Trends in Food Science & Technology 147 (2024) 104415

Contents lists available at ScienceDirect

Trends in Food Science & Technology

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

Industrial food quality and consumer choice: Artificial intelligence-based tools in the chemistry of sensory notes in comfort foods (coffee, cocoa and tea)

Eloisa Bagnulo [1], Giulia Strocchi [1], Carlo Bicchi, Erica Liberto [*]

Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Italy

ARTICLE INFO

Handling Editor: Prof. F. Toldra

Keywords:
Chemometrics
Coffee
Cocoa
Tea
Artificial intelligence
Quality
Sensory notes

ABSTRACT

Background: Food quality is a multifaceted, evolving concept encompassing various aspects throughout the production chain. The shift from traditional analytics to comprehensive strategies is driven by the need to meet this extended quality definition.

Scope and approach: Foodomics, specifically focusing on connecting chemical composition to sensory properties, is vital for comfort foods like coffee, cocoa, and tea, chosen for enjoyment rather than nutrition. In foodomics, larger and more complex datasets demand Artificial intelligence-based tools for decoding encrypted information.

Key findings and conclusions: Global coffee, cocoa, and tea supply involve numerous small farms affected by socio-political instability and climate change. Financial motives drive fraudulent practices, leading to unfair competition, loss of consumer confidence, and safety issues. AI-based tools enhance data understanding for knowledge gain, but challenges include the misalignment between academia and industry, limited industrial samples for AI application, academic training gaps, algorithm complexity, and decision-making misinterpretation.

1. Introduction

Foodomics, an interdisciplinary field, uses various techniques to study food at a molecular level, analyzing composition, safety, quality, flavour, and their impact on human health (Cifuentes, 2017). Global food markets, increased demand for safe, minimally processed, and healthy food require new approaches to identify quality markers. Food fraud poses a major problem due to costs, public health implications, intentional or unintentional acts, affecting regulators and industry. Food quality is crucial for the food industry's success, influenced by contextual and situational factors, evolving over time, covering the entire production chain (total quality) (Fig. 1) (Bhatia & Ahanger, 2021). The concept of Food Quality, introduced in 2004 by the FAO and updated to ISO 9001:2015, ensures expected properties. ISO 22000:2018 extends the concept, integrating food quality and safety management systems (ISO 9001:2015; ISO 22000:2018). Food quality dimensions depend on consumer expectations and industry standards. Industry seeks an objective, precisely defined standard, while consumer quality is subjective, influenced by expectations, perception, and acceptance based on the moment or situation.

1.1. Food quality and its impact on industry production

Ensuring high-quality products helps food companies build a strong reputation, attract new customers, and stand out from competitors (Chen, Zhang, & Luo, 2021).

Food quality is crucial for safety and compliance; failure to meet standards can lead to recalls, fines, legal issues, and reputation damage (Bhatia & Ahanger, 2021). Optimizing supply chains and maintaining food quality reduces waste, minimizes losses, and enhances operational efficiency. Premium pricing is achievable for high-quality, organic, natural, or specialty products. To ensure the quality and safety of products, food companies have implemented a set of practices, procedures and guidelines called Food. To ensure product quality and safety, food companies implement Food Quality Management Systems (QMS) (ISO 22000:2018). A robust Food QMS builds trust, lowers recall risks, and ensures safety and quality for manufacturers, processors, distributors, and retailers.

* Corresponding author.
E-mail address: erica.liberto@unito.it (E. Liberto).

[1] Eloisa Bagnulo and Giulia Strocchi have equally contributed to the work.

https://doi.org/10.1016/j.tifs.2024.104415
Received 22 December 2023; Received in revised form 14 February 2024; Accepted 3 March 2024
Available online 9 March 2024
0924-2244/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Fig. 1. The different aspects of the food total quality concept.

1.2. Food quality from the consumer view

Consumer-perceived food quality is vital, directly impacting satisfaction and the overall dining experience. Evaluation factors include taste, appearance, freshness, nutritional value, safety, and dining ambience (Fig. 2). Preferences lean towards balanced options with essential nutrients, avoiding excessive unhealthy ingredients like trans fats, added sugars, and high sodium levels. Authentic flavours, culinary traditions, and ingredient origin contribute to perceived quality, particularly in niche or specialty dishes. Ethical considerations such as sustainability, fair trade, and animal welfare are increasingly influencing consumer choices (Mascarello, Pinto, Parise, Crovato, & Ravarotto, 2015). Consumer satisfaction aligns when expected and experienced quality match. Brands serve as quality indicators, yet economic crises affect brand choice. A recent IFIC survey highlights flavour as the top determinant of food quality, followed by price, healthiness, convenience, and sustainability (IFIC, 2017). Positive word-of-mouth enhances reputation and attracts new customers, while poor food quality leads to dissatisfaction, negative reviews, and sales decline.

1.3. Comfort food and flavour

Coffee, cocoa, and tea serve as comfort and "social" foods, contributing to human well-being (Lemarcq, 2020; Preedy, 2013, 2015). Comfort food, offering psychological comfort, adds complexity to eating behaviour, involving all sensory systems and gut-brain interactions (Spence, 2017). In recent years, a strong scientific effort has been made to better understand how all the different perceptions interact in the "building up" of a food flavour (Fig. 3). Food flavour is therefore defined as a multisensory phenomenon that results from the integration of taste, smell and other sensory (e.g. somatosensory) information into a perceived characteristic of the food (Prescott, 2015; Spence, 2016a, 2016b). It is therefore clear that the interactions between flavour components determine what we perceive in food and what we like, and that they influence consumer choices.

The quality of these foods depends on factors like botanical variety, weather, agricultural practices, storage, and processing. Higher quality products command higher prices, determined by sensory profile and nutritional properties linked to chemical composition.

This means that the production chain must be supplied with products of uniform quality standards, i.e. uniform raw materials with specific quality parameters from the country of origin. This is a serious problem for industries that process coffee, chocolate and tea manufacturers, as they need masses of raw or semi-finished products of consistent quality to meet ever-increasing demand. Standardizing quality is challenging due to crops being produced in non-EU countries by independent farmers, leading to batch heterogeneity impacted by socio-political instability and climate change (Boeckx, Bauters, & Dewettinck, 2020; Nowogrodzki, 2019). Objective and robust tools are key for industries to develop methods for recognizing and assessing food quality markers and

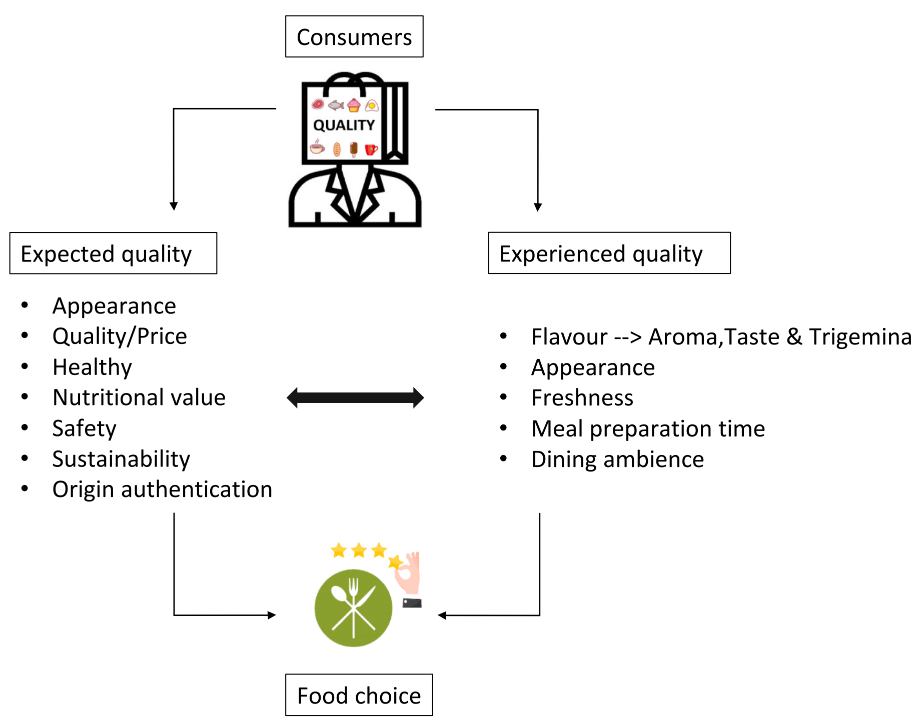

Fig. 2. Factors affecting the consumer's choice.

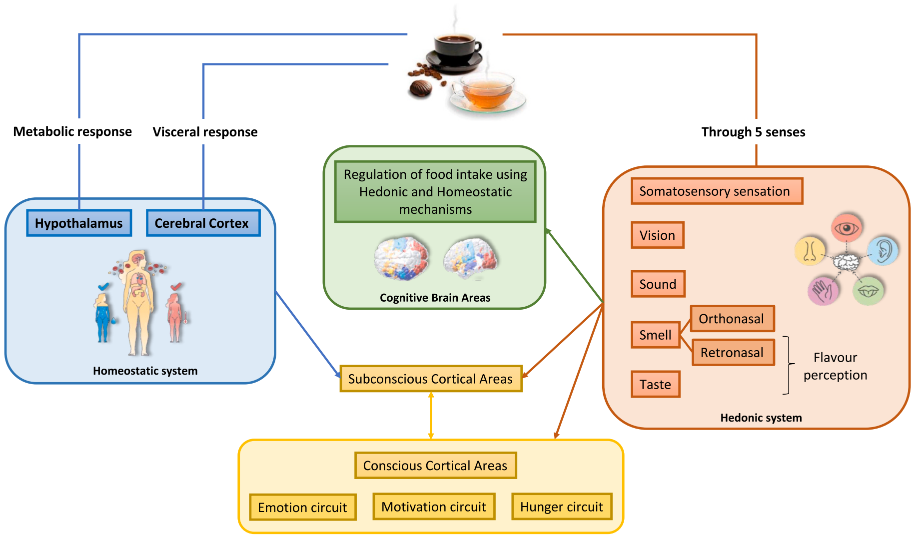

Fig. 3. Visual vision of flavour's perception multisensory phenomena.

their development is driven by the globalization of the food market, food safety, and rising consumer demand for safe, less-processed, and healthful food (Gao, Tello, & Peterson, 2023; Jiménez-Carvelo, González-Casado, Bagur-González, & Cuadros-Rodríguez, 2019; Zeng et al., 2023).

1.4. AI-based tools play a pivotal role in ensuring coffee, cocoa and tea flavour quality

Artificial intelligence comes from the Greek, where some things were able to move on their own. The modern concept aims to solve problems with the help of computers executing code that mimics the cognitive processes of the human brain. Depending on the field (i.e. information technology, business, social, chemistry) there are different terms and approaches used in data science and very similar multiple definitions of the same terms exist and are often used interchangeably, e.g. data mining, pattern recognition, machine learning, artificial intelligence. However, there is not full agreement in the scientific community on these definitions that support an open interpretation. All terms are commonly used in chemical data analysis (Ayres, 2021; Jiménez-Carvelo, 2019; Szymańska, 2018). In each case, it is the application of a mathematical process that is used to solve the problem under investigation, assuming it is needed, and in analytical chemistry it could be summarised as chemometrics (Amigo, 2021).

Chemometrics is a discipline that deals with the application of mathematical and statistical methods to chemical data. The main aim of chemometrics is to extract meaningful information from complex data sets, enabling new insights into food systems by exploring chemical data, researching trends and building predictive functional models, and drawing meaningful conclusions from experiments to make informed decisions (Fig. 4) (Andre & Soukoulis, 2020; Jiménez-Carvelo et al., 2019.

Chemometrics can be applied: a) to plan factors to optimize food processing and analysis by the design of experiments (DOE); b) to extract useful information from the experimental data (Data mining) (Fig. 4).

The industrial revolution, characterized by the integration of digital technologies, automation and data exchange, plays a crucial role in quality management and AI-based tools are an important enabler for achieving and maintaining quality standards, creating new business models, increasing plant productivity and improving product quality. This review will delve into the potential and constraints of AI to address real industrial challenges in ensuring the flavour quality of coffee, cocoa and tea. Details about the algorithms used are covered elsewhere (Ayres, Gomez, Linton, Silva, & Garcia, 2021; Casale et al., 2020; Granato et al., 2018; Jimenez-Carvelo, 2021).

