditional flavor characterization approach that relies on well-trained and experienced experts but is often subjective. With rapid advances in analytical techniques, it has become more accessible to explore the flavor characteristics of foods objectively at the molecular level. These techniques mainly include gas chromatography-mass spectrometry (GC-MS) and its derivative techniques such as gas chromatography–olfactometry-mass spectrometry (GC-O-MS), gas chromatograph-ion mobility spectrometry (GC-IMS), and two-dimensional gas chromatography-time-of-flight-mass spectrometry (GC×GC-TOF-MS) (Wei et al., 2023). A detailed flavor profile is obtained based on the reported retention time of compounds, peak intensity, and mass-to-charge ratio, with suitable preprocessing steps (e.g., baseline correction, normalization, and peak alignment). Furthermore, key aroma-active compounds can be identified using dilution analytical approaches, including aroma extract dilution analysis (AEDA) and dynamic headspace dilution analysis

(DHDA), combined with two important parameters, flavor dilution (FD) factor and odor activity value (OAV) (Song & Liu, 2018). Additionally, external calibration or standard addition methods facilitate the quantification of volatile compounds in foods (Tufariello et al., 2019). Collectively, the application of these techniques enables flavor quality control.

Flavor data obtained from analytical techniques is vast, but also complex. Having a large amount of data does not mean that it can provide the desired answers unless the data is converted into useful information (Jiménez-Carvelo et al., 2019). Simply listing volatile compounds in tables or depicting them using heatmaps/Principal Component Analysis (PCA) cannot fully demystify flavor data. Machine learning (ML) refers to a group of computer-based methods capable of learning complex patterns from large datasets to predict outcomes, which improves identification accuracy, simplifies model predictions, and avoids subjectivity in data analysis (Ji et al., 2023). Currently, the application of ML in the food industry is gaining ground (Mavani et al., 2022; Oztuna Taner & Çolak, 2024). Regarding food flavor studies, Zeng et al. (2023) proposed the concept of food flavor analysis 4.0 and defined it as combining modern flavor analysis methods with ML. Our recent review pointed out that data-driven modeling techniques aid cheese flavor analysis (Luo et al., 2024). ML techniques have recently begun to be applied in flavor studies. Authors usually focus on one specific aspect, such as marker volatiles identification or sample classification (Qiu et al., 2024; Xiang et al., 2025; Zhang et al., 2025), which cannot fully demystify the flavor data. A systematic framework for flavor data interpretation through ML is necessary, however, to the best of our knowledge, this information is unavailable. Besides, food scientists generally

lack experience in modeling, and detailed guidance on the implementation of ML techniques would be beneficial.

This study aims to provide a framework for flavor data interpretation through ML (Figure 1). The main tasks of flavor interpretation are (i) to determine how factors influence flavor profiles, and (ii) to find marker compounds. To address these tasks, four widely used ML techniques are suggested, including PCA, Redundancy Analysis (RDA), Partial Least Squares (PLS), and Random Forest (RF). Specifically, PCA is used as an exploratory approach, and RDA is used to quantify the significance of influencing factors. On the other hand, PLS and RF are conducted for feature importance analysis and predictions. To demonstrate these data mining techniques, two representative datasets were used as case studies: semi-quantitative flavor data from Melkonian et al. (2023) and quantitative flavor data from Wang et al. (2024). Overall, this study provides technical tools and guidance on flavor data interpretation for food scientists.

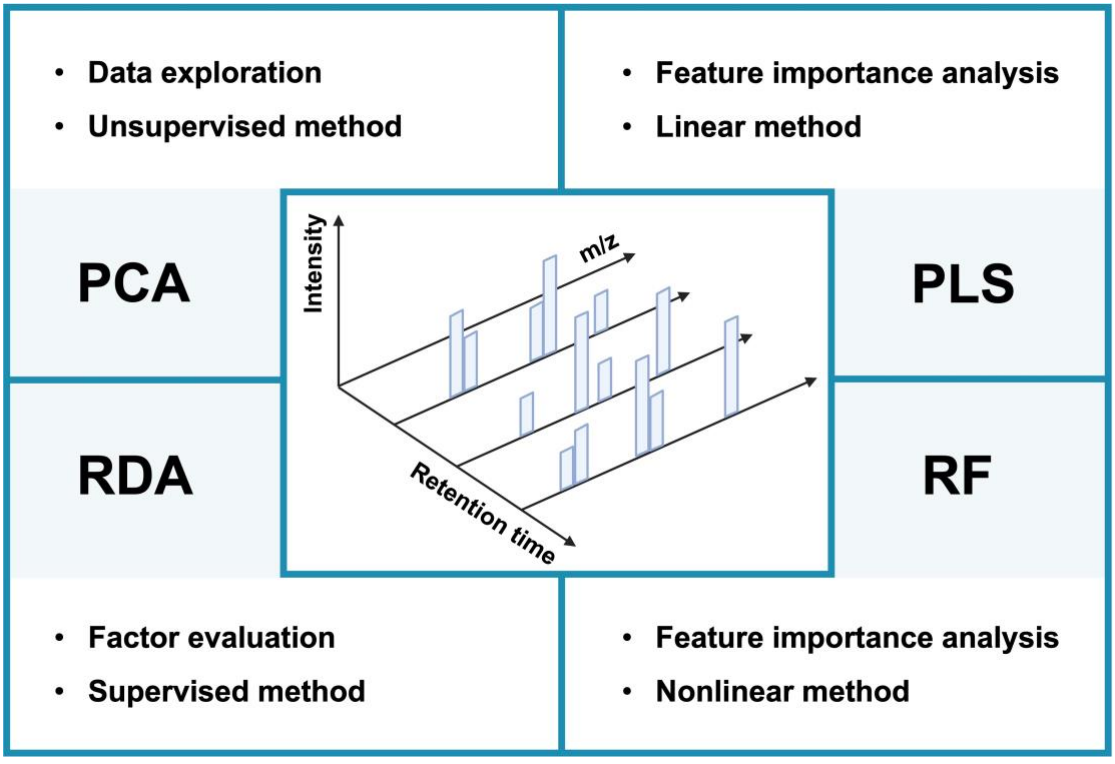


Figure 1. Framework for flavor data interpretation in this study, including Principal Component Analysis (PCA), Redundancy Analysis (RDA), Partial Least Squares (PLS), and Random Forest (RF).

2. Materials and methods

2.1.Flavor datasets

The flavor dataset can be considered as a matrix consisting of *n* rows (samples) and *p* columns (variables, in this case volatile compounds), with each cell containing a numerical value of the measurement (peak area, height, or concentration). Given a dataset, if the number of samples is greater than the number of volatile compounds, it is considered a large dataset, and vice versa. Usually, the quantities of volatile compounds reported in the literature are semi-quantitative based on the use of a suitable internal standard. Alternatively, accurate quantification of volatile compounds is achievable by building a reliable calibration curve with external standards. However, this method is time-consuming and requires a calibration curve for each matrix and compound, limiting the number of samples that can be measured. Both cases are discussed in this study.

Case Study 1: semi-quantitative flavor data. Melkonian et al. (2023) studied the role of microbial interactions in cheese flavor formation by employing a strain dropout strategy. Specifically, the authors prepared a defined starter culture under non-industrial conditions by hand (referred to as the condition HP). Thereafter, the effects of leaving out *Streptococcus thermophilus* (ST) or the *Lactococcus* blend (LB) (corresponding to condition LOST or LOLB, respectively) were investigated in a year-long Cheddar cheese ripening experiment. Volatile compounds were analyzed using solid-phase microextraction and GC-MS. A total of 64 samples with different microbiota

composition and ripening time were measured, and 28 volatile compounds were reported.

