Letters in Food Research 1(4): e25050
Journal homepage: https://www.lettersinfoodresearch.com

LETTERS IN FOOD RESEARCH

REVIEW

Machine learning in food authentication: A sustainable solution to global food challenges

Malik, M. and *Dewan, A.

Department of Food Technology, Guru Jambheshwar University of Science and Technology, Hisar (125001), Haryana, India

Article history:
Received: 9 July 2025
Received in revised form: 13 August 2025
Accepted: 13 August 2025
Available Online: 6 September 2025

Keywords:
Adulteration,
Edible oils,
Food safety,
Honey,
Neural network,
Origin

OPEN ACCESS
Citation: Malik, M. and Dewan, A. (2025). Machine learning in food authentication: A sustainable solution to global food challenges. *Letters in Food Research*, 1(4), e25050. https://doi.org/10.26656/lifr.1.e25050

Abstract

Authenticity of a food product is one of the major global food challenges. Ensuring consumers about origin, variety, composition, processing method, shelf life and purity of a food product is challenging for food producers and manufacturers. Traditional methods of food authentication are time consuming and labour intensive and can handle only limited datasets. While machine learning (ML) algorithms can serve as a sustainable and rapid approach through handling of larger datasets thereby overcoming the demerits of traditional methods of food authentication. However, no comprehensive review article covering all published literature of last 5 years is available till now. Therefore, this review was designed to discuss the implementation of various ML algorithms in authentication of various food commodities such as honey, beverages like beer, juices and wine, edible oils, seafoods, spices, cereals, meat, fish, egg, fruits, vegetables, milk and milk products etc. Neural networking (NN), decision trees, random forests, support vector machines (SVM), principal component analysis (PCA) and autoencoder with amalgamation with spectroscopic and chromatographic techniques provide a robust and rapid approach for building trusts between producers and consumers and fair-trade practices. ML has transformative potential in addressing food integrity challenges with profound successful implementation in food industry to enhance the food safety and quality. This information will help researchers and regulatory bodies to ensure consumers and traders for food quality and safety.

1. Introduction

Rising evidence of food frauds results in the loss of consumer trust, leading to a decline in brand loyalty, economic losses and potential health risks. Food safety and quality control is a crux for ensuring the authenticity of food products. Food authenticity is essential to build and maintain consumer trust by preventing fraudulent practices such as substitution, dilution, mislabeling, adulteration and counterfeiting issues that have become increasingly prevalent in the global food supply chain (Brooks *et al*., 2021). According to the Grocery Manufacturers Association (GMA), food fraud costs the world between $40 billion annually in revenue loss (GMA, 2017). Given its detrimental impact on human health, food product adulteration has raised awareness of the need for strict food safety and quality control procedures in the fields of food chemistry and related sciences (Bhat *et al*., 2025).

Traditional approach in detecting food frauds includes analytical analysis employing chromatographic techniques, spectroscopic characterization and polymerize chain reaction (PCR). However, these techniques are reasonably effective, but require specialized operators, time consuming, and labor-intensive (Fernando *et al*., 2024). Technological innovations of blending traditional analytical techniques and machine learning (ML) algorithms have revolutionized food authentication process by offering rapid, more accurate and automated diagnosis for detecting food fraud and ensuring traceability (Chhetri *et al*., 2024). Without explicit programming, ML algorithms involve a range of computational techniques that allows systems to recognize patterns in large datasets and provide predictions. Methods like supervised learning (neural networking (NN), decision trees, random forests, support vector machines (SVM)), and unsupervised learning (Hierarchical cluster analysis (HCA), K-means clustering, principal component analysis (PCA) and autoencoder) have shown incredible promise in the analysis of complex food-related data (Tseng *et al*., 2023).

Strong prediction models that make use of spectral

*Corresponding author.
Email: *23aastha10@gmail.com*

eISSN: 3093-7361 /
Published by Rynnye Lyan Resources

Figure 1. Global food challenges and frauds in different food commodities such as fruits and vegetables, cereals, pulses and oilseeds, spices and edible oils, beverages, honey, milk and milk products and meat, seafood and poultry products.

data, image analysis, and chemical composition assessment have been developed as a result of integrating machine learning (ML) into food authentication. For instance, food samples have been successfully classified and adulterants identified using near-infrared (NIR) and Raman spectroscopy in conjunction with ML methods such as SVM and convolutional neural networks (CNN) (Liang *et al*., 2022). Animal based food such as fish, seafood, meat, eggs, honey, milk and milk products and plant-based foods like olive oil, cereals, pulses, oilseeds, spices, nuts, fruit juices, beer, wine, coffee and tea beverages are among the processed food products that rank highest on the food fraud/adulteration scale (Karabagias, 2024). Researchers from all over the world closely examine these products to ascertain their authenticity. Different ML models integrated with different computing techniques has gained significant attention due to its successful application in addressing global challenges of food frauds (Figure 1). However, although numerous research investigations have been carried out to check the efficiency of ML application in food authentication. Very few review articles have focused on the comprehensive compilation of both plant-based and animal-based food authentication via ML algorithms. Thus, the review was planned to comprehensively review and summarize the application of different ML algorithms in both plant and animal-based food authentication. This review attempts to summarize successful application of different ML models with diverse analytical techniques in food authentication in the global food sector by exploring current innovations. This review will serve as a value information and scientific data for food scientists/ researchers/policy makers who are working on various strategies for food authentication.