2. DOE in cocoa, coffee and tea flavour research

Design of Experiments (DOE) is crucial for efficient experimentation, providing maximum relevant information at minimal cost (Casale et al., 2020). This approach considers parameter interactions, offers global knowledge, minimizes experiments for optimal information, and

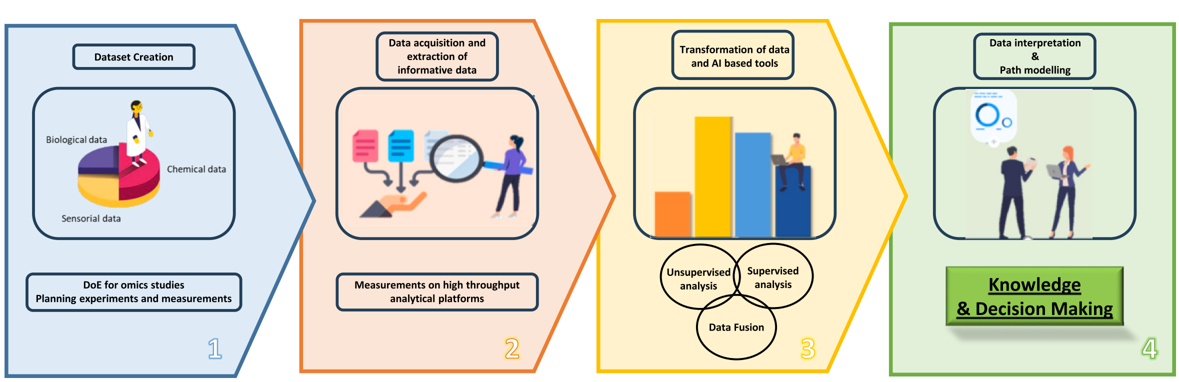

Fig. 4. The goals of chemometrics from the planning experiments to extraction of relevant information for a deep knowledge and proper decision maker at industrial level.

facilitates mathematical model construction. In flavour and sensory analysis, DOE efficiently designs experiments, establishing relationships between inputs (ingredients, properties) and output variables (sensory properties, preferences). Consequently, the optimisation of the product based on one or more product properties, such as sensory attributes or consumer preference, may be performed based on the desirability function and methods that enable the formulator (industry) to develop a food product that is optimised with respect to given properties (Andruszkiewicz, Corno, & Kuhnert, 2021; McClure, Spinka, & Grün, 2021). For example, in coffee, DOE explored brewing conditions' influences physicochemical and sensory characteristics (Frost, Ristenpart, & Guinard, 2020; Zakaria et al., 2023). McClure et al. evaluated cocoa processing using a randomized optimal experimental design, observing significant changes in important bitter compounds during roasting (i.e., theobromine, caffeine, epicatechin, catechin, procyanidin B2) from cocoa beans with eight roast profiles across three origins (McClure et al., 2021). Fermentation and treatment of raw cocoa, coffee and tea material can benefit from DOE application. Artificial systems of controlled fermentation to model the responses of bean components have recently been studied by the application of DOE (Gutiérrez-Ríos et al., 2022). John et al. applied DOE to study controlled fermentation effects on cocoa biochemistry. Response surface model was drawn up with a D-optimal design concluding that temperature, acetic acid and incubation time had the greatest effects of the factors considered observing significant interactions between factors for each response, highlighting their collective action in modulating cocoa biochemistry (John et al., 2020). Wei et al. used experimental design to reveal the mechanism responsible for the improvement of taste and colour of yellow tea (YT) optimizing yellowing. The tea leaves after picking are dried at a lower temperature for a short time, while the leaves are still warm and moist, they are then 'yellowed' by wrapping the tea. They adopted a response surface methodology (RSM) using Box-Behnken design (BBD) based on a composite central design that revealed that an increased temperature and relative humidity, a reduced time, enhanced yellowness and sweetness by over 40.5%, and improved consumer acceptability (Wei et al., 2023). Despite DOE advantages, its industrial-level application remains limited (Casale et al., 2020).

3. Data mining in coffee, cocoa and tea quality definition

Data mining is the process of discovering patterns, trends, correlations, or useful information from large datasets. It involves various techniques and methods to analyse and extract knowledge from data, often with the goal of making informed decisions, improving processes, or gaining insights into complex systems (Ji et al., 2023; Jimenez-Carvelo, 2021; Zeng et al., 2023).

Analytically, data processing is divided into profiling and fingerprinting based depending on investigative objectives. Profiling characterizes a sample's chemical composition through detailed analysis, often coupling chromatography with mass spectrometry for unique identification. Fingerprinting treats the entire signal output as a sample's unique fingerprint for rapid comparative analysis, without isolating information from the bulk data matrix. If signal outputs are from multidimensional platforms like GC-MS, GC $\times$ GC-MS, LC-MS, LCxLC, e-Nose, or PTR-MS, information is latent and can be targeted in successive steps. (Cuadros-Rodríguez, Ortega-Gavilán, Martín-Torres, Arroyo-Cerezo, & Jiménez-Carvelo, 2021; Cuadros-Rodríguez, Ortega-Gavilán, Martín-Torres, Medina-Rodríguez, et al., 2021; Cuadros-Rodríguez, Ruiz-Samblás, Valverde-Som, Pérez-Castaño, & González Casado, 2016). Despite increasing fingerprinting use, analytical challenges may compromise data quality. Insufficient chromatographic resolution may arise from complex samples or the preference for quicker runs. Issues like baseline drift, peak shape changes, retention time shifts, and co-elution of compounds with similar mass spectra are common in complex food analysis. These problems can irreversibly affect data quality, and not all software can address them comprehensively. Pre-processing techniques correct analytical issues, like baseline/background contributions or retention time shifts, before chemometric treatment and model building (Jiménez-Carvelo et al., 2019). Tools such as Multi Curve Resolution-Alternative Least Square (MCR-ALS) were adopted to resolve co-elution (overlapped peaks). MCR-ALS, like parallel factor analysis (PARAFAC) and other multi-way approaches (Tucker 3 or *N*-PLS, *N*-SIMCA, etc.), affords a computational gain in chemical information relating to specific analytes, for both targeted and untargeted approaches (Casale et al., 2020).

Profiling and fingerprinting are widely used in sensomics and flavouromics, "omics" disciplines (or sub-disciplines) that refer to the scientific study of flavour at a molecular level (Charve, Chen, Hegeman, & Reineccius, 2011; Dunkel et al., 2014). They combine principles from various fields, including chemistry, biology and sensory science, to explore and quantify the compounds that contribute to the taste and smell of foods and beverages and require an integrationist's mindset that adopts more comprehensive strategies capable of comprehensively mapping the Chemical Flavour Code (Bressanello et al., 2021; Cordero, Kiefl, Reichenbach, & Bicchi, 2019; Peterson, 2008). In addition, machine learning applied to fingerprinting and/or profiling technologies has revealed strong relationships and correlations between aroma and taste. Several research papers have demonstrated the potential of using fingerprint workflows and concepts to identify features with high correlation to a biological outcome. Flavouromic workflows based on fingerprints have been applied to discover new components responsible to enhance or suppress taste/smell attributes of foods (Gao, Tello, & Peterson, 2021; Lin, Tello, Simons, & Peterson, 2022; Ronningen, Miller, Xia, & Peterson, 2018; Sittipod, Schwartz, Paravisini, & Peterson, 2019; Sittipod, Schwartz, Paravisini, Tello, & Peterson, 2020).

Different authors confirmed that odourants and non-odourants interact in the expression of a perceived sensory attribute (Dunkel et al., 2014; Guichard, Barba, Thomas-Danguin, & Tromelin, 2019; Wang, Chambers, & Kan, 2018). Guichard et al. also showed that odourants that promote targeted taste perception could be used to modulate the overall taste profile in foods and beverages. The investigation on odour and taste networks using cheminformatics in commercial multi-fruit juices showed strong associations of the network visualization between odour (green, grassy and herbal) and taste (bitterness) descriptors (Barba, Beno, Guichard, & Thomas-Danguin, 2018; Guichard et al., 2019). Very recently, Nicolotti et al. introduced the concept of an "artificial intelligence olfactory machine" through SEBES (Sensomics-Based Expert System) an analytical approach that attempts to simulate the human sense of smell by defining the patterns of key odourants responsible for food aroma (Nicolotti, Mall, & Schieberle, 2019; Squara et al., 2023).

AI algorithms used are unsupervised or supervised learning. Unsupervised learnings are devoted to exploring data to find trends and patterns, most typically using principal component analysis (PCA) or clustering algorithms which can perform dimensionality reduction and clustering. Supervised Learning belongs to a category of classifier and predictive algorithms where the system learns to make predictions from a labelled dataset (Ayres et al., 2021; Casale et al., 2020). Among supervised pattern recognition methods, Partial Least Square Discriminant Analysis and Regression (PLS-DA, PLSR) are the most popular, even if Support Vector Machine (SVM), Random Forest (RF) and Artificial Neural Networking (ANN) are also used in cocoa, coffee and tea field with the target listed below (Fig. 5), while Fig. 6 summarizes some advantages and disadvantages of the AI tools (Bressanello et al., 2018, 2021; Gao et al., 2021; Perotti et al., 2020; Strocchi, Bagnulo, et al., 2022; Yu, Low, & Zhou, 2018).

✓ Authentication & traceability;
✓ Adulteration;
✓ Compositional properties and quality;
✓ Processing and quality control;
✓ Prediction of food properties.

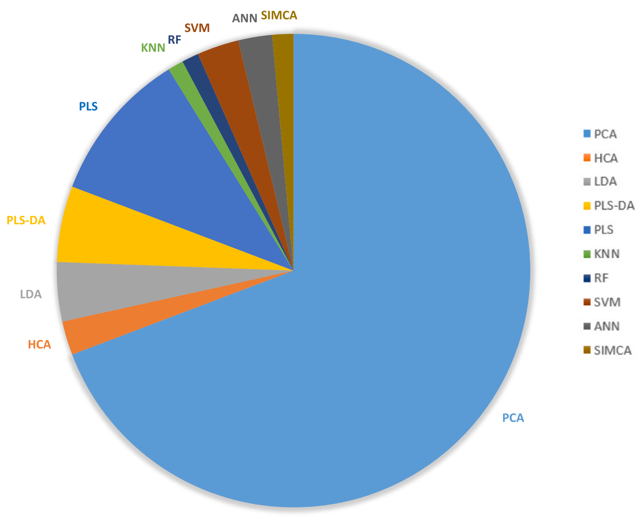

Fig. 5. Overview of the data analysis tools most used in coffee, cocoa and tea flavour studies from 2013 to 2023.

Hence, the decision between profiling or fingerprinting approaches, along with the choice of particular AI tools, depends on the research objectives. Profiling methods offer accuracy and specificity, whereas fingerprinting approaches offer flexibility and the potential to identify new target markers.

3.1. Authentication & traceability: origins and species

Cocoa, coffee, and tea authenticity is crucial due to the growing market for high-quality products. Different geographical origins result in significant variations in quality, taste, and commercial value. Ensuring traceability is essential for authenticity, benefiting import-export trade and global consumers (Li et al., 2021). The marketing of these comfort foods relies on distinct flavours from various origins, highlighted on packaging to attract customers. Increasingly, consumers, especially with mono-origin products, prioritize sustainability and the fair trade market (Bilfield, 2022). In food authentication, fingerprinting and profiling strategies are increasingly used and accepted to monitor the integrity of food products (Cuadros-Rodríguez, et al., 2021). Indeed, an adequate number of pure and authentic samples are required to create a representative database of the "real" food population to determine the degree of similarity of the fingerprints of diagnostic chemical characteristics of an unknown sample compared to a representative reference sample (Cuadros-Rodríguez, Ruiz-Samblás, Valverde-Som, Pérez-Castaño, & Casado, 2016). Furthermore, the identification of flavour quality requires analytical methods capable of providing detailed diagnostic profiles that correlate with the sensory characteristics that can be monitored and quantified for objective evaluation in quality control (QC) (Bressanello et al., 2021). Hyphenated chromatographic platforms, coupled with artificial intelligence, are essential for decoding complex datasets, uncovering trends, and assessing associations in food chemical composition (Badmos, Fu, Granato, & Kuhnert, 2020; Bressanello et al., 2018; Jimenez-Carvelo, 2021; Quelal-Vásconez, Lerma-García, Pérez-Esteve, Talens, & Barat, 2020).