Case Study 2: quantitative flavor data. Wang et al. (2024) obtained a quantitative and dynamic understanding of the key aroma compounds for *Monascus*-fermented cheese across a 90-day ripening period. Volatile compounds were analyzed using GC-MS. Key aroma compounds were identified using aroma extract dilution analysis (AEDA) and odor activity value (OAV) analysis. Besides, external standard curves were constructed using GC-MS for the quantification of volatile compounds. A total of 12 samples with different ripening time were quantified, and 36 volatile compounds were reported.

2.2.Preprocessing and data mining

Min-max normalization was conducted to transform the data of various dimensions into the range [0,1]. Thereafter, from a mathematical perspective, the normalized volatile compound data can be regarded as *X*-variables. In addition, the experimental condition of the sample is considered the *Y*-variable. In these case studies, this was the ripening time and/or microbiota composition.

Preprocessing and data mining were performed in RStudio (version 4.4.0) with packages “vegan”, “mixOmics”, “randomForest”, “caret”, “ggfortify”, “ggplot2”, “dplyr”, and “ggrepel”.

2.3.Machine learning techniques

ML techniques can be classified into unsupervised and supervised ones, where the latter relies on the existence of labels in the training set (Alloghani et al., 2020). In this study, four widely used techniques were introduced below, i.e., PCA, RDA, PLS, and RF.

PCA is an unsupervised technique used for dimensionality reduction and pattern recognition (Cozzolino et al., 2019), meaning that only the information of the *X*-variables is used for data mining. In principle, PCA linearly transforms the original data into a new coordinate system; these new coordinates are called the principal components (PCs), of which the first PC captures the largest proportion of variance of the original data followed by the second PC and so on. Usually, the first two or three PCs are plotted for visualization.

RDA is a supervised technique with similarities to PCA. Compared with PCA, RDA also uses the class label information (i.e., the *Y*-variable) for data mining. In principle, RDA combines regression and canonical analysis to understand how much variance in the response variables (*X*-variables) can be explained by the explanatory variable (*Y*-variable). An essential element of RDA is the use of permutation tests, which are used to assess the statistical significance of the observed relationships (Zuur et al., 2007). Specifically, the permutation test involves randomly permuting the *X*-variables and recalculating the RDA statistic (e.g., the explained variance or canonical eigenvalues) to create a null distribution for comparison. This test allows to determine whether the observed relationships are statistically significant or simply due to chance.

PLS is a supervised technique and takes into account the information of both the *X*- and *Y*-variables. In principle, PLS is used to find relationships between predictor variables (*X*-variables) and the response variable (*Y*-variable) by maximizing the covariance. Depending on whether the *Y*-variable is continuous or categorical, PLS can be further classified into regression and classification models. Once the mathematical relationship is established, the model can be tested on new samples for prediction. The variable importance in projection (VIP) score in PLS models is a statistical tool used to measure the importance of each variable in explaining the variation in both the predictor and response variables (Wold et al., 1993). VIP scores are particularly useful in models with many variables, as they assist in feature selection by highlighting those that are most relevant. Mathematically, let us assume the PLS model has *K* components. For each predictor variable *j* (in this case volatile compound), the VIP score is calculated as (Pérez-Enciso & Tenenhaus, 2003):

$$VIP_j = \sqrt{\frac{p}{\sum_{k=1}^{K} SSY_k} \sum_{k=1}^{K} SSY_k \left(\frac{w_{jk}^2 \cdot SSX_k}{\sum_{j=1}^{p} w_{jk}^2} \right)}$$

where p is the total number of variables; w_{jk} is the weight of the jth variable on the kth component; SSX_k is the sum of squares explained by the kth component for the *X*-variables; SSY_k is the sum of squares explained by the kth component for the *Y*-variable. The VIP score aggregates the weighted sum of squares for each variable across all components, normalized by the total explained variance. Variables with VIP scores larger than 1 are typically considered as more influential in the model.

Different from PLS, **RF** is a nonlinear technique, which achieves prediction by creating multiple decision trees with random sampling and aggregation. Also, RF can be used to rank variable importance, with the key advantage being that it covers the impact of

each predictor variable individually as well as in multivariate interactions with other predictor variables. In principle, RF computes the importance of the variable by permuting out-of-bag (OOB) samples and calculating the mean decrease accuracy (MDA) as the index. The larger the MDA value, the more important the volatile compound. Mathematically, for each variable j (in this case volatile compound), the MDA is calculated as (Han et al., 2016):

$$MDA_j = \frac{1}{n_{tree}} \sum_{t=1}^{n_{tree}} (EP_{tj} - E_{tj})$$

where n_{tree} is the number of trees in the forest; E_{tj} is the OOB error on tree t before permuting the values of X_j; EP_{tj} is the OOB error on tree t after permuting the values of X_j.

2.4.Model evaluation

To evaluate the classification performance of Partial Least Squares-Discriminant Analysis (PLS-DA) and RF models, a 10-fold cross validation was employed for Case Study 1 (16 samples for each ripening time/microbiota composition class). Considering the limited number of samples in Case Study 2 (three samples for each ripening time class), a leave-one-out cross validation was used. The RF model was constructed with 500 trees and the number of variables at each split was optimized during RF model training. The PLS-DA model was optimized by adjusting the number of latent variables to achieve the best accuracy. The model accuracy was calculated using the formula below:

$$Accuracy = \frac{Correct\ predictions}{Total\ predictions} \times 100\%$$

Confusion matrices for both PLS-DA and RF models with the best accuracy were plotted using GraphPad Prism 10.0.

3. Results and discussion

First, the four ML techniques were applied to a large dataset (Case Study 1) to explain their application potential with respect to data exploration, factor evaluation, and feature importance analysis. Next, the extent to which these ML techniques are applicable to a small dataset (Case Study 2) was evaluated. Then, the classification performance of PLS-DA and RF was assessed in both case studies. Moreover, specific suggestions for applying the proposed framework in real-world scenarios were given to finalize the discussion.

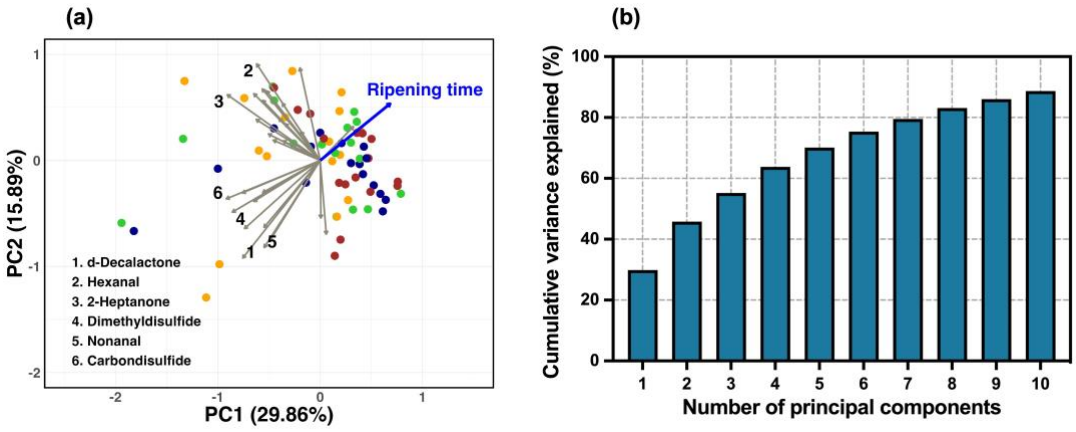
3.1.Data exploration by PCA

As an exploratory technique, PCA is often the first step in data mining. The obtained result is generally presented in a two-dimensional biplot, which is a combined graphical representation of samples and variables in a reduced-dimensional space (Buvé et al., 2022). For food scientists, it is important to understand how to interpret PCA biplots and what information can be extracted for flavor interpretation. The **variance explained** by each PC is one of the key messages, with a higher percentage indicating that the PC captures more variance of the original dataset. Secondly, **scores** are the coordinates of samples in the new coordinate system. Samples that are close to each other have similar characteristics with respect to the original variables. Thirdly, **loadings** are vectors that show the contribution of each original variable to the PCs, with a longer vector suggesting a stronger influence on the PCs. Besides, **supplementary variables** can be added to the biplots to aid interpretation, which are

not used in the PCA computation but help in understanding how additional information relates to the PCs without altering the original analysis (Greenacre et al., 2022).