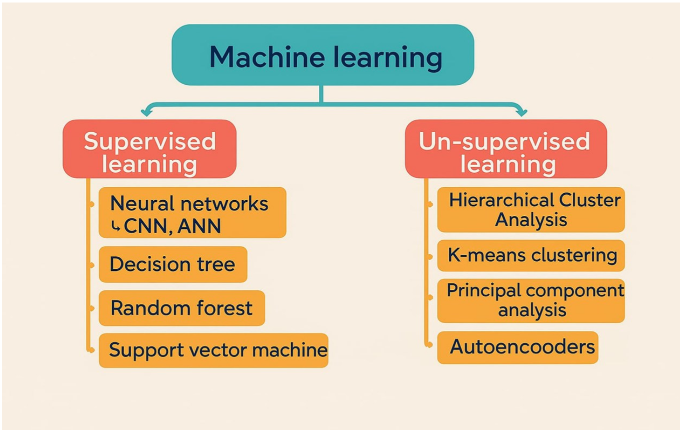

Figure 2. Types of Machine learning, a) Supervised learning (Neural networks (CNN, ANN), Decision tree, random forest and support vector machine) and b) Un-Supervised learning (Hierarchical cluster analysis, K-means clustering, principal component analysis and autoencoders).

2. Machine learning

The idea of machine learning (ML) involves generating and employing algorithms based on data (Meenakshi, 2020). Wu *et al*. (2022a) further defined that there are various methods for creating a model in a computer that gathers on its own and gets better with practice. This process is known as ML. ML has been utilized in agricultural fraud detection and verification of food items and producing excellent outcomes. Based on theoretical frameworks, machine learning approaches can be segmented into two groups: supervised learning (SL), and unsupervised learning (UL) (Figure 2).

2.1 Supervised learning

Supervised learning (SL) employs a labelled dataset to categorize; in other words, it trains algorithm to forecast the association of the input with output (P(Y|X)) using the independent variable (x-data) and the matching dependent (y-data) (Uddin *et al*., 2023). This kind of method is frequently employed to address regression and

https://doi.org/10.26656/lifr.1.e25050

categorization issues since it considers the precise understanding of the output target. The most common tasks carried out by SL techniques are classification tasks (Amini and Rahmani, 2023). For instance, steps where the model process the labelled data and predicts the category or group of new output include fraud detection, digit identification, disease diagnosis, face recognition, object identification, and image classification (Dargan *et al*., 2020). It examines specific characteristics in the test data and draws a significant inference.

For regression and categorization, Neural Networking (NN), Decision Trees, Random Forests, and Support Vector Machines (SVM) are the most widely used SL techniques. These techniques record varying degrees of precision and work exceptionally well for authenticating various food commodities. Regression is a technique used to illustrate the relationship between dependent and independent variables in the form of polynomial, logistical, and linear regression algorithms.

However, deep learning is an advanced subset of SL with robust, when combined with machine vision systems or portable devices, have shown promising results in various aspects of food authentication, including adulteration identification, variety identification, freshness detection, and food quality assessment (Shen *et al*., 2024a). The integration of deep learning with hyperspectral imaging (HSI) has further enhanced the potential for rapid food safety and quality evaluation. This is due the fact that spectral signatures of food substances are responsive to various factors such as water content, geographical origin, and harvesting time (Gul *et al*., 2024).

2.1.1 Neural networks

Neural networks (NNs) tools such as Convolutional neural networks (CNNs), Artificial Neural Networks (ANNs), K-nearest neighbour (K-NNs) excel at extracting relevant features from food images, enabling accurate detection and recognition of various food items (Tüzemen *et al*., 2023; Nazir *et al*., 2025).

2.1.1.1 CNNs

The deep architecture of CNNs allows for hierarchical feature learning, with deeper layers capturing more complicated and abstract features, indicating advanced classification performance (Alzubaidi *et al*., 2021). CNNs' versatility in data extraction, pattern analysis, item identification, process optimization, waste management, and pattern recognition has been demonstrated by their application in a variety of industries. CNNs have significantly changed the face of the food sector in an era of rapid technology advancement and increased focus on enhancing food safety, quality, and sustainability. Acquiring high-quality datasets, incorporating multimodal inputs for food quality evaluation, and satisfying the need for Graphics Processing Units (GPUs) to optimize deep learning models for food photographic examination are just a few of the hurdles faced by food companies (Nazir *et al*., 2025). In food image recognition tasks, CNNs have revealed superior accuracy compared to conventional ML approaches via handcrafted features (Nazir *et al*., 2025; Alzubaidi *et al*., 2021).

2.1.1.2 ANNs

ANNs model was created to replicate the functionality of the human brain, drawing inspiration from the operational principles of biological neurons (Tüzemen *et al*., 2023). This algorithm consists of interconnected neurons with weights, thresholds, and an activation function (Taki *et al*., 2023). A perceptron is a mathematical representation of a biological neuron, is structured with input, hidden, and output layers (Feng *et al*., 2020). As the number of neurons increases, both performance accuracy (Roshani *et al*., 2021) and computational demands (Farooq *et al*., 2021) rise. The categorization model has found applications in various domains owing to its high precision, capacity to handle intricate interactions, robustness, computerization, and simplicity (Dais *et al*., 2021; Houssein *et al*., 2022). ANNs have been extensively employed in studies focused on food authentication (Viejo *et al*., 2019; Amoghin *et al*., 2024; Sharma *et al*., 2024; Yang *et al*., 2024). However, ANN face several challenges, including the necessity to convert data into binary format, a significant risk of overfitting (Aggarwal *et al*., 2018; Menegatti *et al*., 2023), substantial computational requirements (Farooq *et al*., 2021), and a lack of transparency in decision-making processes, often known as the "black box" problem (Sarker, 2021).