The interpretation of the data obtained can be carried out through the use of AI tools to identify patterns and determine key chemical fingerprints or chemical identity cards of the product's origins. When applying different sampling methods to describe in-cup sensory properties related to the origins and species, PCA was used by Bressanello et al. to explore the chemical information on coffee aroma and flavour derived from HS-SPME of ground coffee and SBSE/SPME sampling in solution in combination with GC–MS to assess their compatibility with cupping evaluation for quality control. In this application, they were able to distinguish the species and, within them, the origin. Data processing showed that, despite their differences, the three methods provide the same type of chemical information to distinguish the samples and that they can be used interchangeably to recognize the chemical-sensory identity card of origin (Bressanello et al., 2017).

Volatilome analysis on cocoa, coupled with PCA, reveals that geographical origin (Africa, America, and Southeast Asia) has a more significant impact on product differentiation than industrial processes. This is particularly important for the chocolate industry, where beans and liquors can enter different processing steps (Marseglia, Musci, Rinaldi, Palla, & Caligiani, 2020). This concept is also described in studies by Acierno and Bagnulo (Acierno, Alewijn, Zomer, & van Ruth, 2018; Bagnulo et al., 2023). Bagnulo et al. applied HS-SPME-GC-MS in combination with PCA and PLS-DA in the origin identitation of 160 cocoa bean and liqueur samples by using fingerprinting and profiling strategies that enabled origin differentiation with efficient classification models (Bagnulo et al., 2023). In the study of Acierno et al. Flow Infusion-Electrospray Ionization- Mass Spectrometry (FI-ESI-MS) was tested to assess the geographical origin of fifty-seven dark chocolates, categorized by bean origin (Africa-15, Asia-11, South America-31) (Acierno et al., 2018). PCA discriminated between African and Asian bean-to-bar chocolates but did not show a clear trend for South American chocolates. kNN results enabled to classify African and Asian

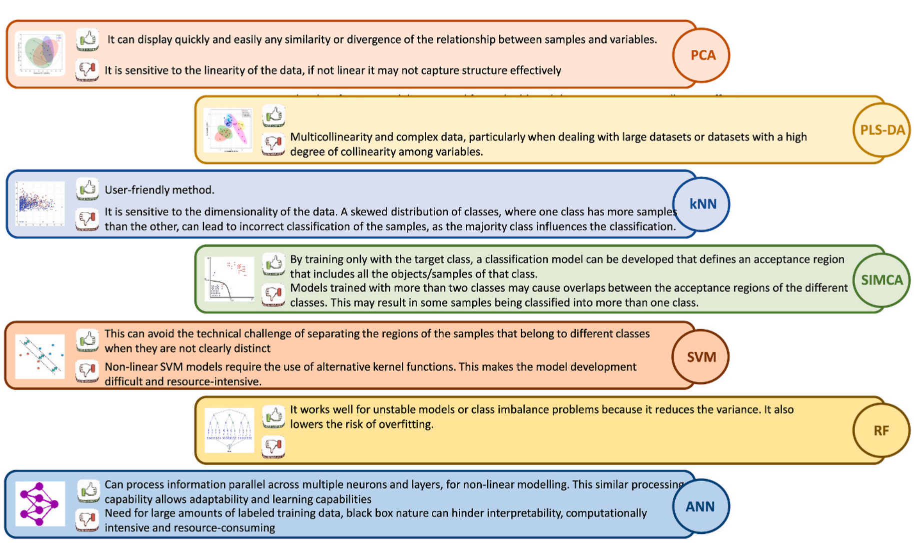

Fig. 6. Advantages and disadvantages of the most used data mining tools.

chocolates *versus* South America. Authors attributed the inability to simultaneously separate the three continents and the weak differentiation of South American samples to brand-related factors like recipe and industrial processing. However, factors like season, experimental conditions, batch, and instrument effects in a large dataset over several years pose challenges, hampering origin classification. Kumar et al. used a four-year dataset with 297 L C-MS fingerprints of cocoa from 10 countries to investigate these challenges using PCA and LDA. PCA provided limited origin separation, while LDA showed a strong non-linear dependence on compound quantity. However, they achieved origin identitation and avoid overfitting by optimizing multivariate analysis. LDA classification was non-linearly influenced by the components number. A compound selection criterion based on the Gaussian distribution of intensities across samples was employed to reach an enhancement of the origin clustering of samples (Kumar et al., 2021).

Food authentication and food fraud detection are two sides of the same coin. Fraudulent modification of food quality entails marketing products with a composition different from label indications or adding substances to alter specific features for financial gain. Food origin falsification is often employed to enhance market demand (Momtaz, Bubli, & Khan, 2023). In this context, supervised classification algorithms are extensively used. Class modelling, exemplified by techniques like Soft Independent Modelling of Class Analogies (SIMCA), focuses on a single category, often the main category of interest, employing a "one-class classification" approach. This method assesses if a sample matches specific characteristics of the target class; if not, it's excluded and not classified. This approach yields reliable results, crucial for meeting quality and regulatory standards for sample authentication (Jimenez-Carvelo et al., 2021; Rodionova, Titova, & Pomerantsev, 2016). In contrast, classification techniques, such as PLS-DA, KNN, and LDA, determine the likely class affiliation of a sample from predefined classes, assigning each sample to its most probable category, even for objects not associated with the analysed classes (Fig. 6) (Casale et al., 2020, Jimenez-Carvelo et al., 2021). The performance of the model hinges primarily on several factors, including the output type (whether it is classification or regression), sample size, the pre-processing technique used, and the specific algorithms employed. Another group of supervised classification models adopts experience-driven methodologies. These models utilize iterative classification techniques aimed at reducing errors through specific training sets utilized in constructing SVM, RF, and ANN models (Fig. 6). There is not a universally 'best' classification method; each task requires the most relevant and suitable AI tools to address the specific question posed. Protected Designation of Origin (PDO) and protected geographical indication (PGI) products are of higher quality compared to non-PDO ones and, thereby, they have higher prices (Bilfield, 2022; Nowogrodzki, 2019). There are various techniques for authentication and traceability, with methods based on elemental composition and isotope ratio analysis being the most commonly used. However, these analytical methods cannot reflect the changes in flavour quality from different regions (Shuai et al., 2022; Silva Fernandes, de Sousa Fernandes, Pistonesi, & Gonçalves Dias Diniz, 2023). For this reason, ever more studies have investigated metabolites based on LC-MS/GC-MS fingerprints as indicators to identify the geographical origin. Gu and co-workers have traced the origins of Chinese green tea using a two-dimensional (2D) LC-DAD fingerprints coupled with chemometrics (Gu et al., 2022). A total number of 62 chemical components were extracted from raw fingerprints of 78 tea samples from two Chinese regions (Zhejiang (ZJ) and Shandong (SD)) by multivariate curve resolution-alternating least squares (MCR-ALS) algorithm previously pre-processed by baseline corrections. Orthogonal partial least squares-discriminant analysis (OPLS-DA) performed on the areas from MCR-ALS profiles of a training set of 80% of total samples and cross-validated using two types of scaled data showed a predictive ability $Q^2_{cum} > 0.75$. The application on a test set of 20 % of samples afforded a total recognition rate of 92%. A comparative analysis of different machine learning tools including convolutional neural network (CNN), LDA, and SVM was used to choose the best model using data from fast and non-destructive terahertz (THz) spectroscopy for coffee bean origin classification (Yang et al., 2021). Due to the data dimensionality, PCA and genetic algorithm (GA) were applied to reduce features for LDA and SVM improving prediction accuracy respectively: SVM 50%, PCA-SVM:65% and GA-SVM:75%, the best classification was achieved by CNN with a 90% accuracy in the prediction set.

Many studies were focused on developing methods to detect, identify and quantify coffee substitutes, including grains, nuts and legumes, using metabolomics methods. For instance, Fourier-transform mid infrared (FT-MIR), combined with machine learning, was used to identify and quantify adulterants in coffee (Flores-valdez, Gabriela, Osorio-revilla, & Gallardo-vel, 2020). SIMCA modelling was applied to classify different types of adulteration (coffee-: coffee husks, corn, barley, soy, oat, rice). The model optimisation involved spectral pre-treatments: normalization (multiplicative scatter correction, MSC), Savitzky–Golay filter (5 points for smoothing) and baseline correction (offset). To quantify the adulterant percentage, quantitative models were developed using algorithms such as One-variable (PLS1), Multiple variable Partial Least Squares (PLS2) algorithms and Principal Component Regression (PCR). PLS1 exhibited a superior predictive model, although the standard error in prediction varied across adulterant types (SEP: barley > soy > oat > coffee-husks > corn > rice). However, the authors incorrectly claim SIMCA modelling as an AI-discriminant tool.

Products derived from cocoa are also exposed to different adulterations related to insufficient production and increasing market demand. The sensory quality of cocoa products may be impacted by shell content in cocoa powders because of economically motivated adulteration, potential cross-contamination or machinery failures during the peeling process. FTIR (Fourier transform infrared) or NIR (near-infrared spectrometry) fingerprints were widely used to reveal adulteration. Oliveira et al. predicted the content of cocoa shells in cocoa powder by AI methods using NIR. They employed PLSR and different spectrum pre-processing techniques, finding reflectance smoothing with standard normal variate as the most effective. The Ensemble Monte Carlo Variable Selection (EMCVS) was used to select the most informative wavelengths, making NIR spectroscopy reliable for cocoa fraud detection (Oliveira, Badaró, Esquerre, Kamruzzaman, & Barbin, 2023). In any case, these methods are only classificative because they do not enable to a chemical speciation linked to the origins.

Despite the large number of modern analytical tools available to detect food authenticity and fraudulent behaviour (i.e. adulteration of raw materials or finished products, fraudulent declaration of geographical origin), the industry and official controls need fast, more meaningful and economical methods to quantify quality characteristics/attributes, the application of which leads to very complex and large data sets that are not easy to interpret. However, the algorithms used need to be validated on representative samples in terms of number and diagnostics to have robust mathematical models that can be useful as AI decision tools. Scientific papers do not always report the correct approach or concept, very often the exploratory analysis (e.g. PCA) is considered as a classification method and/or there is an unclear distinction between discriminative and modelling tools, and very often supervised learning ends up with internal cross-validation without testing the model on an external sample set (Amigo, 2021; Casale et al., 2020; Jimenez-Carvelo et al., 2021; Rodionova et al., 2016). Furthermore, academic research and industry are not always aligned because, on the one hand, industry has difficulties implementing new methods in processing and, on the other hand, researchers often develop methods that do not always meet the requirements of user-friendliness and cost management.

3.2. Post-harvest processing: fermentation, drying and roasting

Post-harvest treatments have a decisive influence on the final product's quality. Several diverse metabolic processes occur during post-harvest processing, altering the chemical composition. Fermentation, drying and roasting are the most important processing steps for the quality of these hedonic foods. The first two steps are usually carried out in the country of origin and play a decisive role in the flavour profile of the leaves and beans, while roasting generally takes place directly in the industrial plants in the exported countries (Elhalis, Cox, & Zhao, 2023; Febrianto & Zhu, 2023; Herrera-Rocha, Fernández-Niño, Cala, Duitama, & Barrios, 2023).

3.2.1. Fermentation

Fermentation remains a spontaneous and non-standardized process impacting on cocoa, tea and coffee flavour, which is mainly linked to the activity of environmental microorganisms. Heterogeneous and dynamic bacterial and yeast communities are responsible for the biochemistry associated with precursors of flavour (Fu et al., 2024; Herrera-Rocha et al., 2023). During fermentation, endogenous proteases break down the proteins into amino acids and short-chain oligopeptides, while invertase breaks down the polysaccharides into glucose and fructose, thus forming important precursors that are required for the subsequent chemical reactions in the following steps of the process. In addition, polyphenol oxidase reduces the content of polyphenols through oxidation, which affects the flavour of the products by reducing astringency and bitterness (Herrera-Rocha et al., 2023; John et al., 2020; Megias-Perez, Moreno-Zambrano, Behrends, Corno, & Kuhnert, 2020).