A concrete example is given to demonstrate PCA using the dataset of Case Study 1. As shown in Figure 2a, the first two PCs explained 45.75% of the total variance, which means that half of the information in the *X*-variables that contain the GC-MS data on the volatiles is captured. The percentage of variance explained by other PCs is presented in the scree plot (Figure 2b). Usually, the first question of interest is how many components need to be considered. In exploratory studies, the answer is "the first few components" because the purpose is usually just to "have a look at the data" (Bro & K. Smilde, 2014). When using PCA for other purposes, 80% variance explained is often used as a default threshold for choosing the number of components (Kherif & Latypova, 2020). Next to this, it's important to interpret the similarities and differences between samples using scores. Samples were displayed with points of different colors according to the microbiota composition (Figure 2a). The result indicated that most samples were located on the right side of the biplot, but no clear trend was noted in the grouping of samples with different microbiota compositions. Besides, the contribution and correlation of variables can be obtained by observing the length of the vectors on PCs and the cosine of the angle between the vectors, respectively. Variables (in this case volatile compounds) were represented by grey arrows from the origin in different directions. To make the figure easy to read, only the top 6 volatile compounds were labeled. The horizontal (PC1) and vertical (PC2) axes reflect the relative contribution of compounds, quantified by their loadings. PC1 is dominated by sulfur-containing compounds, with carbondisulfide (loading = -0.92) and dimethyldisulfide (loading = -0.86) exhibiting large contributions. Aldehydes (e.g., hexanal, loading = 0.91) and

lactones (e.g., d-decalactone, loading = -0.92) mainly govern PC2, but show opposite trends. These findings suggest that PC1 distinguishes samples through sulfur-related biochemical pathways, while PC2 reflects contrasting trends in lipid oxidation and flavor development. In terms of correlation, when two variables point in the same direction, such as dimethyldisulfide and d-decalactone, a positive correlation is suggested. Conversely, for two variables pointing in an opposite direction, a negative correlation is indicated. In addition, the relationship between samples and variables can be obtained by projecting a sample point onto a loading vector. For example, the projections of two orange points on the lower left of the biplot fall in the same direction of vectors such as d-decalactone and nonanal, which means that the content of the two volatile compounds in these samples is above average. Furthermore, to enrich the interpretation, ripening time was added as a supplementary variable to the biplot, as indicated by the blue arrow in Figure 2a. The interpretation of such a supplementary variable is exactly the same as that of volatile compounds. Samples located on the upper right of the biplot showed longer ripening time. Besides, there was a clear effect of ripening time on volatile compounds. For example, compounds such as carbondisulfide and dimethyldisulfide were negatively correlated with ripening time, whereas hexanal was not correlated. In summary, PCA suggests a trend in the volatiles with respect to ripening time, but no clear insight into the effect of microbiota composition on volatile compounds was obtained.



271

272 Figure 2. PCA biplot representing changes in volatile compounds of Cheddar cheese
273 samples in a year-long ripening (a) and scree plot showing the cumulative variance
274 explained by the first ten principal components (b). Samples are displayed by points with
275 different colors according to the microbiota composition: lime green (●) for commercial
276 starter culture, dark blue (●) for starter culture prepared under non-industrial conditions
277 by hand, orange (●) for starter culture prepared by leaving out *Streptococcus*
278 *thermophilus*, and brown (●) for starter culture prepared by leaving out the *Lactococcus*
279 blend. Volatile compounds are displayed by gray arrows and the top 6 are labeled. To
280 enrich the interpretation, ripening time is added as a supplementary variable, as indicated
281 by the blue arrow.

282

283 *3.2.Factor evaluation by RDA*

284 RDA is routinely applied in ecology to determine which environmental factors
285 significantly explain community responses (Blanchet et al., 2014). However, its
286 application in flavor studies is still limited. As inspired, RDA is proposed to evaluate
287 the effect of factors on flavor profiles. Unlike PCA, RDA is a supervised technique that
288 constrains the information of the *Y-variable* to the main axes (components). The
289 obtained results can be represented by a triplot of samples (points), response variables
290 (arrows), and explanatory variables (arrows for continuous variables and centroids or
291 ellipses for categorical variables) (Ramette, 2007). Optionally, supplementary variables
292 can be added to enrich interpretation. By examining the above elements, the triplot can
293 provide a comprehensive understanding of the relationships between them and help

uncover patterns within the data. Besides, permutation tests are performed to assess the statistical significance of the observed relationships, with a p-value below 0.05 indicating significant differences (Borcard et al., 2018).

Consistent with PCA, the dataset of Case Study 1 will be used as an example to demonstrate RDA. To better understand the effect of microbiota composition on flavor profiles (which was unclear by PCA), the microbiota composition was used as an explanatory variable (in this case a categorical variable). First, RDA was constructed considering all conditions, i.e., control, HP, LOST, and LOLB, as shown in Figure 3a. The result showed that microbiota composition, which is illustrated by the ellipsoids, had a significant effect on flavor profiles ($p=0.003$). To figure out the exact cause, a series of RDAs were performed using different combinations of microbiota compositions, as shown in Figure 3b-e. No significant differences ($p=0.473$) were noted between samples from the control and HP group. This indicated that the starter culture that was prepared by making a combination of pure cultures (HP) led to similar volatile compositions as the commercial starter culture (Figure 3b). Next, the flavor profiles were compared between samples of the groups HP and LOST (Figure 3c), HP and LOLB (Figure 3d), and LOST and LOLB (Figure 3e). The result indicated that leaving out *Streptococcus thermophilus* from the defined starter culture had a significant effect on flavor profiles ($p=0.008$), whereas no significant differences were noted by leaving out the *Lactococcus* blend ($p=0.287$). This finding was consistent with the result reported in the original work by Melkonian et al. (2023). Furthermore, a significant difference ($p=0.001$) between samples from the group LOST and LOLB was noticed. Although this finding was not noted in the original work, it can be related to the same observation of a significant effect of leaving out *S. thermophilus*. Besides, in agreement

with the findings obtained in PCA, a clear effect of ripening time (as indicated by the blue arrows) on flavor profiles was observed in all RDA triplots.

The specific effects of the removal of a strain on volatile compounds were observed from the relationships between samples (points with color), response variables (gray arrows), and explanatory variables (ellipses with color). The top 5 volatile compounds were labeled for each RDA. Taking the RDA of the combination of LOST and LOLB as an example, a distinct separation of samples was noticed. The distance between the sample points approximates the dissimilarities regarding volatile compounds. Samples close to each other or within the ellipse present similar flavor profiles. Additionally, samples located in the direction of a volatile compound vector indicate a relatively high quantity of that volatile. For instance, ethanol was the primary volatile compound to separate samples and was abundant in samples from the LOST group. Regarding the supplementary variable (ripening time), it was found to be positively correlated with hexanal and negatively correlated with d-decalactone. Overall, RDA is a powerful technique to quantify the statistical significance of influencing factors.