2.1.1.3 K-nearest neighbour

K-nearest neighbor (K-NN) is a non-parametric method employed to categorize data points based on their resemblance to existing data (Wang *et al*., 2023). This classification approach assumes that similar data points are situated near each other in the feature space. It employs Euclidean distance to gauge proximity and assigns the most prevalent category (Wang *et al*., 2023). K-NN collects examples and classifies new instances based on similarities. It identifies k samples in the training dataset that are in proximity to the point being classified (Thomas *et al*., 2020). Selecting the ideal k value is essential to strike a balance between underfitting and overfitting. Excessively large or small k values may incorporate points from other classes or be susceptible to noise (Montesinos Lopez *et al*., 2022). The benefits of K-NN include high precision, low computation time, no learning cost, absence of optimization, and user-friendliness. However, large datasets increase computation time, diminishing its appeal for classification problems (Takahashi *et al*., 2023). K-NN is widely utilized in engineering and pattern identification. Its straightforward nature and precision have resulted in applications across various domains of Agriculture.

2.1.2 Decision trees

Decision tree is a prevalent SL method for classification. It is structured like a tree, comprising root, leaf, and internal nodes (Ge *et al*., 2023). Internal nodes represent feature tests, branches show test outcomes, and leaf nodes display final classification accuracy and class labels (Costa and Pedreira, 2023). The construction of a Decision Tree is an iterative process, breaking down data based on characteristics and dividing it into classes until reaching a specification (Neelakandan and Paulraj, 2020). Decision trees are becoming increasingly popular in cereal, pulses, oilseed and meat science (Barbon *et al*., 2018; Chu *et al*., 2024). Wakhid *et al*. (2020) employed a decision tree classifier for coffee classification, achieving 97% accuracy. Decision Tree Algorithms can handle various data types, including numeric, categorical, and ratings data, and can manage missing data. The algorithm's resemblance to human thinking makes it suitable for agricultural product classification. Decision Tree is extensively used in classification and prediction, with studies demonstrating >90% classification accuracy, showcasing its robustness (Mustapha *et al*., 2024).

2.1.3 Random forest

Random forest, an advanced version of a decision tree, constituting several trees for categorization and regression tasks. It uses decision trees with randomly chosen training data and features. The tree-growing procedure is repeated until achieving the highest accuracy (Faqe Ibrahim *et al*., 2023). During prediction, each tree forecasts its target class, and the classifier proposes the most accurately predicted class. Random forest employs the mean of decision trees with significant variance to establish a more consistent model, less prone to overfitting (Abdulhafedh, 2022). Reducing bootstrap samples increases randomness but may affect performance (Chen *et al*., 2023). Ideally, the bootstrap sample size should correspond the original training data set. Random forest corrects overfitting of a single decision tree (Pragasam *et al*., 2023). It collects uncorrelated decision trees, combining them to minimize the deviation and acquire precise predictions. The decision class is selected based on the majority of trees (Chaplot *et al*., 2023).

2.1.4 Support vector machine

Support vector machine (SVM) is a SL algorithm utilized for data classification and regression tasks. It is renowned for its exceptional classification capabilities, employing a hyperplane as a decision boundary to differentiate between data classes (Cervantes *et al*., 2020). The algorithm identifies the optimal decision boundary separating vectors from distinct categories, where vectors are numerical lists representing spatial coordinates (ElHaj *et al*., 2023). SVM determines the hyperplane that divides the data into two subspaces or categories with the best possible deviation (Jalal *et al*., 2022). The algorithm classifies data by establishing a task that divides data points into two groups, aiming to either minimize errors or maximize the margin between classes. A wider space adjacent to the dividing task leads to minimize the errors, as it allows for better distinction between labels (Sen *et al*., 2020). The increased margin and linear separation enhance classification accuracy. SVM has proven to be a powerful ML tool and has successfully application in different types of food commodities authentication (Richter *et al*., 2019; Feng *et al*., 2019; Song *et al*., 2020a; Chung *et al*., 2021; Ye *et al*., 2022).

2.2 Unsupervised learning

A branch of ML that explores the hidden patterns and structures in the set of unlabelled data is known as unsupervised learning (UL). UL relies on cluster analysis, lacks an extensive and accurately labelled dataset for the inputs, and lacks previous knowledge for grouping and structure (Liu and Lang, 2019). UL, in contrast SL, finds out to produce predictions (P(X)), identify underlying patterns, and forecast output based solely on unlabelled input X values (Uddin *et al*., 2023). In this case, the unlabelled input method is requested to search for undetectable characteristics and group the data according to how similar they are. In contrast to supervised learning, performance in unsupervised learning is typically domain-dependent and qualitative.

There are three categories of unsupervised learning problems: anomaly detection, dimensionality reduction, and clustering (Yazici *et al*., 2023).

Algorithms for anomaly detection look for occasional events or findings in a dataset that stand out from most of the data. They don't match the dataset's typical trends. Algorithms for detecting anomalies make the assumptions that there are significantly more normal cases than anomalies and that the abnormalities differ qualitatively from normal cases (Yazici *et al*., 2023). In order to simplify data and improve its interpretability, reduction of dimensionality projects an array of data onto a lower-dimensional space with little information loss. It has a smaller prediction error than the whole model and selects the most relevant attributes to train, resulting in a subset of the dataset (Yazici *et al*., 2023). Clustering is the process of arranging a group of examples that have never been categorized before. In turn, these instances are not assigned a class attribute, and they are grouped based on certain similarity criteria. Therefore, a similarity measure is utilized to calculate the inclusion of cases for clusters suggested by a few techniques, and cases are then allocated to their associated clusters based on the similarity measure (Yazici *et al*., 2023).