Traditional Chinese tea is classified into six categories based on the processing and fermentation degree (ISO 20715:2023). Fused analytical data of the total amount of catechins, total polyphenols, theanine, free amino acids and caffeine in tea leaves treated by PCA have shown how these components distinguished the different processing, with the total amount of polyphenols, catechins and theanine decreasing as the degree of fermentation increased (Fu et al., 2024). The aroma quality of pile-fermented variety Yinghong No. 9 dark tea was studied by combining GC $\times$ GC-QTOFMS, electronic nose (E-nose) and GC-olfactometry (GC-O) and compared to the sun dried tea. Pile fermentation, a common method for dark tea, alters organic compounds through heat, microorganisms, and enzymes. PCA and Venn diagram analysis identified volatile metabolite differences between processing methods. OPLS-DA and Vulcano plot were then used to analyse aroma component differences, revealing 32 differential volatile compounds with odour activity values (OAVs) > 1. Pile fermentation reduced pungency and enhanced woody, stale, and sweet aromas (Wen et al., 2023). Automated machine learning and computer vision-based applications have been proposed to grade tea fermentation, using an image database ($n = 6000$) with recorded temperatures, humidity and time. Features were selected by PCA and used in kNN, SRC (Sparse Representation Classifier), and SVM models whose performances have been validated by the k (10) fold cross-validation. Accuracy in tea fermentation level detection of the proposed algorithm achieved 87.39% (k-NN), 89.72% (SRC), and 98.75% (SVM) (Bhargava, Bansal, Goyal, & Shukla, 2023).

Several research groups studied the effects of fermentation on cocoa in detail (Balcázar-Zumaeta, Castro-Alayo, Cayo-Colca, Idrogo-Vásquez, & Muñoz-Astecker, 2023; Megias-Perez et al., 2020). This spontaneous and natural process, which involves the succession of yeasts, LAB and acetic acid bacteria (AAB), is responsible for the metabolic “metamorphosis” in the beans, which is essential for the formation of the basic flavour and aroma precursors (such as alcohols, aldehydes, organic acids and esters) for high-quality chocolate (De Vuyst & Leroy, 2020). Despite the studies, the process is complex and little is known in particular about how external factors and their interactions (i.e. temperature, fermentation time, pH, microbial influence) affect the formation of metabolites in the beans, as most studies are based on spontaneous fermentations, over which they have almost no control. Temperature, acetic acid, and incubation time don’t always exhibit linear effects on pH, peptide diversity, and flavanol content during controlled fermentation, as well as on the volatile profiles (Cevallos-Cevallos, Gysel, Maridueña-Zavala, & Molina-Miranda, 2018; John et al., 2020). AI approaches showed a change in total polyphenol content and an increase of epicatechin in fermented beans, an improvement of low molecular weight carbohydrates, and a degradation of proteins to oligopeptides after 72 h of spontaneous fermentation (Kumari et al., 2018; Megias-Perezet al., 2020). The low astringency and bitterness due to the reduction of phenols and flavan-3-ols, and the epicatechin/catechin, and fructose/glucose ratios contribute to the classification of beans according to cocoa processing; moreover, the level of fermentation of peptides determine if a bean is under-fermented, adequately fermented, or over-fermented (Balcázar-Zumaeta et al., 2023). Time-related changes in volatile profiles of bulk and fine-flavour cocoa during fermentation were studied by Cevallos-Cevallos et al. Employing cluster analysis and PCA, they examined volatiles from Criollo, Forastero, and Nacional cocoa varieties. PCA illustrated improved clustering of the three cultivars after fermentation completion (5 days), displaying distinct volatile patterns. Criollo cocoa exhibited floral, fruity, and woody aromas, with characteristic volatiles like linalool, epoxylinalool, and benzene-ethanol. Nacional cocoa displayed fruity, green, and woody aromas, featuring volatiles such as 2-nonanone and valencene. Forastero cocoa, representing bulk cacao, released floral and sweet aroma volatiles such as epoxylinalool and pentanoic acid (Cevallos-Cevallos, 2018). After harvesting, coffee beans undergo a post-harvest process to make them more stable, versatile and roastable. The green coffee beans are processed using one of three techniques: dry, wet or semi-dry, with the aim of removing the pulp of the cherry. The quality of coffee brewing is decisively influenced by the microbiota during fermentation and various studies are dedicated to improving the quality or exploring new flavours in the cup (Elhalis et al., 2023; Siridevi, Havare, Basavaraj, & Murthy, 2019). In particular, this process is applied to Robusta or to other coffee species/hybrids, that are less sensitive and/or high yielding, to obtain a flavour quality closer to Arabica (Afriliana et al., 2019; Febrianto & Zhu, 2023; Ribeiro et al., 2017; Saunshia, Sandhya, Lingamallu, Padela, & Murthy, 2018). In natural fermentation, yeasts are ubiquitous microbiota essential to initiate coffee fermentation preventing oxygenic filamentous fungal growth, degrading mucilage through the secretion of pectinolytic enzymes which release flavour precursors. Aswathi et al. investigated microbial metabolites and the quality of greenery and roast resulting from the fermentation of pulped natural/honey coffee blend (HC) and washed coffee (WC) of *Coffea canephora* (Robusta), using 1H NMR, GC-MS and sensory investigations. PCA revealed the impact of fermentation on carbohydrates, organic acids, and secondary/specialized metabolites, despite similarities in chemical-physical traits between HC and WC beans. HC processing yielded increased volatile levels, enriching flavours with sweetness, tea rose, and chocolate notes (Aswathi, Shirke, Praveen, Chaudhari, & Murthy, 2023).

In recent decades, a great deal of research has been carried out to investigate the microbial ecology of cocoa, coffee and tea fermentation and its effects on flavour formation. However, climate change affects the dynamism of the yeast and microbial community, which can play a positive/negative role in the flavour complexity of these foods. On the one hand, yeast and microorganisms can improve the flavour or be responsible for new sensory characteristics that can be industrially exploited to create new products/blends, but, on the other hand, the changes in microbial ecology can favour the formation of off-flavours. In any case, this makes increasingly difficult to obtain standardized products from batch to batch over time. Fermentation must be properly controlled to improve the overall process efficiency, consistency and product quality. Bioreactors, which are widely used in the food industry, e.g. for alcoholic beverages, yoghurt and vinegar, can be a promising alternative. Recently, bioreactors have been used in coffee fermentation, but, in concrete, their implementation in the production chain in the countries of origin is not easy to apply. The use of starter cultures could be a friendly and less costly way to standardise the schedule of the fermentation process and reduce the negative effects of uncontrolled processes (Elhalis et al., 2023; Febrianto & Zhu, 2023). AI tools are, and will increasingly be, useful to optimize and control the fermentation process and for benchmarking between processes and/or flavour quality in food flavour analysis 4.0 (Bhargava et al., 2023; Zeng et al., 2023).

3.2.2. Drying & roasting

Drying is the next step after fermentation in the processing of coffee, cocoa and tea. Reducing humidity is essential to prevent mould and spoilage of leaves and beans. However, the process requires care to avoid the loss of bioactive components, overcooked leaves for tea and to preserve quality during storage (Herrera-Rocha et al., 2023; Nowogrodzki, 2019). The drying can be carried out either directly in open air or mechanically by driers, however in general the mode and progression of drying in the plantation differs because of practical and economic considerations including drying capacities, energy costs, and crop quality (Velásquez & Banchón, 2022). Despite its impact, drying is little studied. The dynamic changes of volatile profiles of green tea under different temperatures and times were investigated by Yang and co-workers using gas phase electronic nose (GC-E-Nose) and gas chromatography-ion mobility spectrometry (GC-IMS) combined with multivariate statistical analysis (Yanqin Yang et al., 2022). The authors discriminated by GC-E-Nose three drying stages despite the volatile profiles changed dynamically with the increase of the drying temperature. By applying PLS-DA to GC-IMS data they were able to select chemical components (3-methylbutanal, isopropyl alcohol, methyl benzoate, and heptanal) affecting the discrimination.

At the same time, inappropriate artificial drying of the cocoa beans to speed up the post-harvest process can impair quality, for example by producing a smoky off-flavour that cannot be removed in the subsequent steps of chocolate production. A top down analytical procedure was applied by Perotti et al. and Scavarda et al. where advanced fingerprinting obtained by the HS-SPME-GCxGC-MS from a set of representative smoky and non-smoky samples revealed the chemicals responsible for the off-flavour (Perotti et al., 2020; Scavarda et al., 2021). Perotti et al. applied an augmented visualization by a computer vision on untargeted fingerprints to define digital images differences between cocoas at molecular level (Caratti et al., 2023). Ten volatiles in the chromatographic pattern differed significantly between smoky and non-smoky samples, and they were independent of the origin or processing step considered (raw cocoa beans or liquors) (Perotti et al., 2020). The results served to develop a 1D-GC method for routine application in analysing cocoa beans and liquors. The reduction of dimensionality was obtained by PCA and the classification of samples reached by building up a PLS-DA model then cross-validated (5 C V); the resulting total classification rate was 97%. Achieving these results in liquors was not obvious since the target volatiles are also formed during roasting. Scavarda et al. then developed a quicker analytical decision maker for routine controls based on HS-SPME-MS-enose combined with machine learning tools to obtain diagnostic mass-spectral patterns to detect smoked samples. They tested two different model classifications PLS-DA and SIMCA, the latter models provided the best results, with sensitivities exceeding 90% and a high class specificity range of 89–100% depending on the matrix investigated (beans or liquors). The discrimination ability of the decision maker was then cross-validated by quantitative analysis through HS-SPME-GC–MS (Scavarda et al., 2021).

Roasting is a very important step of the whole production chain in particular for coffee and cocoa. The roasting profile is indeed crucial for the industry of the field because it affects chemical and physical properties of the final product, e.g. the external colour of the beans, weight loss, chemical composition, and the developed sensory characteristics of final products. During the roasting process, drying is the first effect of thermal energy, followed by non-enzymatic browning chemical reactions, such as Maillard reaction, Strecker degradation, and oxidation of lipids and polyphenols, which result in volatile and non-volatile chemical compounds that contribute to the flavour and aroma of the roasted beans (Velásquez & Banchón, 2022).

In general, in industry, the most adopted indicator to control the degree of roasting is colour or dry matter weight (for coffee) although these indicators are not related to the final flavour of beans. Several chemical markers including free amino acids, alkylpyrazines, chlorogenic acids, and their ratios were indicated as indirect markers of the degree of roasting. PCA and Fisher weight ratio algorithm on aroma components enabled to identify 16 volatiles better associated with colour degree that were used to build up a multiple linear regression (MLR) model with high goodness of fit and ± 2 error in colour prediction (Ruosi et al., 2012). Liberto et al. used direct mass spectral fingerprints both as a marker and as an analytical decision maker (ADM) in combination with OLPS regression to predict the colour degree of roasting with a high fitting (rpred 0.9472) and a satisfactory standard error of prediction (SEP 2.53) (Liberto et al., 2013). However, the limitation of this study was the sampling time by HS-SPME that was not well aligned for online monitoring of roasting, which was between 8 and 15 min depending on the required degree (light, medium, dark). Other direct mass spectroscopic approaches diagnostic at the molecular level, such as PTR-TOF-MS or Single-Photon Ionization Mass Spectrometry, are more consistent with the real-time monitoring of the roasting process. (Heide, Czech, Ehlert, Koziorowski, & Zimmermann, 2020).

Temperature and time settings of roasting are also fundamental in the developing cocoa flavour although companies define them as part of the production chain and, differently from coffee, often are standardized ranging from 120 to 160 °C for 20–40 min. Roasting produces aroma components and modulates the cocoa flavour because, besides Maillard reactions, the evaporation of acetic acid contributes to reduce acidity (Aprotosoaie, Vlad Luca, & Miron, 2016; Herrera-Rocha et al., 2023). Lemarcq et al. compared HS-SPME-GC-MS aroma and LC-MS profiles of dark chocolate derived from convectionally-roasted beans to chocolate produced from microwave-roasted cocoa beans revealing that chocolate made from microwave-roasted cocoa beans had a distinct scent profile compared to those submitted to convectional-roasting. Although more prone to oxidation, microwave roasting of cocoa beans was still within acceptable limits, making it a feasible alternative roasting method (Lemarcq, Monterde, et al., 2022). The same authors also studied a roasting profile preserving as much as possible the components that may contribute to psychopharmacological activities without compromising flavour, i.e. the mood pyramid of cocoa, a new concept consisting of four levels (flavan-3-ols, methylxanthines, minor compounds and orosensory properties) (Lemarcq et al., 2020). Using UPLC-HRMS, HS-SPME-GC-MS and sensory analysis, they found that roasting at 130 °C for 30 min had no significant effect on the content of epicatechin, procyanidin B2 and theobromine, while salsolinol increased significantly. In addition, bitterness and astringency were reduced while developing the desired cocoa flavour. Thus, interesting phytochemicals can be obtained without compromising the flavour by selecting suitable roasting time and temperature conditions.