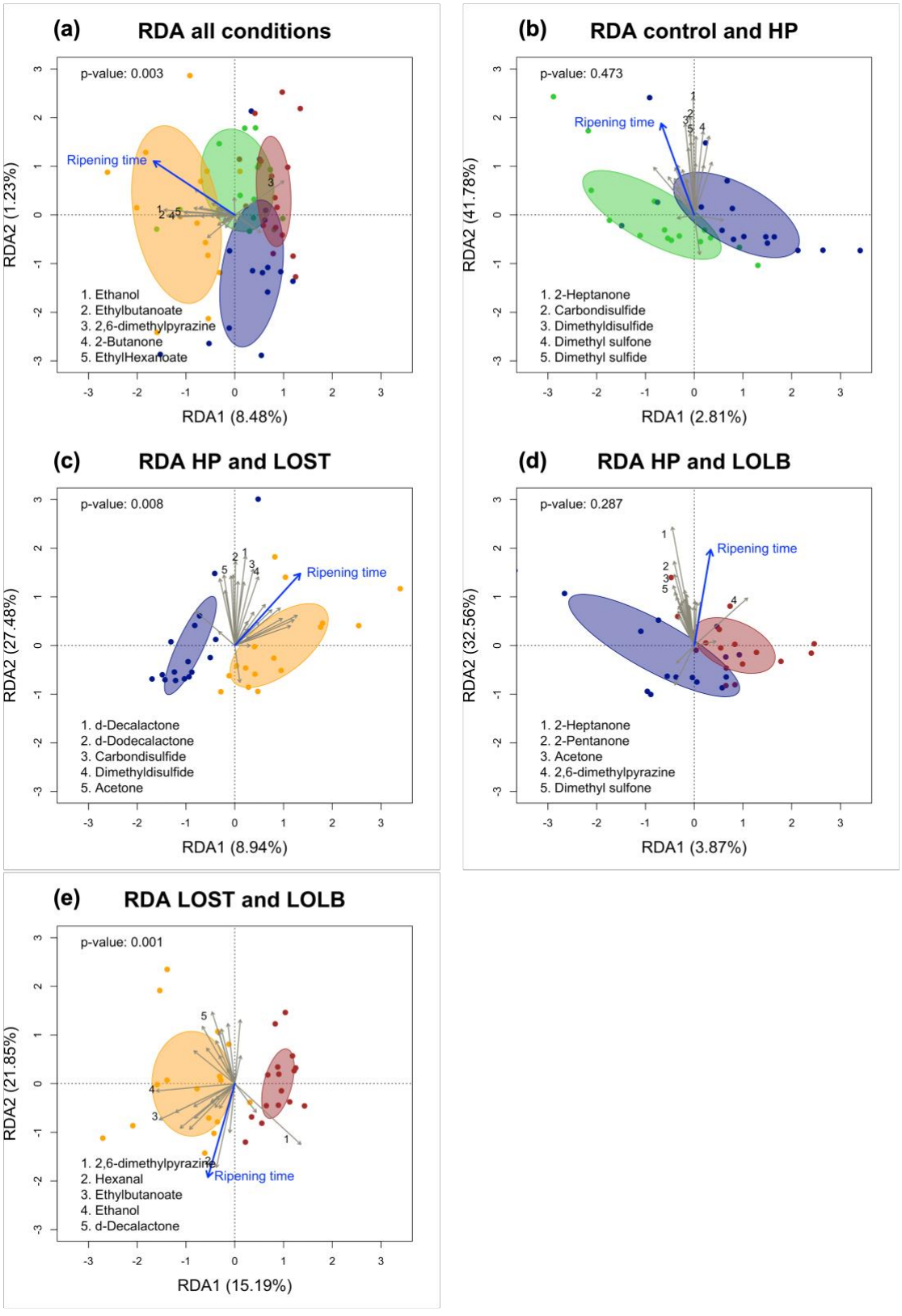


Figure 3. RDA triplots representing the effect of microbiota composition (explanatory variable) on Cheddar cheese flavor profiles during a year-long ripening. Samples are colored according to the microbiota composition: lime green (●) for commercial starter culture (control), dark blue (●) for starter culture prepared under non-industrial

conditions by hand (HP), orange (●) for starter culture prepared by leaving out *Streptococcus thermophilus* (LOST), and brown (●) for starter culture prepared by leaving out the *Lactococcus* blend (LOLB). Volatile compounds (response variables) are represented by gray arrows. To make the plots easy to read, only the top 5 volatile compounds are labeled. Ripening time is added as a supplementary variable to enrich interpretation, as indicated by the blue arrows. RDA is performed under different combinations: (a) all conditions (control, HP, LOST, and LOLB), (b) control and HP, (c) HP and LOST, (d) HP and LOLB, and (e) LOST and LOLB. The p-value of the permutation test is shown in the upper left corner of the triplot, with a value below 0.05 indicating significant differences.

3.3.Feature importance analysis by PLS and RF

After identifying the effects of ripening time and microbiota composition on Cheddar cheese flavor profiles, it is time to perform feature importance analysis to find associated marker compounds. Considering the large number of volatile compounds present in foods, it is costly and time-consuming to analyze all of them. Careful selection of marker compounds through ML enables efficient monitoring of flavor quality. The use of VIP scores in PLS models has been reported to identify marker compounds (e.g., Fan et al., 2019; Zhang et al., 2022). On the other hand, there are only few implementations of the MDA value in RF models for feature importance analysis in the literature (e.g., Xiang et al., 2025; Zhao et al., 2022). In this section, these two techniques are discussed and compared using the dataset of Case Study 1. Since the *Y-variable* used in this study is a categorical variable, a PLS-DA model was performed.

First, 28 volatile compounds were ranked based on the PLS-DA and RF models, respectively. Notably, the obtained VIP scores and MDA were not in the same range. Therefore, in order to make the results comparable, the values were normalized to the range [0,1]. Generally, ‘great than one rule’ is used as a criterion for variable selection,

368 which means that variables with VIP scores above 1 are considered important (Chong
369 & Jun, 2005). For MDA values, there is no agreement on the threshold that indicates
370 variable importance. Thus, volatile compounds were ranked based on the VIP scores
371 before normalization and were divided into two groups: VIP > 1 and VIP < 1.
372 Similarities and differences were noted in the screening of maker compounds
373 concerning the ripening time effect (Figure 4a) and the microbiota composition effect
374 (Figure 4b). Specifically, both PLS-DA and RF models highlighted ethyl hexanoate,
375 2,6-dimethylpyrazine, ethyl decanoate, ethyl butanoate, 2-pentanone, and 3-
376 methylbutanal as marker compounds for both influencing factors (ripening time and
377 microbiota composition). This indicates that these 6 compounds likely have strong and
378 consistent associations with ripening time and microbiota composition, making them
379 robust predictors reported by both models. In terms of differences, only one compound
380 was reported as the least important in PLS-DA, whereas for RF multiple compounds
381 had an equally low importance. Moreover, there were more differences in the marker
382 compounds suggested by the two models for the microbiota composition effect than for
383 the ripening time effect. For example, ethyl octanoate was considered important by
384 PLS-DA, whereas it was the least important in RF. These differences may be explained
385 by the following three aspects.

386 (i) Linearity and nonlinearity of the technique: PLS-DA is a linear technique and
387 focuses on finding linear combination of predictors that best separates the classes.
388 It shows advantages in handling collinear data (Gromski et al., 2015). RF is a
389 nonlinear and ensemble method that builds multiple decision trees using random
390 subsets of data and predictors. It can capture nonlinear relationships and
391 interactions between variables due to its tree-based structure. Although RF is

robust to overfitting by averaging the results of different trees, it still tends to overfit on small datasets due to the limited number of trees.

(ii) Mechanism of variable importance measures: PLS-DA selects variables based on their contribution to the maximization of the explained variance in the response through latent variables (Rossini et al., 2012). In contrast, RF selects variables based on how much they decrease the model's predictive accuracy when permuted (Strobl et al., 2007). A key advantage of RF is that it covers the impact of each variable individually as well as in multivariate interactions with other variables (Strobl et al., 2008).

(iii) Sensitivity to data structure: PLS-DA is sensitive to the structure and distribution of the data, whereas RF is relatively insensitive to variations in data structure due to its random sampling and averaging procedures (Shi et al., 2019).