2.2.1 Hierarchical cluster analysis

Hierarchical cluster analysis (HCA) is a form of unsupervised ML that relies on either a divisive (top to bottom) or agglomerative (bottom to top) methods to build a hierarchy of clusters. By iteratively connecting clusters based on similarity and grouping criteria, HCA creates a graph structure called a dendrogram. The greatest similarity and dissimilarity are evaluated and determined by the HCA algorithm. This process is followed by a clustering procedure that may be based on distance, scale, sample, linkage method, or variable, both within and across classes (Leal *et al*., 2016).

2.2.2 k-means clustering

In unsupervised ML, the k-means clustering algorithm is a method that divides data into k groups based on high within-cluster and low between-cluster resemblance. It arranges information into predetermined clusters according to feature likeness. Each item is assigned to only single cluster, resulting in a specific amount of non-hierarchical, separate groups. The algorithm begins by randomly choosing centroids, denoted by k. Centroids represent the arithmetic mean of all points in a cluster. The Euclidean metric is used to calculate distances between points and centroids (Dubey *et al*., 2016). The process then gives each data point to the closest centroid cluster and updates centroids based on new assignments until it reaches convergence. K-means is efficient in terms of computation, simple to interpret, and works well with large datasets, making it a preferred choice for classification tasks. The main drawbacks of this method include its sensitivity to outliers and the requirement to indicate the number of clusters (k) beforehand.

2.2.3 Principal component analysis

Principal component analysis (PCA) is an unsupervised, non-linear statistical method that minimizes dimensionality and identifies orthogonal elements called Principal Components (PC). This technique decreases variables in the original dataset into PCs, containing most variable data, simplifying and recognizing significant data variations (Granato *et al*., 2018). PCA has enhanced ML classifiers' accuracy in classification tasks, recognized as a robust dimensionality reduction technique (Wan *et al*., 2021). Many studies have explored PCA in cereals, fruits, vegetables, oils, pulses and spices authentication (Richter *et al*., 2019; Yang *et al*., 2019; Lyu *et al*., 2020; Song *et al*., 2020b; Ye *et al*., 2022; Liu *et al*., 2024b; Xiong *et al*., 2024)

2.2.4 Autoencoders

Autoencoders offers the ability to learn complex data features without the need for labeled datasets. These neural network architectures can effectively reduce dimensionality and extract meaningful representations from high-dimensional food data (Berahmand *et al*., 2024). In the context of food authentication, autoencoders can be particularly useful for detecting anomalies or fraudulent products by learning the normal patterns of authentic food samples and identifying deviations from these patterns (Yong and Brintrup, 2022; Alam *et al*., 2023). Interestingly, autoencoders can be combined with other deep learning techniques to enhance their performance in food authentication tasks. For instance, convolutional neural networks (CNNs) can be integrated with autoencoders to leverage the power of both architectures for more effective feature extraction and anomaly detection in food-related data (Alam *et al*., 2023). Additionally, variational autoencoders (VAEs) offer a probabilistic approach to modeling food data distributions, potentially providing more robust representations for authentication purposes (Gomes *et al*., 2021; Yong and Brintrup, 2022).

3. Application of ML in food authentication

Numerous investigations of several ML applications in the food and agriculture sectors were executed (Mustapha *et al*., 2024; Shen *et al*., 2024; Bhat *et al*., 2024). Publication of review and research article has risen to vast extent due to the successful application

REVIEW

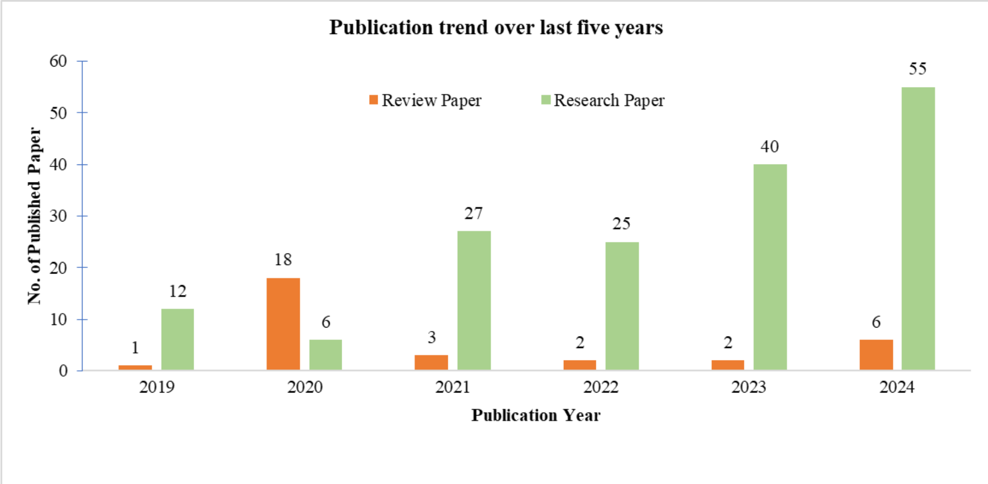

Figure 3. Trends of research and review article publications on machine learning in food authentication (Source: PubMed database).

of ML in food authentication. Figure 3 represents the number of research and review article publication over the period of 5 years from 2019 to 2024 and the data has been generated using PubMed as source. The implementation of ML in food adulteration identification and authentication is quick, economical, and highly accurate. Its programmed authentication process minimizes manual labor and human intervention. The model is simply deployable and scalable for numerous locations (Imdoukh *et al*., 2020). In many ways, ML is better than other conventional techniques (Choudhary and Sethi, 2023).