AI-based tools offer remarkable advantages in monitoring thermal treatments or detecting errors in uncorrected processing along the production chain. AI-driven roasting processes provide significant potential benefits in terms of optimisation and efficiency. However, they also bring challenges related to the complexity requiring expertise in both roasting processes and AI algorithms, including data dependency related to the need for large amounts of data, potential errors due to not proper calibration, and the initial investment in equipment and training that can be a limit for companies. New technologies have been tested and others are under development within the industry in the effort to reduce the carbon footprint, in particular during the roasting process and to control the formation of process toxicants (e.g. furans, acrylamide). DoE applications can help to accelerate technological optimisation to meet R&D requirements. At the same time, data mining is useful to explore the flavour resulting from the new technology (ies), that, for example in coffee studies, is the bottleneck in the industrial strategies investigated to mitigate acrylamide (Strocchi, Rubiolo, et al., 2022).

3.3. Compositional and prediction of food properties

The most typical and widespread method used by the industry to assess the quality of cocoa, coffee and tea is sensory evaluation by trained panel test. Because of the high economic impact on the world market of these comfort foods, it is necessary to search for alternative tools, such as statistical methods, numerical measurements and standardized analyses to confirm their quality, complementing the conventional methods of interpreting the flavour. To address these issues, sensomics and flavouromics approaches adopt instrumental chemical analysis to judge flavour objectively and attempt to correlate aroma and taste molecules to sensory perception (Bressanello et al., 2018; Charve et al., 2011; Dunkel et al., 2014; Lindinger et al., 2008; Ribeiro, Augusto, Salva, & Ferreira, 2012; Sittipod et al., 2020). Extraction of chemical components from the matrix as well as analytical techniques have an impact on the information caught from the foods. Solvent extraction, extraction with supercritical CO2, distillation, headspace analysis by applying different high concentration capacity techniques are, in some cases, a bottleneck of the analytical strategies in terms of the informative dimension extracted. Mass spectrometry (MS), flame ionization detector (FID) sometimes in conjunction with an olfactory port, are used to detect aromatic chemicals separated by GC. The main tastes that determine the final flavour profile are bitterness, acidity, astringency, and sweetness which are analysed by high-/ultra-performance liquid chromatography (HPLC/UPLC), thin-layer chromatography (TLC) or capillary electrophoresis (CE). Additionally, non-destructive methods like nuclear magnetic resonance spectroscopy (NMR) and near-infrared spectroscopy (NIRS) are very popular since they enable the identification of different group(s) of molecules (Lemarcq, Van de Walle, Monterde, Sioriki, & Dewettinck, 2022). These analytical strategies have been used to characterize compositional data of complex cocoa, coffee and tea flavour in relation to different factors affecting it, and AI approaches have been extensively applied to connect instrumental data to human perception.

Compounds that influence the flavour of coffee brew by improving or suppressing bitterness or by positive modulating coffee quality have been revealed from Peterson's group works by applying untargeted flavouromic by LC/MS. These studies employed PCA to recognize outliers and then PLS-DA or OPLS-DA algorithms to identify, through VIP selection, features related to the cup score that are then isolated. The chemical structures are defined by MS and NMR and the relationship with sensory was made by recombinant models that are tested again from a sensory panel to define the coherence with specific attributes (Gao et al., 2021, 2023; Sittipod et al., 2019, 2020). Bressanello et al. and Liberto et al. used a similar approach for the coffee aroma prediction by applying HS-SPME-GC-MS and an HS-SPME-MS-Enose to the chemical definition of several coffee notes. The strategy applied, based on PCA and PLS-DA, was found to be discriminative and informative, identifying aroma compounds characteristic of the selected sensory notes. The predictive ability of PLS regression in defining the sensory scores of each aroma note was used as a validation tool for the chemical signatures of characterized notes with a standard error $\pm$ 2 in prediction of the sensory scores when using MS-enose as screening method (Bressanello et al., 2018; Liberto et al., 2019). The limitations of the method lie in the compromises that must be adopted when using a screening method as complementary to human evaluation in the sensory assessment of incoming raw materials. Gonzalez Viejo estimated coffee intensity and aroma using a low-cost, portable electronic nose (e-nose) in conjunction with two machine learning models built on artificial neural networks. The results demonstrated that it is possible to predict individual aromas as well as to evaluate the intensity of coffees with excellent accuracy (98%). The suggested non-contact, non-destructive, quick, dependable, and affordable technology has been successful in evaluating volatile chemicals in coffee at all stages of production, enabling early detection of any negative features and the constant assurance of high-quality products (Gonzalez Viejo, Tongson, & Fuentes, 2021). However, sensor technology exclusively produces response values for compounds featuring specific functional groups, such as alcohols, and exhibits heightened sensitivity to environmental fluctuations, including temperature and humidity. Spectroscopic techniques, mainly in the Near-Infrared (NIR) and Mid-Infrared (FTIR-Fourier Transform Infrared), have been widely used for coffee quality classification, followed by the application of statistical methods to the chemical data obtained, to develop classification models PLS, SVM, RF, DCNN (Deep convolutional neural networks) based on specific quality criteria. Neither SVM nor RF can solve multi-level classification problems. Therefore, ensemble learning was used for the multiple flavour prediction model achieving accuracy and recall >70% (Belchior et al., 2022; Ribeiro et al., 2012). The cost-benefit ratio of this and other instrumental screening approaches needs to be considered and weighed against the benefits of the potency of human response, which could be better utilized to modulate blends for sensory experiences outside the routine.

To characterize the composition of coffee, cocoa and tea flavour and assess the quality of these foods, several analytical approaches have been hyphenated and often analytical outputs are fused to find more objective tools that relate the chemical components to the sensory quality (Biancolillo et al., 2021; Bressanello et al., 2021; Lemarcq, Van de Walle, et al., 2022; Li et al., 2023). Data fusion refers to the process of combining or merging multiple sources of data to create a more comprehensive and accurate understanding of a particular phenomenon or situation. By fusing data from different sources, we can discover hidden relationships, detect anomalies, and make more informed and accurate conclusions. Generally, 3 levels of data fusion have been described in the literature: low-, mid-, and high-level. In low data fusion data from all analytical sources are directly fused in a single matrix, in mid-level classification algorithms are applied to each data separately to extract characteristics from the data relevant for prediction of the response before fusing them. These characteristics can consist of groups of data such as, for example, a subset of variables, a set of latent variables, or others such as shape or position in an image. In high-level fusion, first some supervised model is fit to each data matrix then results are fused for further treatments (Smolinska, Szymanska, Buydens, & Blanchet, 2019).

A comprehensive evaluation method to discriminate qualitatively between different grades of black tea was developed by combining Micro-NIR, computer vision, and colorimetric sensor array to collect data to apply multiple data fusion to create models by SVM, LS-SVM, EML and PLS-DA algorithms. Results indicated that the LS-SVM model with mid-level data fusion attained an accuracy of 98.57% in the testing set. Support vector regression was applied to determine flavour substances in black tea quantitatively with correlation coefficients for the predicted sets of gallic acid, caffeine, epigallocatechin, catechin, epigallocatechin gallate, epicatechin, gallocatechin gallate and total catechins ranging between 0.811 and 0.942 (Li et al., 2023). A mid-level data fusion on chromatographic fingerprinting strategy was also investigated by Bressanello et al. to delineate chemical patterns correlated to coffee odour and taste attributes and searching networks between features. The study was based on 155 commercial coffees where chemical data were obtained by analyzing volatile (HS-SPME-GC-MS) and non-volatile (LC-UV/DAD) fractions, and sensory data to simulate the main phases of the cupping protocol were measured according to the Specialty Coffee Association (SCA) that evaluates both smell and taste and were here used to describe sample sensory notes. Analytical platforms for fingerprinting were selected in light of routine control laboratory requirements for high batch-to-batch reproducibility, separation efficiency, and informative potentials. AI tools (MFA, PCA, ANOVA) were employed to explore datasets, select relevant features (PLS-DA), predict sensory scores (PLS regression), and investigate feature networks (HCA, correlation). PLS regression models utilized features from PLS-DA analysis for each sensory attribute. Comparison of Q model parameter and RMSEP revealed non-volatile fraction's greater influence on acid, bitter, and woody notes compared to flowery and fruity. Data fusion highlighted aroma's role in sensory perception, yet correlative networks between volatile and non-volatile data warrant further investigation for odour-taste integration potential. (Bressanello et al., 2021).

An example of high-level data fusion was used on cocoa to develop an analytical methodology affording to correlate the sensory poles of chocolate to their chemical characteristics (Biancolillo et al., 2021). Thirty-six different descriptors attributable to chocolate flavour were divided into 4 sensory poles samples and analysed by six different techniques: PTR-ToF-MS, SPME-GC-MS, LC-DAD (to quantify organic acids), UHPLC-QqQ-MS (to quantify polyphenols), 3D front face fluorescence spectroscopy and NIR. A multi-block classification approach (Sequential and Orthogonalized-Partial Least Squares – SO-PLS and SO-PLS-LDA) has been used to exploit the chemical information to predict the sensorial poles of samples. Information is sequentially extracted from various predictor blocks, comprising multiple matrices from different analytical platforms measured on the same samples. Data require pre-processing both within and across blocks. The SO-PLS approach involves a series of standard PLS regression and matrix orthogonalization operations to extract sequentially the complementary information from different data blocks, meaning that the aim is to incorporate blocks of data one at a time and to assess their incremental contribution (Biancolillo et al., 2021; Mishra et al., 2021). This strategy indicated that for chocolate samples, among the diverse analytical techniques applied, the combination of PTR-ToF-MS, Fluorescence and NIR spectroscopy led to the best results (classification rate of 93.6% on test samples). The models on the beans are less satisfactory and the analytical techniques able to predict bean samples resulted in different (UHPLC, HPLC blocks and NIR spectroscopy) meaning that the prediction of sensory poles is not based on the same constituents.

The flavour quality of coffee and cocoa emerges during roasting because of the chemical reactions, although stable compared to other perishable food, during storage, can change and the loss of volatile characteristic compounds and/or the appearance of oxidation products can cause off-flavours. This is particularly important in coffee where fresh notes can be quickly lost over time leading to staling phenomena. Aroma is also one of the important factors that determine the quality of green tea; a harmonious and pleasant aroma is an important standard for evaluating green tea and a vital consumer attribute that influences purchasing decisions. Since green tea is not fermented, it has a relatively low storage stability, especially during distribution and shelf-storage in retail outlets. Temperature, humidity, the presence of oxygen, light and the barrier capacity of the packaging are the key factors contributing to maintain the freshness of the flavour (Lin et al., 2022; Manzocco, Calligaris, Anese, & Nicoli, 2016; Strocchi, Bagnulo, et al., 2022; Strocchi, Bagnulo, Pellegrino, Bicchi, & Liberto, 2023; Strocchi, Müller, et al., 2023).

Teas not properly stored present a stale odour that has been related to the decreased content of aldehydes, and at the same time to the increase of ketones and heterocycles (furans, pyrroles, etc.). Key odourants indicated from a PLS-DA as (*E*,*E*)-2,4-heptadienal, α-terpineol, (*E*, *E*)-2,4-nonadienal and (*E*,*E*)-2,4-decadienal, appeared to be the main responsible for the stale odour. Baking treatment can remove the stale off-flavour typical of stored green tea improving the volatile profiles. Twenty-three compounds were identified following a sensomic workflow by applying HS-SPME-GC-O with OAVs >1, eight of them were further identified as the key volatiles by OPLS-DA (nonanal, (Z)-3-hexenyl hexanoate, benzaldehyde, β-ionone, linalool, β-cyclocitral, (E)-geranylacetone and α-terpineol) contributing to the green, fruity and roasted aromas of green tea (Liu et al., 2023).

Strocchi et al. defined the fingerprint describing oxidation notes in commercial coffee stored in different packaging (i.e., standard with aluminium barrier and Eco-caps). The study was carried out using HS-SPME-GC-MS/FPD in conjunction with machine learning data processing extrapolating 25 volatiles (out of 147) indicative of oxidised coffees regardless of packaging and blending. They noted that of the 25 volatiles that synergistically are responsible for the aging of coffee, some components are consistent with previous studies, and resulted in lower levels in oxidised coffees, while others appear to be specific to the packaging and/or blend, including 2-methylfuran, 2-methylbutanal, 2-acetylfuran, 2,5-dimethylfuran (meaty) and 2-methyl-2-cyclopenten-1-one (Strocchi, Bagnulo, et al., 2022).