Recently, different attempts have been made to identify marker compounds. Li et al. (2024) investigated the OAV of compounds with VIP > 1 to screen key flavor compounds in *baijiu* samples. OAV is considered an experimental validation approach, and this combination is beneficial for limited samples with sufficient resources. In another study by Xiang et al. (2025), a Venn plot was created for potential marker compound screening by combining the VIP score from PLS-DA, MDA value from RF, and fold change analysis result from the volcano plot. The intersection of the three analyses was considered as the marker compound. Integrating the results from different analyses is insightful, but consistent evaluation is required for the results to be reliable. Overall, both the VIP score in the PLS-DA model and the MDA value in the RF model provide valuable insights into which variables are most influential. By combining these two techniques, a more comprehensive understanding of feature importance can be

obtained. Variables that consistently show high importance in both models are robust marker compounds.

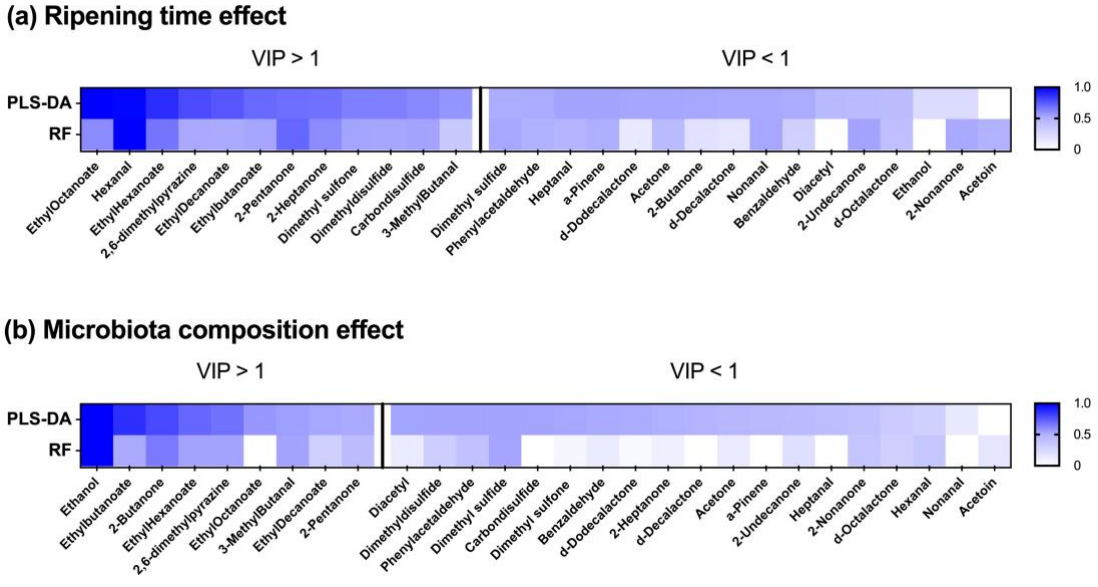


Figure 4. The screening of marker compounds with respect to (a) ripening time and (b) microbiota composition. Volatile compounds were ranked based on the variable importance in projection (VIP) score of the Partial Least Squares Discriminant Analysis (PLS-DA) model and the mean decrease accuracy (MDA) value of the Random Forest (RF) model. Variables with VIP scores above 1 are considered important. The results are presented as normalized [0,1] VIP scores and MDA values.

3.4.Flavor interpretation with small datasets

The above discussions are based on a dataset with more samples than variables (volatiles), whereas in practice, researchers often face challenges with small datasets with fewer samples than variables given measurement costs and other limitations. Therefore, it is essential to discuss flavor interpretation in such a case with a small sample size.

Case Study 2 (12 samples and 36 variables) is taken as an example. Similar to Case Study 1, PCA was first applied for data exploration. As shown in Figure 5a, the first two PCs explained 83.13% of total variance present in the original dataset. The biplot

illustrates that samples were clearly clustered along PC1 and PC2 according to ripening time. Although ripening time was added as a supplementary variable (indicated by a blue arrow) to the biplot, the increase in ripening time did not follow the direction of the arrow. Volatile compounds were represented by gray arrows. By projecting the sample point perpendicularly onto the variable vector, it was noted that 10 d samples had a high content of 2-petanone and styrene, 40 d samples had a high content of 1-dodecanol and octanoic acid ethyl ester, and 90 d, 0 d samples had a low content of butanoic acid. Next, RDA was performed to investigate the effect of ripening time on flavor profiles by adding it as an explanatory variable. However, no significant difference was noticed as indicated by the high p-value. Looking at the RDA triplot (Figure 5b), a distinct separation was noted among samples, but in the direction of RDA1, samples were not arranged in increasing order of ripening time. Most of the volatile compounds were not correlated with ripening time, as shown by the cosine of the angle between them. Besides, it was noted that samples ripened for 10 d or 40 d were located in the same direction of volatile compounds, whereas samples ripened for 0 d or 90 d were located in the opposite direction, which suggests a low content of volatile compounds at 0 d and 90 d. These findings indicate that there might be a nonlinear relationship between ripening time and volatile compounds, which is why RDA as a supervised linear technique cannot demonstrate this relationship.

Lastly, PLS-DA and RF were performed to screen marker compounds regarding ripening time. As shown in Figure 5c, both techniques highlight the importance of compounds such as 2-tridecanone, octanoic acid ethyl ester, butanoic acid, and 2,3-butanediol. In the original work, Wang et al. (2024) used OAV to quantify the potential contribution of individual compounds to the overall flavor, which is the ratio of the

concentration of a single compound to the odor threshold of that compound. Compounds with an OAV above 1 are considered important. The OAVs of the above compounds reported by the models were all greater than 1 in samples at different ripening times. This indicates the feasibility of screening marker compounds by PLS-DA and RF. Notably, the results of PLS-DA should be treated with caution because it is a linear technique, whereas the observed time effect is nonlinear. Although the OAV method helps identify marker compounds in individual samples, it falls short in providing insights that differentiate samples across varying conditions. In summary, when interpreting flavor data with small sample sizes, overfitting is a risk for linear techniques such as PCA, RDA, and PLS-DA, which can result in a distinct separation of samples without finding statistical significance.

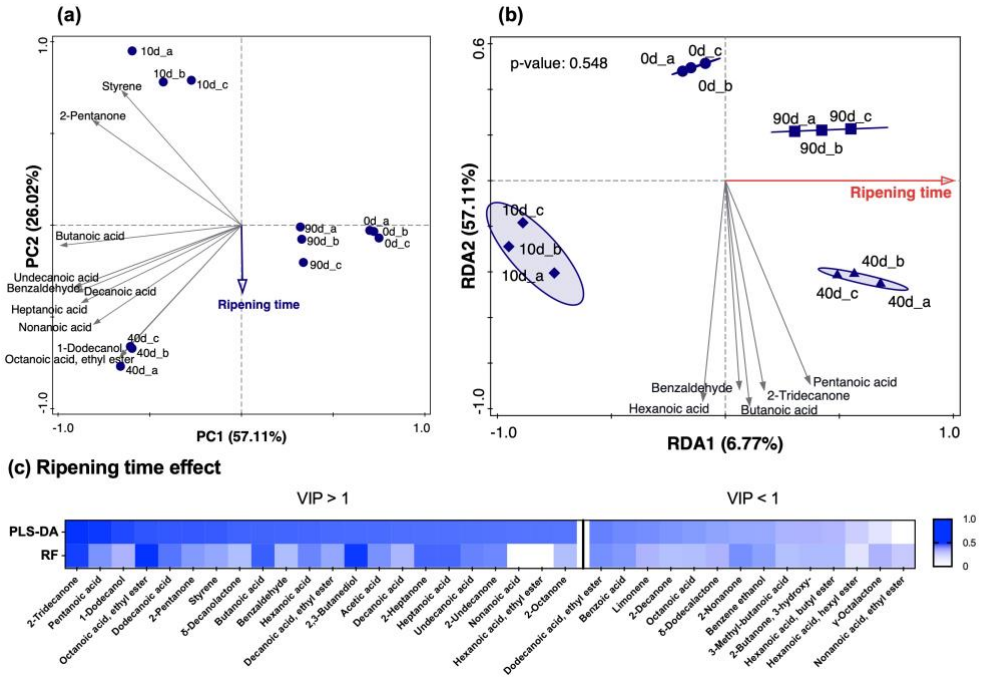


Figure 5. Flavor interpretation in a small dataset, using (a) PCA for data exploration, (b) RDA for factor evaluation, and (c) PLS-DA and RF for feature importance analysis. In (a) and (b), samples are represented by symbols, response variables are represented by gray arrows, the supplementary variable in PCA is displayed by a blue arrow, the explanatory variable in RDA is displayed by a red arrow, and the ripening time in days is indicated by

the number that initiates the name of each sample. In (c), variables with variation in projection (VIP) scores above 1 are considered important.