3.1 Fruits and vegetables

The produce available in markets may not be as pure as it appears, potentially containing harmful chemicals and pesticides. During growth and storage, some plant-based foods can naturally produce toxins such as cyanide, alkaloid, and aflatoxin. Additionally, certain manufacturers deliberately add dangerous substances like melamine, benzoic acid, and nitrite (Wu *et al*. 2019a; Xue *et al*. 2019; Goyal *et al*., 2022). In niche markets catering to health-conscious consumers seeking safe, nutritious, and eco-friendly products, organic food is an appealing option, which can lead to fraudulent practices. The quality of fruits and vegetables, which have high moisture content and are prone to damage and spoilage, can be compromised during rapid harvesting by hand or machine, presenting challenges for transportation and storage (Khatodiya and Malik, 2022). Extended storage periods can affect the sensory attributes and chemical composition of produce, including hardness (indicating maturity and sugar content), soluble solid content (influencing fruit maturity and flavor), organic acid levels, and pH, all of which reflect overall quality. These factors underscore the importance of swift authentication processes to ensure fruit and vegetable quality (Araujo *et al*., 2019; Kang *et al*., 2022a). Table 1 presents various research studies on fruit and vegetable authentication.

In recent years, Deep Learning and ML combined with image processing have emerged as powerful tools for assessing fruit and vegetable quality (Goyal *et al*., 2022). The first category of data utilizes a vision-based approach for quality assessment of parameters such as color, texture, size, shape, defects, and morphological characteristics. ML algorithms classify fruits and vegetables based on their external appearance. The second category examines food products' chemical composition using factors like moisture, pH, temperature, pressure, humidity, and viscosity. ML models are employed for classification and pattern recognition to identify adulterants in food and extract pixel intensity values to distinguish various image patterns (Goyal *et al*., 2022). Zhou *et al*. (2019) explored deep learning applications in food recognition, calorie estimation, quality evaluation of fruit and vegetables, food supply chain management, and contamination detection. The combination of ANN classification models with NIRS or HSI enables swift analysis of both external and internal fruit defects. IRS paired with ANN is commonly used to authenticate chemical compositions in fruits and vegetables (Bai *et al*., 2019). Feng *et al*. (2019) developed a technique to identify fine injures on winter jujubes from four geographical origins by integrating CNN with pixel-wise spectra from HSI at 874 -1734 nm. The CNN model achieved accuracies ranging from 90-100%, surpassing SVM and LR. Liu *et al*. (2019) created a two-branch CNN for classifying near-infrared HSI using spectral and spatial information, testing it on strawberry bruise detection. The proposed CNN model selected effective wavelengths and demonstrated superior performance compared to SPA in feature extraction. The two-branch CNN model surmounted SVM, GLCM-SVM, and one-dimensional CNN models in categorization precision for both whole spectra and effective wavelengths.

Centonze *et al*. (2019) employed non-targeted metabolomics utilizing HS-SPME/MS in conjunction

Table 1. Machine learning in authentication of fruits and vegetables.

Objective	Sample size	Analytical method	Method of feature extraction/data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Total soluble solid detection in 'Aiyuan 38' orange	117	VIS/NIR and E-nose	PLSR, MIT	PLSR, PCA-BPNN	• PCA-BPNN had better TSS detection ability than PLSR with an accuracy of 88.7%	Xu *et al*., 2019
Identify verities or ripeness of grapes for Protected. Designation of Origin wine	405	ATR-FTIR	OCWA	ANN	• ATR fast analysis avoided undesired structural changes • Pectin and polysaccharides were significant in variety and ripeness identification • polyphenols and fructose indicated ripeness	Murru *et al*., 2019
Geographically classified white asparagus	319	ICP-MS	PCA	SVM	• Accuracy of 91.2% was achieved when classified as per country • Relevant elements were lithium, cobalt, rubidium, strontium, uranium • Classification not affected by variety and harvest year	Richter *et al.,* 2019
Differentiated dried tangerine peels stored for different years	-	THz, Imaging	t-test	PCA-SVM	• Accuracy reached over 94% • Feasibility and superiority of THz	Yang *et al.,* 2019
Classification of organic and conventional tomato, bell pepper, onion, and lettuce	364	ICP-OES	F-score and chi-squared	PCA, LDA, SVM, ANN, RF	• LDA and SVM performed better • LDA reached 100% in discriminating against tomato and bell pepper • SVM outperformed others with accuracy of 100% in bell pepper and onion and 97% in tomato	Araujo *et al.,* 2019
Origin classification of RhizomaAtract ylodisMacroce phalae	224	HSI	GLCM, GLRLM	PLS-DA, SVM	• PLS-DA achieved highest accuracy of 97.30% • Unsatisfactory spectra selection by SPA • HSI with data fusion done rapid and nondestructive sorting	Ru *et al.,* 2019
Determination of green apple quality (soluble solids and acid)	100	NIRS	PCA	BPNN, GRNN and their combination	• Hybrid model showed great potential • Forecasting of total acid was accurate than soluble solids	Wu *et al*., 2019c
Polysaccharide and Ergosterol in *Antrodia camphorata* and Matsutake	-	NIRS	PCA	CNN, SPA-MLR, PLS, RBFNN, ResNet, ABRN	• ABRN performed better	Huang *et al*., 2019

REVIEW

Table 1. (cont') Machine learning in authentication of fruits and vegetables.