The reported correlative studies suggest that the integrated approach can successfully be used to complement sensory analysis offering valuable insights into the relationship between objective measurements and subjective perceptions. In general, success in developing these methods requires a high level of consistency and alignment of the sensory panel in product evaluation, as subjectivity in data collection can influence the development of the mathematical model used to predict the score. However, the natural variability of these food matrices and their complexity make it difficult to achieve good representativeness for all commercial products handled at an industrial level. Furthermore, correlative studies often involve complex data analysis techniques and interpretation, requiring expertise in both sensory evaluation and analytical instrumentation. The need to combine different analytical techniques for coverage of multiple sensory attributes results in cost and resource intensiveness in terms of time, equipment, and personnel.

4. Perspectives and critical vision: the good, the bad and the ugly

AI-based tools play a crucial role in addressing real-world challenges in the coffee, cocoa, and tea industries, offering a multifactorial approach. It aids in authenticity detection, adulteration prevention, process control, and maintaining consistent product quality. This enhances data comprehension, particularly regarding the involved chemistry, crucial for product development and brand consistency. Various studies explore the chemistry of sensory notes and their relation to industrial issues using traditional AI algorithms. AI contributes to efficiency, cost-effectiveness, and overall industrial success.

The good. The AI tools enable researchers to analyse information from large amounts of data by breaking down complex issues into manageable components and understanding the relationships between them and the initial hypothesis in order to confirm it or propose a new hypothesis. Indeed, research studies are based on hypothesis-driven approaches that guide the design of experiments to obtain information. However, when there is not enough data or the hypothesis does not work, individuals can look for the causes, explore alternative solutions, observe with a keen eye, and make informed decisions based on careful analysis that generates new hypotheses (hypothesis-generating approaches) for an effective problem solving. In this regard, the extraction of hidden and helpful information from large data set by chemometrics does not fall into the hands of software in which "by clicking a button gives the solution" but needs researchers able to learn the use of different data mining methods depending on the problem to solve contributing to innovation by challenging existing ideas and exploring new possibilities (Jimenez-Carvelo, 2021).

The bad. The approximation in complex problems. Reducing data while preserving vital variation often involves PCA as a diagnostic tool and PLS/PLS-DA as common supervised algorithms. Converging instrumental approaches with sensory evaluation is essential for managing the complexity and multidimensionality of data. Although PCA and PLSR typically require linearity, previous studies rarely found a linear link between flavour perception and chemical composition (Lemarcq et al., 2020). While linear models like PCA and PLSR offer acceptable approximations, their performance can suffer from significant inter-variable nonlinearity. Nonlinear algorithms such as SVM, RF, ANN, deep learning, and hybrids have successfully addressed this issue in food flavour analysis. However, their limited use in coffee, cocoa, and tea research, where high sample numbers are often required, poses challenges from analytical and industrial perspectives, particularly when referencing sensory analysis. Overfitting remains a common challenge. Insufficient critical thinking during model development can contribute to inaccuracies, and overlooking data scaling, normalization, or transformation issues may lead to biased results. Analytical techniques providing numerous measures may not offer detailed molecular profiles correlating with sensory features (i.e., e-nose, NIR etc.). The use of fused data can benefit these algorithms, given an adequate number of representative samples. However, a gap exists in academic training for advanced AI tools, hindering alignment with industry modernization needs (Ayres et al., 2021; Zeng et al., 2023). The disparity between academia and industry hinders the implementation of methods used by academic researchers in companies. Nevertheless, the growing number of publications exploring the relationship between chemical data and sensory properties through instrumental analysis using machine learning reflects industry interests in this context.

The ugly. Misinterpreting chemometric results can lead to flawed conclusions and decisions based on inaccurate data analysis. Users, without a critical approach, might overlook non-linearity in the data, resulting in unreliable predictions. Complex AI algorithms can lack transparency, making it challenging to comprehend underlying data relationships. As noted by other authors, this becomes problematic without critical evaluation, impacting model reliability (Jiménez-Carvelo et al., 2019; Lemarcq et al., 2022; Yu et al., 2018). Ensuring the robustness and reliability of chemometric applications requires awareness of potential pitfalls, including overfitting, assumption violations, and result misinterpretation. Finally, there is misalignment in the literature as to what is covered by the term "artificial intelligence" due to its wide applications in different domains that need a clearer definition in function of the field of application.

Author statement

The authors declare that they have no financial and personal relationships with other people or organizations that can inappropriately influence their work.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

Acknowledgments

The authors thank the PhD programme of the University of Turin and the Fondo per la Ricerca Locale (Ex 60%) Anno 2021 financed from the University of Turin.

References

Acierno, V., Alewijn, M., Zomer, P., & van Ruth, S. M. (2018). Making cocoa origin traceable: Fingerprints of chocolates using flow infusion - electro spray ionization - mass spectrometry. *Food Control, 85*, 245–252. https://doi.org/10.1016/j.foodcont.2017.10.002

Amigo, J. M. (2021). Data mining, machine learning, deep learning, chemometrics: Definitions, common points and trends (spoiler alert: VALIDATE your models!). *Brazilian Journal of Analytical Chemistry, 8*(32), 22–38. https://doi.org/10.30744/brjac.2179-3425.AR-38-2021

Andre, C. M., & Soukoulis, C. (2020). Food quality assessed by chemometrics. *Foods, 9*(7), 897. https://doi.org/10.3390/foods9070897

Andruszkiewicz, P. J., Corno, M., & Kuhnert, N. (2021). HPLC-MS-based design of experiments approach on cocoa roasting. *Food Chemistry, 360*, Article 129694. https://doi.org/10.1016/J.FOODCHEM.2021.129694

Aprotosoaie, A. C., Vlad Luca, S., & Miron, A. (2016). Flavor chemistry of cocoa and cocoa products-an Overview. *Comprehensive Reviews in Food Science and Food Safety, 15*(1), 73–91. https://doi.org/10.1111/1541-4337.12180

Aswathi, K. N., Shirke, A., Praveen, A., Chaudhari, S. R., & Murthy, P. S. (2023). Pulped natural/honey robusta coffee fermentation metabolites, physico-chemical and sensory profiles. *Food Chemistry, 429*(April), Article 136897. https://doi.org/10.1016/j.foodchem.2023.136897

Ayres, L. B., Gomez, F. J. V., Linton, J. R., Silva, M. F., & Garcia, C. D. (2021). Taking the leap between analytical chemistry and artificial intelligence: A tutorial review. *Analytica Chimica Acta, 1161*, Article 338403. https://doi.org/10.1016/j.aca.2021.338403

Badmos, S., Fu, M., Granato, D., & Kuhnert, N. (2020). Classification of Brazilian roasted coffees from different geographical origins and farming practices based on chlorogenic acid profiles. *Food Research International, 134*, Article 109218. https://doi.org/10.1016/J.FOODRES.2020.109218

Bagnulo, E., Scavarda, C., Bortolini, C., Cordero, C., Bicchi, C., & Liberto, E. (2023). Cocoa quality: Chemical relationship of cocoa beans and liquors in origin identitation. *Food Research International, 172*(June), Article 113199. https://doi.org/10.1016/j.foodres.2023.113199

Balcázar-Zumaeta, C. R., Castro-Alayo, E. M., Cayo-Colca, I. S., Idrogo-Vásquez, G., & Muñoz-Astecker, L. D. (2023). Metabolomics during the spontaneous fermentation in cocoa (Theobroma cacao L.): An exploraty review. *Food Research International, 163* (September 2022). https://doi.org/10.1016/j.foodres.2022.112190

Barba, C., Beno, N., Guichard, E., & Thomas-Danguin, T. (2018). Selecting odorant compounds to enhance sweet flavor perception by gas chromatography/olfactometry-associated taste (GC/O- AT). *Food Chemistry, 257*(172−181). https://doi.org/10.1016/j.foodchem.2018.02.152

Bhargava, A., Bansal, A., Goyal, V., & Shukla, A. (2023). Machine learning & computer vision-based optimum black tea fermentation detection. *Multimedia Tools and Applications*, 43335–43347. https://doi.org/10.1007/s11042-023-15453-3

Bhatia, M., & Ahanger, T. A. (2021). Intelligent decision-making in smart food industry: Quality perspective. *Pervasive and Mobile Computing, 72*, Article 101304. https://doi.org/10.1016/j.pmcj.2020.101304

Biancolillo, A., Preys, S., Gaci, B., Le-Quere, J. L., Laboure, H., Deuscher, Z., et al. (2021). Multi-block classification of chocolate and cocoa samples into sensory poles. *Food Chemistry, 340*, Article 127904. https://doi.org/10.1016/J.FOODCHEM.2020.127904

Bilfield, A. (2022). *Brewing sustainability in the coffee and tea industries: From producer to consumers*. Routledge. https://doi.org/10.4324/9781003228851

Boeckx, P., Bauters, M., & Dewettinck, K. (2020). Poverty and climate change challenges for sustainable intensification of cocoa systems. *Current Opinion in Environmental Sustainability, 47*, 106–111. https://doi.org/10.1016/j.cosust.2020.10.012

Bressanello, D., Liberto, E., Cordero, C., Rubiolo, P., Pellegrino, G., Ruosi, M. R., et al. (2017). Coffee aroma: Chemometric comparison of the chemical information provided by three different samplings combined with GC–MS to describe the sensory properties in cup. *Food Chemistry*. https://doi.org/10.1016/j.foodchem.2016.07.088

Bressanello, D., Liberto, E., Cordero, C., Sgorbini, B., Rubiolo, P., Pellegrino, G., et al. (2018). Chemometric modeling of coffee sensory notes through their chemical signatures: Potential and limits in defining an analytical tool for quality control. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.8b01340

Bressanello, D., Marengo, A., Cordero, C., Strocchi, G., Rubiolo, P., Pellegrino, G., et al. (2021). Chromatographic fingerprinting strategy to delineate chemical patterns correlated to coffee odor and taste attributes. *Journal of Agricultural and Food Chemistry, 69*(15), 4550–4560. https://doi.org/10.1021/acs.jafc.1c00509

Caratti, A., Squara, S., Bicchi, C., Tao, Q., Geschwender, D., Reichenbach, S. E., et al. (2023). Augmented visualization by computer vision and chromatographic fingerprinting on comprehensive two-dimensional gas chromatographic patterns: Unraveling diagnostic signatures in food volatilome. *Journal of Chromatography A, 1699*, Article 464010. https://doi.org/10.1016/j.chroma.2023.464010

Casale, M., Malegori, C., Oliveri, P., Liberto, E., Rubiolo, P., Bicchi, C., et al. (2020). Chapter 10: Chemometrics: Basic principles and applications. *Food Chemistry, Function and Analysis, 2020-January*. https://doi.org/10.1039/9781788015752-00403

Cevallos-Cevallos, J. M., Gysel, L., Maridueña-Zavala, M. G., & Molina-Miranda, M. J. (2018). Time-related changes in volatile compounds during fermentation of bulk and fine-flavor cocoa (theobroma cacao) beans. *Journal of Food Quality, 2018*. https://doi.org/10.1155/2018/1758381

Charve, J., Chen, C., Hegeman, A. D., & Reineccius, G. A. (2011). Evaluation of instrumental methods for the untargeted analysis of chemical stimuli of orange juice flavour. *Flavour and Fragrance Journal, 26*(6), 429–440. https://doi.org/10.1002/FFJ.2078

Chen, T., Zhang, J., & Luo, J. (2021). Differential game evolution of food quality safety based on market supply and demand. *Food Science and Nutrition, 9*(5), 2414–2435. https://doi.org/10.1002/FSN3.2128

Cifuentes, A. (2017). Foodomics, foodome and modern food analysis. *TrAC, Trends in Analytical Chemistry, 96*, 1. https://doi.org/10.1016/j.trac.2017.09.001

Cordero, C., Kiefl, J., Reichenbach, S. E., & Bicchi, C. (2019). Characterization of odorant patterns by comprehensive two-dimensional gas chromatography: A challenge in omic studies. *TrAC, Trends in Analytical Chemistry, 113*, 364–378. https://doi.org/10.1016/J.TRAC.2018.06.005

Cuadros-Rodríguez, L., Ortega-Gavilán, F., Martín-Torres, S., Arroyo-Cerezo, A., & Jiménez-Carvelo, A. M. (2021). Chromatographic fingerprinting and food identity/quality: Potentials and challenges. *Journal of Agricultural and Food Chemistry, 69*(48), 14428–14434. https://doi.org/10.1021/acs.jafc.1c05584

Cuadros-Rodríguez, L., Ruiz-Samblás, C., Valverde-Som, L., Pérez-Castaño, E., & Casado, A. G. (2016). Chromatographic fingerprinting: An innovative approach for food 'identitation' and food authentication - a tutorial. *Analytica Chimica Acta, 909*, 9–23. https://doi.org/10.1016/j.aca.2015.12.042

De Vuyst, L., & Leroy, F. (2020). Functional role of yeasts, lactic acid bacteria and acetic acid bacteria in cocoa fermentation processes. *FEMS Microbiology Reviews, 44*(4), 432–453. https://doi.org/10.1093/femsre/fuaa014

Dunkel, A., Steinhaus, M., Kotthoff, M., Nowak, B., Krautwurst, D., Schieberle, P., et al. (2014). *Nature s chemical signatures in human olfaction : A foodborne perspective for future biotechnology* (Vols. 2–22). https://doi.org/10.1002/anie.201309508

Elhalis, H., Cox, J., & Zhao, J. (2023). Coffee fermentation: Expedition from traditional to controlled process and perspectives for industrialization. *Applied Food Research, 3*(1), Article 100253. https://doi.org/10.1016/j.afres.2022.100253

Febrianto, N. A., & Zhu, F. (2023). Coffee bean processing: Emerging methods and their effects on chemical, biological and sensory properties. *Food Chemistry*. https://doi.org/10.1016/j.foodchem.2023.135489. Elsevier Ltd.