3.5.Classification performance of PLS-DA and RF

The classification performance of PLS-DA and RF was evaluated in both case studies. Accuracy was computed by determining the percentage of samples correctly classified. For Case Study 1, both models reported higher accuracy when classifying samples based on ripening time compared with the classification based on microbiota composition (Figure 6a-d). This finding is not surprising, as strong correlations between volatile compounds and ripening time were noticed by RDA (Section 3.2). In contrast, the relationship between microbiota composition and volatile compounds is more complex. Since RF is good at capturing complex patterns, the resulting model could be assumed to have a high accuracy. However, a highest accuracy was reported in the PLS-DA model (Figure 6a-b). The incorrect classification of control samples was the main reason for the low accuracy of the RF model. Notably, 6 samples were wrongly classified as HP samples. According to the RDA result (Figure 3b), there were no significant differences in the flavor profiles of the control and HP samples, which explains why RF had difficulties in correctly classifying samples between the control and HP group. Another reason for the low accuracy of RF might be that RF requires a larger sample size, while the number of samples in this case study was still limited. In terms of classifying LOST samples, both models achieved high accuracy. From the perspective of model optimization, the number of decision trees and the number of variables in the random subset at each node in the forest are important parameters that affect the performance of RF (Li et al., 2017). Regarding PLS-DA, the number of latent variables is important. For Case Study 2, both models correctly classified all samples, as demonstrated by the confusion matrices in Figure 6e-f. In general, small datasets

have an increased chance of overfitting, even more so in the case of the nonlinear RF model. However, for the current small dataset (12 samples), this was not observed. In summary, when volatile compounds are highly correlated with class labels, both PLS-DA and RF can achieve high classification accuracy. RF is more suitable for recognizing complex patterns with sufficient data. For datasets with small sample sizes, the risk of overfitting should be considered.

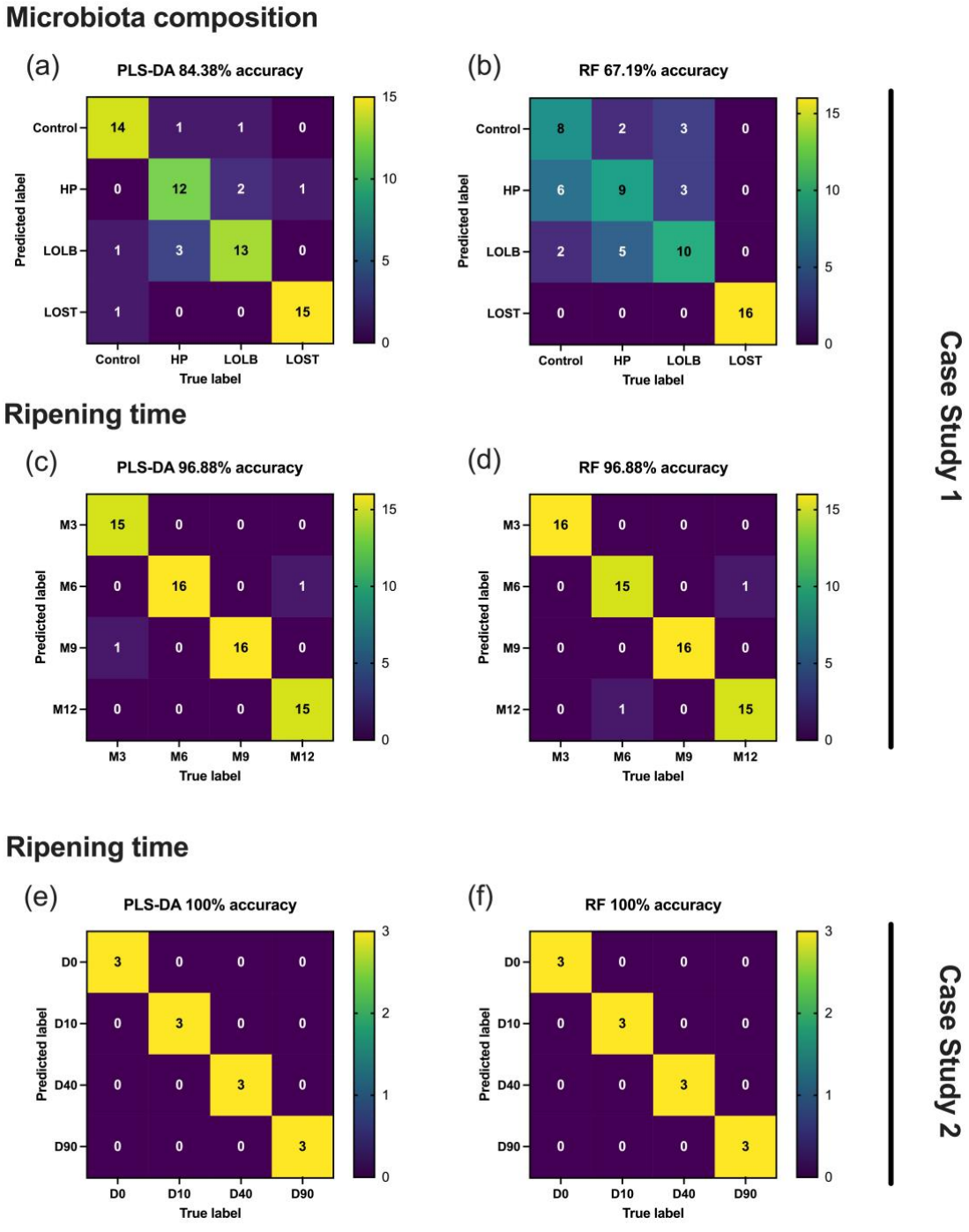


Figure 6. Classification performance of Partial Least Squares Discriminant Analysis (PLS-DA) and Random Forest (RF) in (a-d) Case Study 1 for 64 samples and (e-f) Case Study 2 for 12 samples. For Case Study 1, a 10-fold cross validation was performed, and for Case Study 2 a leave-one-out cross validation was chosen due to limited samples.

3.6.Practical application of the proposed ML-based framework

The two case studies discussed in this work are representative for typical flavor datasets, and specific suggestions for applying the proposed framework in real-world scenarios are given below. The first step is to examine the elements and figure out whether the dataset is considered large or small. Next, proper preprocessing should be implemented such as normalization to balance bias and variance and ensure a robust comparison. For small sample sizes, special attention should be paid to overfitting. The following solutions can be considered: (i) feature selection, removal of irrelevant and collinear variables (Hawkins, 2004), and (ii) regularization, such as sparse PCA and sparse PLS-DA (Bujak et al., 2016; Ruiz-Perez et al., 2020). Besides, cross validation should be employed to improve the robustness of the model. Both *k*-fold and leave-one-out are popular cross validation techniques (Wong, 2015). In most practical applications, 10-fold cross validation provides a good balance between bias and variance and is the preferred choice unless the dataset is small, where leave-one-out cross validation should be adopted to obtain a reliable accuracy (Cawley & Talbot, 2010). Regarding feature importance analysis, PLS-DA is recommended for hypothesis-driven studies using the prior knowledge, whereas RF is considered for data-driven discovery. For the practical applications, a hybrid approach considering both VIP scores from PLS-DA and MDA value from RF is suggested to improve the reliability of marker volatiles selection.