Objective	Sample size	Analytical method	Method of feature extraction/data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Classify hybrid seeds of okra and loofah individually	6136 and 4128	NIR HSI	-t-SNE	DCNN, PLS-DA, SVM	• DCNN had best stability and highest accuracy of >95%	Nie *et al.*, 2019
Geographical classification of apple	208	FT-NIR and soluble solid content	RFA	PLS, DCNN	• Geographical origin affected soluble solid detection accuracy • Multiple-origin model achieved robust and accurate results	Bai *et al.*, 2019
Details of food nutrients, carbs, proteins and fats	140	HSI	Multimodal architecture, error avoidance	DNN, multimodal DNN	• Degradation of accuracy in both systems caused by outliers • Proposed a new system with accuracy of 88.5% which overperformed both models	Ahn *et al.*, 2019
Discovered shelf-life indicators of *Crassostrea gigas* oyster	150	UHPLC-Q/ TOF MS and UHPLC-QQQ MS	PCA	OPLS-DA	• Identified 8526 peptides • 4 endogenous peptides, SSSTGEVGTYSGTTN, TARNEANVNI, VGIIKGSSSEEA, and TARNEANVNIY as potential markers	Chen *et al.*, 2020
Differentiated organic apples from conventional ones	150	Prototype sensor system	PCA	KPLS-DA, SIMCA, LR, PLS-DA, SVM, LS-SVM, kNN, C4.5, RF, LW-PLSC	• Proposed use of low cost sensor system rather than high-cost spectrometer • SVM and LW-PLSC effectively handled low quality images with accuracy of 93-100 %	Song *et al.*, 2020b
Classified conventionally and organically grown yams	122	Stable isotopes, ICP-MS, Spectrophotometer	PCA	kNN, SVM, Lasso, CART, OPLS-DA and RF	• Stable isotopes δD, $\delta^{18}O$ and 16 elements were higher in conventional • Higher contribution was of Mn, Cr, Se, Na, δD, As, δ15N • RF performed best with accuracy of 97.2%	Lyu *et al.*, 2020
Identification of variety and origin of walnut	192	FT-MIR	PCA	ELM, RF, BPNN, RBF	• Variety identification under the same origin performed the best • BPNN as best and RF as worst	Zhu and Xu, 2020
Identified origin of *Gentiana rigescens*	873	FTIR	PCA	PLS-DA	• Instead of roots, leaves could be a source of active chemicals	Liu *et al.*, 2020

REVIEW

Table 1. (cont') Machine learning in authentication of fruits and vegetables.

Objective	Sample size	Analytical method	Method of feature extraction/data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Predicted anthocyanins, flavonoids, phenolics in Dry goji berries	20	NIR-HSI	SPA, CARS, PCA, WT, DAE	PLS, LS-SVM	• Deep learning has a lot of interest in modeling and feature extraction	Zhang *et al.,* 2020
Tomato detection and mass estimation	77	Camera	-	CNN	• Detection and segmentation modules showed good performance in terms of accuracy and robustness with a mean detection accuracy of 99.02%, and precision of 99.7%	Lee *et al.,* 2020
Measured *E. coli* on fresh-cut potato	-	VIS-NIR	MSE	PLS, BPNN	• BPNN demonstrated improved robustness and accuracy (97.6%)	Li *et al.,* 2021a
Classified potato as per freshness and cultivar	135	Chromameter, VIS-NIR, Chlorophyll fluorescence	VIF, VIS, GA	PLSR	• Spectrum reflectance measurements on tuber skin showed best identification of 3 varieties with 93% accuracy	Kasampalis *et al.,* 2021
Predicted sugar content in potato (3 seasons, 2 cultivars)	1190	VIS-NIR, Glucose and Sucrose	SFS	kNN, PLS-DA, PLSR	• With HSI, accuracy scores for glucose and sucrose were as high as 95% and 80.1% Data of multiple growth seasons not as reliable as of one season.	Rady *et al.,* 2021
Predicted moisture in steamed, dried purple sweet potato	420	VIS-NIR, Composition	ROI	PLS-DA, PLSR	• PLSR (97.5%) was designed by combining Savitzky–Golay, multiplicative scatter correction, and the first derivative • Suggested to use 5 wavelengths rather than full spectrum	Heo *et al.,* 2021
Authenticated 3 fruit juices as per volatiles	47	HS-GC-MS-IMS	PCA	LDA, SVM, kNN	• LDA provided best results with accuracy of 96.9% for IMS data and 94.5% for MS data	Brendel *et al.,* 2021
Detected potato germination	296	MSI	-	SMTSM, CED	• Proposed method found 89.85% of germinations with 87.08% precision • CED could not detect boundary with slight differences in germination	Yang *et al.,* 2021
Identification of potato cultivars (Irga, Riviera and Colomba)	3	Images	Best First	WEKA 3.8.4	• Highest accuracies reached 99% for the IBk classifier and 98% for multilayer Perceptron	Ropelewska, 2022
Freshness and safety of fruits (apples, bananas, cucumber, lemon, orange, and tomatoes)	40,000	RGB images, (fresh, medium, and rotten),	-	CNN (ResNets and DenseNets)	• ResNets performed better in identifying fruit freshness • Better than other models used in literature • suggested to use data augmentation technique and generative adversarial networks to solve data imbalance	Han *et al.,* 2022a

Table 1. (cont') Machine learning in authentication of fruits and vegetables.