Flores-valdez, M., Gabriela, O., Osorio-revilla, G., & Gallardo-vel, T. (2020). Coffee (Coffea arabica L.) using FT-MIR spectroscopy coupled with chemometrics. *Foods, 9* (7), 851. https://doi.org/10.3390/foods9070851

Frost, S. C., Ristenpart, W. D., & Guinard, J. X. (2020). Effects of brew strength, brew yield, and roast on the sensory quality of drip brewed coffee. *Journal of Food Science, 85*(8), 2530–2543. https://doi.org/10.1111/1750-3841.15326

Gao, C., Tello, E., & Peterson, D. G. (2021). Identification of coffee compounds that suppress bitterness of brew. *Food Chemistry, 350*, Article 129225. https://doi.org/10.1016/J.FOODCHEM.2021.129225

Gao, C., Tello, E., & Peterson, D. G. (2023). Identification of compounds that enhance bitterness of coffee brew. *Food Chemistry, 415*, Article 135674. https://doi.org/10.1016/J.FOODCHEM.2023.135674

Gonzalez Viejo, C., Tongson, E., & Fuentes, S. (2021). Integrating a low-cost electronic nose and machine learning modelling to assess coffee aroma profile and intensity. *Sensors, 21*(6), 1–16. https://doi.org/10.3390/s21062016

Granato, D., Putnik, P., Kovačević, D. B., Santos, J. S., Calado, V., Rocha, R. S., et al. (2018). Trends in chemometrics: Food authentication, microbiology, and effects of processing. *Comprehensive Reviews in Food Science and Food Safety, 17*(3), 663–677. https://doi.org/10.1111/1541-4337.12341

Gu, H. W., Yin, X. L., Peng, T. Q., Pan, Y., Cui, H. N., Li, Z. Q., et al. (2022). Geographical origin identification and chemical markers screening of Chinese green tea using two-dimensional fingerprints technique coupled with multivariate chemometric methods. *Food Control, 135*, Article 108795. https://doi.org/10.1016/J.FOODCONT.2021.108795

Guichard, E., Barba, C., Thomas-Danguin, T., & Tromelin, A. (2019). Multivariate statistical analysis and odor-taste network to reveal odor-taste associations. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.9b05462

Gutiérrez-Ríos, H. G., Suárez-Quiroz, M. L., Hernández-Estrada, Z. J., Castellanos-Onorio, O. P., Alonso-Villegas, R., Rayas-Duarte, P., et al. (2022). Yeasts as producers of flavor precursors during cocoa bean fermentation and their relevance as starter cultures: A review. *Fermentation, 8*(7). https://doi.org/10.3390/fermentation8070331

Heide, J., Czech, H., Ehlert, S., Koziorowski, T., & Zimmermann, R. (2020). Toward smart online coffee roasting process control: Feasibility of real-time prediction of coffee roast degree and brew antioxidant capacity by single-photon ionization mass spectrometric monitoring of roast gases. *Journal of Agricultural and Food Chemistry, 68*(17), 4752–4759. https://doi.org/10.1021/ACS.JAFC.9B06502/SUPPL_FILE/JF9B06502_SI_001.PDF

Herrera-Rocha, F., Fernández-Niño, M., Cala, M. P., Duitama, J., & Barrios, A. F. G. (2023). Omics approaches to understand cocoa processing and chocolate flavor development: A review. *Food Research International*. https://doi.org/10.1016/j.foodres.2023.112555

IFIC, International Food Information Council. (2017). *Food & health survey* (Vol. 68). International Food Information Council.

ISO 2200. (2018). *International Organization for Standardization. Food safety management systems Requirements for any organization in the food chain*. Retrieved 2 September 2023, from https://www.iso.org/standard/65464.html.

ISO 9001. (2015). *International Organization for Standardization. Quality management systems requirements*. Retrieved 2 September 2023, from https://www.iso.org/standard/62085.html.

ISO 20715. (2023). *Tea.International Organization for Standardization. Classification of tea types*. Retrieved 15 October 2023 from https://www.iso.org/obp/ui/en/#iso:std:iso:20715:ed-1:v1:en.

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

Jimenez-Carvelo, A. M. (2021). Data mining/machine learning methods in foodomics. *Current Opinion in Food Science, 37*, 76–82. https://doi.org/10.1016/J.COFS.2020.09.008

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

John, W. A., Böttcher, N. L., Behrends, B., Corno, M., D'souza, R. N., Kuhnert, N., et al. (2020). Experimentally modelling cocoa bean fermentation reveals key factors and their influences. *Food Chemistry, 302*, Article 125335. https://doi.org/10.1016/J.FOODCHEM.2019.125335

Kumar, S., D'Souza, R. N., Behrends, B., Corno, M., Ullrich, M. S., Kuhnert, N., et al. (2021). Cocoa origin classifiability through LC-MS data: A statistical approach for large and long-term datasets. *Food Research International, 140*, Article 109983. https://doi.org/10.1016/J.FOODRES.2020.109983

Kumari, N., Grimbs, A., D'Souza, R. N., Verma, S. K., Corno, M., Kuhnert, N., et al. (2018). Origin and varietal based proteomic and peptidomic fingerprinting of Theobroma cacao in non-fermented and fermented cocoa beans. *Food Research International, 111*, 137–147. https://doi.org/10.1016/J.FOODRES.2018.05.010

Lemarcq, V., Monterde, V., Tuenter, E., Walle, D. Van de, Pieters, L., Sioriki, E., et al. (2022). Flavor diversification of dark chocolate produced through microwave roasting of cocoa beans. *Lwt, 159*(February). https://doi.org/10.1016/j.lwt.2022.113198

Lemarcq, V., Tuenter, E., Bondarenko, A., Van de Walle, D., De Vuyst, L., Pieters, L., et al. (2020). Roasting-induced changes in cocoa beans with respect to the mood pyramid. *Food Chemistry, 332*(June), Article 127467. https://doi.org/10.1016/j.foodchem.2020.127467

Lemarcq, V., Van de Walle, D., Monterde, V., Sioriki, E., & Dewettinck, K. (2022). Assessing the flavor of cocoa liquor and chocolate through instrumental and sensory analysis: A critical review. *Critical Reviews in Food Science and Nutrition*. https://doi.org/10.1080/10408398.2021.1887076

Li, L., Chen, Y., Dong, S., Shen, J., Cao, S., Cui, Q., et al. (2023). Rapid and comprehensive grade evaluation of Keemun black tea using efficient multidimensional data fusion. *Food Chemistry X, 20*, Article 100924. https://doi.org/10.1016/J.FOCHX.2023.100924

Li, S., Tian, Y., Jiang, P., Lin, Y., Liu, X., & Yang, H. (2021). Recent advances in the application of metabolomics for food safety control and food quality analyses. *Critical Reviews in Food Science and Nutrition, 61*(9), 1448–1469. https://doi.org/10.1080/10408398.2020.1761287

Liberto, E., Bressanello, D., Strocchi, G., Cordero, C., Ruosi, M. R., Pellegrino, G., et al. (2019). HS-SPME-MS-enose coupled with chemometrics as an analytical decision maker to predict in-cup coffee sensory quality in routine controls: Possibilities and limits. *Molecules, 24*(24). https://doi.org/10.3390/molecules24244515

Liberto, E., Ruosi, M. R., Cordero, C., Rubiolo, P., Bicchi, C., & Sgorbini, B. (2013). Non-separative headspace solid phase microextraction-mass spectrometry profile as a marker to monitor coffee roasting degree. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/jf303067q

Lin, H., Tello, E., Simons, C. T., & Peterson, D. G. (2022). *Identification of non-volatile compounds generated during storage that impact flavor stability of ready-to-drink coffee*.

Lindinger, C., Labbe, D., Pollien, P., Rytz, A., Juillerat, M. A., Yeretzian, C., et al. (2008). When machine tastes coffee: Instrumental approach to predict the sensory profile of espresso coffee. *Analytical Chemistry, 80*(5), 1574–1581. https://doi.org/10.1021/ac702196z

Liu, P. P., Feng, L., Xu, Y. Q., Zheng, L., Yin, P., Ye, F., et al. (2023). Characterization of stale odor in green tea formed during storage: Unraveling improvements arising from reprocessing by baking. *Lwt, 174*(July 2022), Article 114458. https://doi.org/10.1016/j.lwt.2023.114458

Manzocco, L., Calligaris, S., Anese, M., & Nicoli, M. C. (2016). The stability and shelf life of coffe products. In P. Subramaniam, & P. Wareing (Eds.), *The stability and shelf life of foods* (pp. 375–396). Elsevier.

Marseglia, A., Musci, M., Rinaldi, M., Palla, G., & Caligiani, A. (2020). Volatile fingerprint of unroasted and roasted cocoa beans (Theobroma cacao L.) from different geographical origins. *Food Research International, 132*(February), Article 109101. https://doi.org/10.1016/j.foodres.2020.109101

Mascarello, G., Pinto, A., Parise, N., Crovato, S., & Ravarotto, L. (2015). The perception of food quality. Profiling Italian consumers. *Appetite, 89*, 175–182. https://doi.org/10.1016/J.APPET.2015.02.014

McClure, A. P., Spinka, C. M., & Grün, I. U. (2021). Quantitative analysis and response surface modeling of important bitter compounds in chocolate made from cocoa beans with eight roast profiles across three origins. *Journal of Food Science, 86*(11), 4901–4913. https://doi.org/10.1111/1750-3841.15924

Megias-Perez, R., Moreno-Zambrano, M., Behrends, B., Corno, M., & Kuhnert, N. (2020). Monitoring the changes in low molecular weight carbohydrates in cocoa beans during spontaneous fermentation: A chemometric and kinetic approach. *Food Research International, 128*(July 2019), Article 108865. https://doi.org/10.1016/j.foodres.2019.108865

Mishra, P., Roger, J. M., Jouan-Rimbaud-Bouveresse, D., Biancolillo, A., Marini, F., Nordon, A., et al. (2021). Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC, Trends in Analytical Chemistry, 137*, Article 116206. https://doi.org/10.1016/j.trac.2021.116206

Momtaz, M., Bubli, S. Y., & Khan, M. S. (2023). *Mechanisms and health aspects of food adulteration: A comprehensive review*.

Nicolotti, L., Mall, V., & Schieberle, P. (2019). Characterization of key aroma compounds in a commercial rum and an Australian red wine by means of a new sensomics-based expert system (SEBES) - an approach to use artificial intelligence in determining food odor codes. *Journal of Agricultural and Food Chemistry, 67*(14), 4011–4022. https://doi.org/10.1021/ACS.JAFC.9B00708/ASSET/IMAGES/LARGE/JF-2019-00708K_0001.JPEG

Nowogrodzki, A. (2019). How climate change might affect tea. *Nature, 566*(7742), S10–S11. https://doi.org/10.1038/D41586-019-00399-0

Oliveira, M. M., Badaró, A. T., Esquerre, C. A., Kamruzzaman, M., & Barbin, D. F. (2023). Handheld and benchtop vis/NIR spectrometer combined with PLS regression for fast prediction of cocoa shell in cocoa powder. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 298*(April). https://doi.org/10.1016/j.saa.2023.122807

Perotti, P., Cordero, C., Bortolini, C., Rubiolo, P., Bicchi, C., & Liberto, E. (2020). Cocoa smoky off-flavor: Chemical characterization and objective evaluation for quality control. *Food Chemistry, 309*(September 2019), Article 125561. https://doi.org/10.1016/j.foodchem.2019.125561

Peterson, R. T. (2008). Chemical biology and the limits of reductionism. *Nature Chemical Biology, 4*(11), 635–638. https://doi.org/10.1038/nchembio1108-635, 2008 4:11.