Post print version of paper published in Food Chemsitry, 2025, Volume 483, Article 144000.

The content is identical to the published paper, but without the final typesetting by the publisher.

4. Conclusions

A framework based on ML is proposed in this study to interpret flavor data by using PCA, RDA, PLS, and RF. PCA should be first employed for data exploration. Next, RDA is implemented to quantify the statistical significance of factors on flavor profiles based on permutation tests. Both the VIP score in PLS-DA models and the MDA value in RF models are useful to rank feature importance. The combination of these two techniques offers a comprehensive understanding of marker compounds. Regarding classification performance, both PLS-DA and RF can achieve high accuracy when volatile compounds are highly correlated with class labels. RF is more suitable for recognizing complex patterns when sufficient data are available. Overfitting is a risk when dealing with datasets with small sample sizes. Specific suggestions are given for applying the proposed framework in real-world scenarios. Although the flavor data used in this study to demonstrate those data mining techniques was obtained from GC-MS, the proposed flavor interpretation framework can be applied to other flavor data sources to demystify food flavor, such as GC-O-MS, GC-IMS, and GC×GC-TOF-MS.

Journal homepage: http://www.sciencedirect.com/journal/food-chemistry

Original file available at: https://www.sciencedirect.com/science/article/pii/S0308814625012518

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation program “E-MUSE” under the Marie Skłodowska-Curie grant agreement no. 956126. Simen Akkermans was funded by the Research Foundation Flanders (FWO) under grant 1224623N. Special thanks to Ms. Yadong Wang and Prof. Bei Wang for providing the original flavor data for Case Study 2.

Code availability statement

The R code used to reproduce the results of this study is available at:

https://gitlab.kuleuven.be/u0153264/demystifying-food-flavor

References

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In M. W. Berry, A. Mohamed, & B. W. Yap (Eds.), *Supervised and Unsupervised Learning for Data Science* (pp. 3–21). Springer International Publishing. https://doi.org/10.1007/978-3-030-22475-2_1

Blanchet, F. G., Legendre, P., Bergeron, J. A. C., & He, F. (2014). Consensus RDA across dissimilarity coefficients for canonical ordination of community composition data. *Ecological Monographs*, *84*(3), 491–511. https://doi.org/10.1890/13-0648.1

Borcard, D., Gillet, F., & Legendre, P. (2018). Canonical Ordination. In D. Borcard, F. Gillet, & P. Legendre (Eds.), *Numerical Ecology with R* (pp. 203–297). Springer International Publishing. https://doi.org/10.1007/978-3-319-71404-2_6

Bujak, R., Daghir-Wojtkowiak, E., Kaliszan, R., & Markuszewski, M. J. (2016). PLS-Based and Regularization-Based Methods for the Selection of Relevant Variables in Non-targeted Metabolomics Data. *Frontiers in Molecular Biosciences*, *3*, Article 35. https://doi.org/10.3389/fmolb.2016.00035

Buvé, C., Saeys, W., Rasmussen, M. A., Neckebroeck, B., Hendrickx, M., Grauwet, T., & Van Loey, A. (2022). Application of multivariate data analysis for food quality investigations: An example-based review. *Food Research*

International, *151*, Article 110878. https://doi.org/10.1016/j.foodres.2021.110878

Cawley, G. C., & Talbot, N. L. C. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research*, *11*, 2079–2107.

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

Cozzolino, D., Power, A., & Chapman, J. (2019). Interpreting and Reporting Principal Component Analysis in Food Science Analysis and Beyond. *Food Analytical Methods*, *12*(11), 2469–2473. https://doi.org/10.1007/s12161-019-01605-5

Fan, M., Fan, Y., Rao, Z., Li, Y., Qian, H., Zhang, H., Wu, G., Qi, X., & Wang, L. (2019). Comparative investigation on metabolite changes in '*wu mi*' production by *Vaccinium bracteatum* Thunb. Leaves based on multivariate data analysis using UPLC–QToF–MS. *Food Chemistry*, *286*, 146–153. https://doi.org/10.1016/j.foodchem.2019.01.144

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*, Article 100. https://doi.org/10.1038/s43586-022-00184-w

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta*, *879*, 10–23. https://doi.org/10.1016/j.aca.2015.02.012

Han, H., Guo, X., & Yu, H. (2016). Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 219–224. https://doi.org/10.1109/ICSESS.2016.7883053

Hawkins, D. M. (2004). The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12. https://doi.org/10.1021/ci0342472

Ji, H., Pu, D., Yan, W., Zhang, Q., Zuo, M., & Zhang, Y. (2023). Recent advances and application of machine learning in food flavor prediction and regulation. *Trends in Food Science & Technology*, *138*, 738–751. https://doi.org/10.1016/j.tifs.2023.07.012

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

Kherif, F., & Latypova, A. (2020). Principal component analysis. In A. Mechelli & S. Vieira (Eds.), *Machine Learning* (pp. 209–225). Academic Press. https://doi.org/10.1016/B978-0-12-815739-8.00012-2

Li, Q., Gu, Y., & Wang, N. (2017). Application of Random Forest Classifier by Means of a QCM-Based E-Nose in the Identification of Chinese Liquor Flavors. *IEEE Sensors Journal*, *17*(6), 1788–1794. IEEE Sensors Journal. https://doi.org/10.1109/JSEN.2017.2657653

Li, S., Li, T., Han, Y., Yan, P., Li, G., Ren, T., Yan, M., Lu, J., & Qiu, S. (2024). Machine learning discrimination and prediction of different quality grades of sauce-flavor *baijiu* based on biomarker and key flavor compounds screening.

Food Chemistry: X, *24*, Article 101877. https://doi.org/10.1016/j.fochx.2024.101877

Luo, H., Akkermans, S., Verheyen, D., Wang, J., Polanska, M., & Van Impe, J. F. M. (2024). Tuning and modeling cheese flavor. *Comprehensive Reviews in Food Science and Food Safety*, *23*(5), Article e13420. https://doi.org/10.1111/1541-4337.13420

Mavani, N. R., Ali, J. M., Othman, S., Hussain, M. A., Hashim, H., & Rahman, N. A. (2022). Application of Artificial Intelligence in Food Industry—A Guideline. *Food Engineering Reviews*, *14*, 134–175. https://doi.org/10.1007/s12393-021-09290-z

Melkonian, C., Zorrilla, F., Kjærbølling, I., Blasche, S., Machado, D., Junge, M., Sørensen, K. I., Andersen, L. T., Patil, K. R., & Zeidan, A. A. (2023). Microbial interactions shape cheese flavour formation. *Nature Communications*, *14*, Article 8348. https://doi.org/10.1038/s41467-023-41059-2

Oztuna Taner, O., & Çolak, A. B. (2024). Dairy factory milk product processing and sustainable of the shelf-life extension with artificial intelligence: A model study. *Frontiers in Sustainable Food Systems*, *8*, Article 1344370. https://doi.org/10.3389/fsufs.2024.1344370

Pérez-Enciso, M., & Tenenhaus, M. (2003). Prediction of clinical outcome with microarray data: A partial least squares discriminant analysis (PLS-DA) approach. *Human Genetics*, *112*(5), 581–592. https://doi.org/10.1007/s00439-003-0921-9

Qiu, S., Han, H., Zeng, H., & Wang, B. (2024). Machine learning based classification of yogurt aroma types with flavoromics. *Food Chemistry*, *438*, Article 138008. https://doi.org/10.1016/j.foodchem.2023.138008

Ramette, A. (2007). Multivariate analyses in microbial ecology. *FEMS Microbiology Ecology*, *62*(2), 142–160. https://doi.org/10.1111/j.1574-6941.2007.00375.x