Objective	Sample size	Analytical method	Method of feature extraction/data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Identified grape juice adulteration in pineapple, orange, and apple juices	-	FTIR	PCA, HCA	PLS, LASSO, SVM, RF, LDA	• LDA and SVM performed well (100%) • FT-IR spectra mainly influenced by type of fruit, lesser by brand	Calle *et al.,* 2022
Detection of pesticide residue Level in Grape	1071	Vis-NIR- and NIR- HSI	PCA	LR, SVM, RF, CNN, ResNet	• For Vis-NIR spectra, the best model was ResNet, with the accuracy of over 93% • For NIR spectra, LR was the best, with the accuracy of over 97%, but SVM, CNN, and ResN*et al*so showed closed and fine results	Ye *et al.,* 2022
Separated organic tomatoes from non-organic ones	160	MS	m/z alignment	DT	• attained 92% accuracy in classifying fruit juices according to their respective groups • 12 biomarkers linked to non-organic and 4 to organic	Oliveira *et al.,* 2023
Geo origin of Chinese yam	-	H NMR	LBP	kNN, DT, SVM	• better than original NMR results • SVM was best with 100% accuracy	Hu *et al.,* 2023
Geographical authenticity of *Yimucao*	63	Stable isotopes, % C and N, extracts	PCA, OPLS-DA	RF, SVM, adaptive boosting and neural network	• RF performed the best (98.4%) • $\delta^{18}O$, $\delta^{2}H$, and $\delta^{15}N$ were potential markers • latitude, sunshine duration, and RH were key factors	Liu *et al.,* 2024b
Geographical verification of *Pleuropteru smultiflorus* th unb.	357	4 Stable isotopes, 16 functional compounds, 40 elements	PCA	LDA, RF, SVM, kNN	• 94.4-100 • 30 potential factors were found, including Cu, 12 rare earth elements, δ2H, and 16 functional chemicals • altitude, high temperature, and moisture were key drivers	Xiong *et al.,* 2024
Composition and wine quality of grapes	369	Vis-NIR, Raman Spectra	PCA	LS, PLS, SVM, GPR, NN, RFR, MPLS	• GPR and SVM performed well in analyzing pH	Ebrahimi *et al.,* 2024
Predicted sensory traits in sweet potato	207	NIR	-	L-SVM, PCR, PLS, RF, XGBoost, ENR	• Intensity of orange color best predicted by all models • PLS was best • ENR, XGBoost and RF performed worse	Nantongo *et al.,* 2024

REVIEW

Table 1. (cont') Machine learning in authentication of fruits and vegetables.

Objective	Sample size	Analytical method	Method of feature extraction/data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Discriminated cultivars and adulteration in maca powder	108	NIRS	PCA, SG	LDA, PLSR	• When only pure yellow, red and black Maca were discriminated against, 94.1% accuracy was seen • 65.5% accuracy was observed in discriminating adulterated cultivars	Zaukuu *et al.,* 2024
Identified origin and shape of *Lanxangia tsaoko*	-	FT-NIR	MCC, PCA, SPA, CARS	PLS-DA, SVM, ResNet	• PLS-DA and SVM achieved 100% accuracy in classifying origin • ResNet classified origin and shape with 100% accuracy without need of preprocess and feature extraction	He *et al.,* 2024
Estimated polyphenol oxidase and peroxidase enzyme activity in bell pepper	90	Vis/NIR	PLSR	SVM, ANN, MLR, GA, PSO, ACO, ICA	• SVM-PSO was the best model to find the most appropriate wavelength • ANN-as best method for wavelength selection	Amoghin *et al.,* 2024
Quality of Persian pomegranate	-	Vis-NIR and others	SG, SNV	PLSR	• Percentage of fruit juice and pH were key predictors Vis/NIR-PLSR achieved 95% accuracy	Hemmati *et al.,* 2024
Classified raw tomatoes as per lycopene	150	HPLC, Vis-SWNIR	PCA	LR, LDA, RF, ANN, SVM	• lycopene had a positive correlation with 760 nm and a negative correlation with 485, 560, and 585 nm • 500-750 nm dominated classification • ANN-95%, SVM-90%, RF-80%	Sharma *et al.,* 2024
Detected pesticides in spinach	-	RS	-	SERSFormer	• Raman signals were significantly enhanced by gold-silver core-shell nanoparticles • models categorized five different pesticides with 98% accuracy.	Hajikhani *et al.,* 2024

with an electronic nose to identify volatile components in oranges (*Citrus sinensis* L. Osbeck) from Italy, South Africa, and Spain. The researchers utilized various multivariate statistical models, including PCA/LDA, SELECT/LDA, and PLS-DA, to process volatile compound data. SELECT/LDA yielded the highest prediction accuracy in both cross-validation and external validation (97.8% and 95.7%, respectively). Vavoura *et al*. (2022) explored the differentiation of fresh, unprocessed orange juice based on botanical origin (7 varieties) using physicochemical parameters and instrumental analyses. These methods encompassed flavonoid detection via HPLC–DAD and volatile compound identification through HS-SPME/GC–MS. MANOVA and LDA revealed that orange juice variety significantly influenced analytical parameters, achieving an 89.3% categorization rate in cross-validation. Amino acids, crucial metabolites for distinguishing between plants of the same fruit, were examined in fruits using NMR and investigated with chemometric tools. Botoran *et al*. (2019) identified ten amino acids employing PCA and LDA chemometrics that assisted in differentiation and classifying various juice varieties from different plant sources. Tsouvaltzis *et al*. (2020) assessed chilling injury in eggplant fruit by combining visible and Near-Infrared (NIR) HSI with classification techniques such as PLS-DA, SVM, and KNN to categorize eggplant by storage temperature. Babellahi *et al*. (2020) showed that HSI with PLS-DA could distinguish between cold-stored green peppers affected by chilling injury and fresh ones.