Preedy, V. R. (2013). In V. R. Preedy (Ed.), *Tea in health and disease prevention*. London, UK: Academic Press. https://doi.org/10.1016/C2010-0-64948-0.

Preedy, V. R. (2015). In V. Preedy (Ed.), *First editCoffee in health and disease prevention*. Academic Press. https://doi.org/10.1016/C2012-0-06959-1.

Prescott, J. (2015). Multisensory processes in flavour perception and their influence on food choice. *Current Opinion in Food Science, 3*, 47–52. https://doi.org/10.1016/j.cofs.2015.02.007

Quelal-Vásconez, M. A., Lerma-García, M. J., Pérez-Esteve, É., Talens, P., & Barat, J. M. (2020). Roadmap of cocoa quality and authenticity control in the industry: A review of conventional and alternative methods. *Comprehensive Reviews in Food Science and Food Safety, 19*(2), 448–478. https://doi.org/10.1111/1541-4337.12522

Ribeiro, J. S., Augusto, F., Salva, T. J. G., & Ferreira, M. M. C. (2012). Talanta Prediction models for Arabica coffee beverage quality based on aroma analyses and chemometrics. *Talanta, 101*, 253–260. https://doi.org/10.1016/j.talanta.2012.09.022

Ribeiro, L. S., Da Cruz Pedroso Miguel, M. G., Evangelista, S. R., Martins, P. M. M., van Mullem, J., Belizario, M. H., & Schwan, R. F. (2017). Behavior of yeast inoculated during semi-dry coffee fermentation and the effect on chemical and sensorial properties of the final beverage. *Food Research International, 92,*, 26–32. https://doi.org/10.1016/J.FOODRES.2016.12.011

Rodionova, O. Y., Titova, A. V., & Pomerantsev, A. L. (2016). Discriminant analysis is an inappropriate method of authentication. *Trends in Analytical Chemistry, 78*, 17–22. https://doi.org/10.1016/j.trac.2016.01.010

Ronningen, I., Miller, M., Xia, Y., & Peterson, D. G. (2018). *Identi fi cation and validation of sensory-active compounds from data-driven research: A flavoromics approach*. https://doi.org/10.1021/acs.jafc.7b00093

Ruosi, M. R., Cordero, C., Cagliero, C., Rubiolo, P., Bicchi, C., Sgorbini, B., & Liberto, E. (2012). A further tool to monitor the coffee roasting process: Aroma composition and chemical indices. *Journal of Agricultural and Food Chemistry, 60*, 11283–11291.

Saunshia, Y., Sandhya, M. K. V. S., Lingamallu, J. M. R., Padela, J., & Murthy, P. (2018). Improved fermentation of cocoa beans with enhanced aroma profiles. *Food Biotechnology, 32*(4), 257–272. https://doi.org/10.1080/08905436.2018.1519444

Scavarda, C., Cordero, C., Strocchi, G., Bortolini, C., Bicchi, C., & Liberto, E. (2021). Cocoa smoky off-flavour: A MS-based analytical decision maker for routine controls. *Food Chemistry, 336*(May 2020), Article 127691. https://doi.org/10.1016/j.foodchem.2020.127691

Shuai, M., Peng, C., Niu, H., Shao, D., Hou, R., & Cai, H. (2022). Recent techniques for the authentication of the geographical origin of tea leaves from camellia sinensis: A review. *Food Chemistry, 374*, Article 131713. https://doi.org/10.1016/J.FOODCHEM.2021.131713

Silva Fernandes, J., de Sousa Fernandes, D. D., Pistonesi, M. F., & Gonçalves Dias Diniz, P. H. (2023). Tea authentication and determination of chemical constituents using digital image-based fingerprint signatures and chemometrics. *Food Chemistry, 421*, Article 136164. https://doi.org/10.1016/J.FOODCHEM.2023.136164

Siridevi, G., Havare, D., Basavaraj, K., & Murthy, P. S. (2019). Coffee starter microbiome and in-silico approach to improve Arabica coffee. *Lebensmittel-Wissenschaft und -Technologie, 114*, Article 108382. https://doi.org/10.1016/J.LWT.2019.108382

Sittipod, S., Schwartz, E., Paravisini, L., & Peterson, D. G. (2019). *Identifi cation of fl avor modulating compounds that positively impact co ff ee quality* (Vol. 301). https://doi.org/10.1016/j.foodchem.2019.125250. July.

Sittipod, S., Schwartz, E., Paravisini, L., Tello, E., & Peterson, D. G. (2020). *Identi fi cation of compounds that negatively impact Co ff ee flavor quality using untargeted liquid chromatography/mass spectrometry analysis* (Vols. 0–7). https://doi.org/10.1021/acs.jafc.0c01479

Smolinska, A., Szymanska, E., Buydens, L., & Blanchet, L. (2019). General framing of low-, mid-, and high-level data fusion with examples in the life sciences. *Data Handling in Science and Technology, 3*, 51–79. https://doi.org/10.1016/B978-0-444-63984-4.00003-X

Spence, C. (2016a). Enhancing the experience through smell. *Food Science and Technology, 30*(2), 32–35. https://doi.org/10.1002/fsat.3002_11.x

Spence, C. (2016b). Oral referral: On the mislocalization of odours to the mouth. *Food Quality and Preference, 50*, 117–128. https://doi.org/10.1016/j.foodqual.2016.02.006

Spence, C. (2017). Comfort food: A review. *International Journal of Gastronomy and Food Science, 9*, 105–109. https://doi.org/10.1016/j.ijgfs.2017.07.001

Squara, S., Caratti, A., Fina, A., Liberto, E., Spigolon, N., Genova, G., et al. (2023). Artificial intelligence decision-making tools based on comprehensive two-dimensional gas chromatography data: The challenge of quantitative volatilomics in food quality assessment. *Journal of Chromatography A, 1700*, Article 464041. https://doi.org/10.1016/J.CHROMA.2023.464041

Strocchi, G., Bagnulo, E., Pellegrino, G., Bicchi, C., & Liberto, E. (2023). Evaluation of the behaviour of phenols and alkaloids in samples of roasted and ground coffee stored in different types of packaging: Implications for quality and shelf life. *Food Research International, 174*(P1), Article 113548. https://doi.org/10.1016/j.foodres.2023.113548

Strocchi, G., Bagnulo, E., Ruosi, M. R., Ravaioli, G., Trapani, F., Bicchi, C., et al. (2022). Potential aroma chemical fingerprint of oxidised coffee note by HS-SPME-GC-MS and machine learning. *Foods, 11*(24). https://doi.org/10.3390/foods11244083

Strocchi, G., Müller, A. B., Kuhnert, N., Martina, K., Bicchi, C., & Liberto, E. (2023). Diterpenes stability of commercial blends of roasted and ground coffees packed in copolymer coupled with aluminium and eco-friendly capsules. *Food Research International, 174*(P1), Article 113577. https://doi.org/10.1016/j.foodres.2023.113577

Strocchi, G., Rubiolo, P., Cordero, C., Bicchi, C., & Liberto, E. (2022). Acrylamide in coffee: What is known and what still needs to be explored. A review. *Food Chemistry, 393*, Article 133406. https://doi.org/10.1016/J.FOODCHEM.2022.133406

Szymańska, E. (2018). Modern data science for analytical chemical data-A comprehensive review. *Analytica Chimica Acta, 1028*, 1–10. /1–10.10.1016/j.aca.2018.05.038.

Velásquez, S., & Banchón, C. (2022). Influence of pre-and post-harvest factors on the organoleptic and physicochemical quality of coffee: A short review. *Journal of Food Science and Technology, 60*(10), 2526–2538. https://doi.org/10.1007/s13197-022-05569-z

Wang, H., Chambers, E., & Kan, J. (2018). Sensory characteristics of combinations of phenolic compounds potentially associated with smoked aroma in foods. *Molecules, 23*(8). https://doi.org/10.3390/molecules23081867

Wei, Y., Yin, X., Zhao, M., Zhang, J., Li, T., Zhang, Y., et al. (2023). Metabolomics analysis reveals the mechanism underlying the improvement in the color and taste of yellow tea after optimized yellowing. *Food Chemistry, 428*, Article 136785. https://doi.org/10.1016/J.FOODCHEM.2023.136785

Wen, S., Jiang, R., An, R., Ouyang, J., Liu, C., Wang, Z., et al. (2023). Effects of pile-fermentation on the aroma quality of dark tea from a single large-leaf tea variety by GC × GC-QTOFMS and electronic nose. *Food Research International, 174*(P1), Article 113643. https://doi.org/10.1016/j.foodres.2023.113643

Yang, Y., Chen, J., Jiang, Y., Qian, M. C., Deng, Y., Xie, J., et al. (2022). Aroma dynamic characteristics during the drying process of green tea by gas phase electronic nose and gas chromatography-ion mobility spectrometry. *Lwt, 154*(June 2021), Article 112691. https://doi.org/10.1016/j.lwt.2021.112691

Yang, S., Li, C., Mei, Y., Liu, W., Liu, R., Chen, W., et al. (2021). Determination of the geographical origin of coffee beans using terahertz spectroscopy combined with machine learning methods. *Frontiers in Nutrition, 8*(June), 1–10. https://doi.org/10.3389/fnut.2021.680627

Yu, P., Low, M. Y., & Zhou, W. (2018). Development of a partial least squares-artificial neural network (PLS-ANN) hybrid model for the prediction of consumer liking scores of ready-to-drink green tea beverages. *Food Research International, 103*, 68–75. https://doi.org/10.1016/J.FOODRES.2017.10.015

Zakaria, N. H., Whanmek, K., Thangsiri, S., Chathiran, W., Srichamnong, W., Suttisansanee, U., et al. (2023). Optimization of cold brew coffee using central composite design and its properties compared with hot brew coffee. *Foods, 12*(12). https://doi.org/10.3390/FOODS12122412/S1

Zeng, X., Cao, R., Xi, Y., Li, X., Yu, M., Zhao, J., et al. (2023). Food 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

Glossary

AAB: Acetic acid batecria
ADM: Analytical decision maker
AI: Artificial inteligence
ANN: Artificial neuronal networks
BBD: Box-behnken design
CNN: Convolutional neural network
CQA: Chlogenic acids
CV: Cross-validation
DCNN: Deep convolutional neural networks
DOE: Design of experiments
EMCVS: Ensemble monte carlo variable selection
e-Nose: Electronic nose
FID: Flame ionization detection
FTIR: Fourier transform infrared
FT-MIR: Fourier-transform mid infrared
GA: Genetic algorithm
GC-MS: Gas chromatograph-mass spectrometry
GC-O: Gas chromatography-olfactometry
GCxGC-MS: Comprehensive gas chromatograph-mass spectrometry
HCA: Hierarchical cluster analysis
HS-SPME: Headspace solid phase microextraction
IMS: Ion mobility mass spectrometry
LAB: Lactic acid bacteria
LCxLC: Comprehensive liquid chromatograph
LDA: Linear discriminant analysis
LS-SVM: Least square support vector machine
MCR-ALS: Multi curve resolution-alternative least square
MFA: Multiple factor analysis
MLR: Multiple linear regression
MSC: Multiplicative scatter correction
NIR: Near-infrared spectrometry
NMR: Nuclear magnetic resonance
N-PLS: Multiway partial least square
OAV: Odour activity value
OPLS: Orthogonal projection to latent structures
PARAFAC: Parallel factor analysis
PCA: Principal component analysis
pCoQA: *p*-coumaroylquinic acid
PCR: Principal component regression
PDO: Protected designation of origin
PGI: Protected geographical indication
PLS or PLSR: Partial least squares regression
PLS-DA: Partial least squares discriminant analysis
PTR-MS: Proton transfer reaction mass spectrometry
QMS: Quality management systems
RF: Random forest
RMSEP: Root-mean-squared error prediction
RSM: Response surface methodology
SEBES: Sensomics-based expert system
SEP: Standard error in preditction
SIMCA: Soft independent modelling of class analogy
SRC: Sparse representation classifier
SVM: Support vector machine