Rossini, K., Verdun, S., Cariou, V., Qannari, E. M., & Fogliatto, F. S. (2012). *PLS* discriminant analysis applied to conventional sensory profiling data. *Food Quality and Preference*, *23*(1), 18–24. https://doi.org/10.1016/j.foodqual.2011.01.005

Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K., & Narasimhan, G. (2020). So you think you can PLS-DA? *BMC Bioinformatics*, *21*, Article 2. https://doi.org/10.1186/s12859-019-3310-7

Shi, L., Westerhuis, J. A., Rosén, J., Landberg, R., & Brunius, C. (2019). Variable selection and validation in multivariate modelling. *Bioinformatics*, *35*(6), 972–980. https://doi.org/10.1093/bioinformatics/bty710

Song, H., & Liu, J. (2018). GC-O-MS technique and its applications in food flavor analysis. *Food Research International*, *114*, 187–198. https://doi.org/10.1016/j.foodres.2018.07.037

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, Article 307. https://doi.org/10.1186/1471-2105-9-307

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, Article 25. https://doi.org/10.1186/1471-2105-8-25

679 Tufariello, M., Pati, S., D'Amico, L., Bleve, G., Losito, I., & Grieco, F. (2019).
680 Quantitative issues related to the headspace-SPME-GC/MS analysis of volatile
681 compounds in wines: The case of *Maresco* sparkling wine. *LWT*, *108*, 268–
682 276. https://doi.org/10.1016/j.lwt.2019.03.063
683 Wang, Y., Zeng, H., Qiu, S., Han, H., & Wang, B. (2024). Identification of key aroma
684 compounds and core functional microorganisms associated with aroma
685 formation for Monascus-fermented cheese. *Food Chemistry*, *434*, Article
686 137401. https://doi.org/10.1016/j.foodchem.2023.137401
687 Wei, G., Dan, M., Zhao, G., & Wang, D. (2023). Recent advances in chromatography-
688 mass spectrometry and electronic nose technology in food flavor analysis and
689 detection. *Food Chemistry*, *405*, Article 134814.
690 https://doi.org/10.1016/j.foodchem.2022.134814
691 Wold, S., Johansson, E., & Cocchi, M. (1993). PLS: partial least squares projections
692 to latent structures. In *3D QSAR in Drug Design: Theory, Methods and*
693 *Applications.* (pp. 523–550). Kluwer ESCOM Science Publisher.
694 Wong, T.-T. (2015). Performance evaluation of classification algorithms by *k*-fold
695 and leave-one-out cross validation. *Pattern Recognition*, *48*(9), 2839–2846.
696 https://doi.org/10.1016/j.patcog.2015.03.009
697 Xiang, X., Wang, K., Wang, F., Yang, Q., Huang, J., Zhou, Q., & Wang, Q. (2025).
698 Enhancing beef tallow flavor through enzymatic hydrolysis: Unveiling key
699 aroma precursors and volatile compounds using machine learning. *Food*
700 *Chemistry*, *477*, Article 143559.
701 https://doi.org/10.1016/j.foodchem.2025.143559
702 Zeng, X., Cao, R., Xi, Y., Li, X., Yu, M., Zhao, J., Cheng, J., & Li, J. (2023). Food
703 flavor analysis 4.0: A cross-domain application of machine learning. *Trends in*

Food Science & Technology, *138*, 116–125. https://doi.org/10.1016/j.tifs.2023.06.011

Zhang, B., Wang, Y., Wang, J., Zhang, Y., Wang, W., Cao, J., Kong, B., & Teng, W. (2025). The establishment of ham grade, sensory scores and key flavor substances prediction models for Jinhua ham via *E*-nose combined with machine learning. *Food Chemistry*, *472*, Article 142847. https://doi.org/10.1016/j.foodchem.2025.142847

Zhang, T., Ayed, C., Fisk, I. D., Pan, T., Wang, J., Yang, N., & Sun, Q. (2022). Evaluation of volatile metabolites as potential markers to predict naturally-aged seed vigour by coupling rapid analytical profiling techniques with chemometrics. *Food Chemistry*, *367*, Article 130760. https://doi.org/10.1016/j.foodchem.2021.130760

Zhao, S., Niu, C., Yang, X., Xu, X., Zheng, F., Liu, C., Wang, J., & Li, Q. (2022). Roles of sunlight exposure on chemosensory characteristic of broad bean paste by untargeted profiling of volatile flavors and multivariate statistical analysis. *Food Chemistry*, *381*, Article 132115. https://doi.org/10.1016/j.foodchem.2022.132115

Zuur, A. F., Ieno, E. N., & Smith, G. M. (Eds.). (2007). Principal component analysis and redundancy analysis. In *Analysing Ecological Data* (pp. 193–224). Springer. https://doi.org/10.1007/978-0-387-45972-1_12

Figure captions

Figure 1. Framework for flavor data interpretation in this study, including Principal Component Analysis (PCA), Redundancy Analysis (RDA), Partial Least Squares (PLS), and Random Forest (RF).

Figure 2. PCA biplot representing changes in volatile compounds of Cheddar cheese samples in a year-long ripening (a) and scree plot showing the cumulative variance explained by the first ten principal components (b). Samples are displayed by points with different colors according to the microbiota composition: lime green (●) for commercial starter culture, dark blue (●) for starter culture prepared under non-industrial conditions by hand, orange (●) for starter culture prepared by leaving out *Streptococcus thermophilus*, and brown (●) for starter culture prepared by leaving out the *Lactococcus* blend. Volatile compounds are displayed by gray arrows and the top 6 are labeled. To enrich the interpretation, ripening time is added as a supplementary variable, as indicated by the blue arrow.

Figure 3. RDA triplots representing the effect of microbiota composition (explanatory variable) on Cheddar cheese flavor profiles during a year-long ripening. Samples are colored according to the microbiota composition: lime green (●) for commercial starter culture (control), dark blue (●) for starter culture prepared under non-industrial conditions by hand (HP), orange (●) for starter culture prepared by leaving out *Streptococcus thermophilus* (LOST), and brown (●) for starter culture prepared by

leaving out the *Lactococcus* blend (LOLB). Volatile compounds (response variables) are represented by gray arrows. To make the plots easy to read, only the top 5 volatile compounds are labeled. Ripening time is added as a supplementary variable to enrich interpretation, as indicated by the blue arrows. RDA is performed under different combinations: (a) all conditions (control, HP, LOST, and LOLB), (b) control and HP, (c) HP and LOST, (d) HP and LOLB, and (e) LOST and LOLB. The p-value of the permutation test is shown in the upper left corner of the triplot, with a value below 0.05 indicating significant differences.

Figure 4. The screening of marker compounds with respect to (a) ripening time and (b) microbiota composition. Volatile compounds were ranked based on the variable importance in projection (VIP) score of the Partial Least Squares Discriminant Analysis (PLS-DA) model and the mean decrease accuracy (MDA) value of the Random Forest (RF) model. Variables with VIP scores above 1 are considered important. The results are presented as normalized [0,1] VIP scores and MDA values.

Figure 5. Flavor interpretation in a small dataset, using (a) PCA for data exploration, (b) RDA for factor evaluation, and (c) PLS-DA and RF for feature importance analysis. In (a) and (b), samples are represented by symbols, response variables are represented by gray arrows, the supplementary variable in PCA is displayed by a blue arrow, the explanatory variable in RDA is displayed by a red arrow, and the ripening time in days is indicated by the number that initiates the name of each sample. In (c), variables with variation in projection (VIP) scores above 1 are considered important.

771 Figure 6. Classification performance of Partial Least Squares Discriminant Analysis
772 (PLS-DA) and Random Forest (RF) in (a-d) Case Study 1 for 64 samples and (e-f) Case
773 Study 2 for 12 samples. For Case Study 1, a 10-fold cross validation was performed,
774 and for Case Study 2 a leave-one-out cross validation was chosen due to limited samples.