3.2 Beverages

Authenticating beverages is a complex field that requires expertise in various areas, including instrumentation, agriculture, food technology, and statistics (Kamiloglu, 2019). Table 2 presents numerous studies focused on authenticating different alcoholic and non-alcoholic drinks. High-quality fruit juices and nectars, especially those freshly squeezed from expensive fruits, are often targets for adulteration. Common methods include dilution with water or cheaper fruit juices, and the addition of sweeteners like sugar syrups, acids, or coloring agents (Esteki *et al*., 2018). In the realm of wine and other alcoholic beverages such as rum, vodka, brandy, and liquor, a primary form of fraud involves substituting ethanol with less expensive methanol. Wine's diverse characteristics, from grape variety to production techniques, provide numerous opportunities for authentication (Dewan *et al.*, 2025).

Even non-alcoholic beverages like coffee, tea and cocoa are susceptible to adulteration. Ground and roasted coffee may be mixed with extraneous additives such as brown sugar, chicory, corn, barley, date seed powder, tamarind seed powder, wheat, soybean, or rye to reduce costs. These additions can alter coffee's sensory characteristics, including acidity, bitterness, along with flavor and aroma profile (Dewan *et al.,* 2024a). Moreover, they may trigger allergic reactions in consumers sensitive to certain gluten-containing cereals used in the adulteration process (Dewan *et al.,* 2024b). Researchers are increasingly turning to ML methods to

Table 2. Machine learning for authenticity of beverages.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/Algorithm for classification/ discrimination	Observation and accuracy	Reference
Identified volatiles in beer	72	SPME-GC-MSD	PCA	ANN	• No hops-derived volatiles in bottom-fermented beers • 4-Ethyguaiacol and trans-β-ionone positively while styrene negatively affected sensory profile • ML achieved 98% accuracy	Viejo *et al.,* 2019
Discriminated Brazilian lager beer	40	pH, ^{1}H NMR	PCA	PLS-DA, SIMCA	• Different style beers could be distinguished using PLS-DA • SIMCA also had accuracy of more than 90%	Silva *et al.,* 2019
Classified clonal varieties of green coffee	15	Raman spectroscopy	MC, MSC	LDA, MDA, QDA, RDA, PLS-DA, SIMCA	• MSC was better than MC • MSC-LDA-98.7% • MDA, RDA, QDA, PLS–DA, SIMCA-100%	Luna *et al.,* 2019

REVIEW

Table 2. (cont') Machine learning for authenticity of beverages.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/Algorithm for classification/ discrimination	Observation and accuracy	Reference
Classified craft beers (IPA, Lager and Weiss)	3	HPLC-DAD-MS/MS	-	LDA	• Derivatives of caffeic and coumaric acids distinguished craft beers • 3-Caffeoylquinic acid, 3-*p*-coumaroylquinic acid, 4-*p*-coumaroylquinic acid, 5-caffeoylquinic acid, coumaric acid, kaempferol-3-O-rutinoside, proanthocyanidin B dimer III and proanthocyanidin B dimer V were food predictors	Cheiran *et al*., 2019
Identified beer olfactory details	5	E-nose	-	CNN, SVM	• CNN-SVM achieved accuracy of 96.7%	Shi *et al*., 2019
Classified liquor samples	5	Chemiresistive sensor	PCA	RF, kNN	• RF performed better with an accuracy of 738%	Schroeder *et al.*, 2019
Authentication of geographical origin of wine	540	FT-MIR, FT-NIR	PCA	SIMCA, DA	• DA categorized 86%, 85%, and 77% of Australian, Chilean, and Chinese Cabernet Sauvignon, while SIMCA classified 97%, 97%, and 92%	Hu *et al.*, 2019
Assessed beer quality	20	E-nose, NIR, RoboBEER	PCA	ANN	• accuracy of 97% in terms of fermentation type • high precision in physicochemical and colorimetric analysis, as well as customer preference and acceptance	Viejo and Fuentes, 2020
Geographical and varietal authentication of wine	221	A-TEEM, HPLC	PCA	XGBDA	• 100% in varietal and 99.7% in geographical classification	Ranaweera *et al.*, 2021
Classified red wines of 4 countries	83	LC-MS, ORAC and phenols	PCA, CFS, RFI	SVM	• attained accuracy of 93.97%	Costa *et al.,* 2021
Identified quality index for coffee beans	3	Color using spectrophotometer	MaxPooling layers	CNN	• difficulty in distinguishing under-roasted and properly roasted coffee • successfully distinguished over-roasted	Przybył *et al.,* 2023
Authentication of origin of grape wine	300	UPLC-Q-TOF-MS	PCA	SVM	• indole, sulfacetamide and caffeine were potential predictors of diff origins	Ta *et al.,* 2023
Origin, variety and processing traceability of black tea	219	^{1}H NMR	PCA	LDA, KNN, SVM, RF	• RF classified tea with an accuracy of 92.7% • key molecular markers of origin were caffeine, malic acid, lysine and β-glucose.	Cui *et al.,* 2023

Table 2. (cont') Machine learning for authenticity of beverages.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/Algorithm for classification/ discrimination	Observation and accuracy	Reference
Origin traceability of white tea	579	NIRS	PCA, LDA, SPA, WT, MSC, SNV	kNN, RF, SVM (180 models)	• LDA best for feature extraction • DPC-CWT-LDA-KNN, DDFP-OS-LDA-KNN and AFWT-OS-LDA-KNN performed better with 88.97%, 93.88% and 97.96% accuracy, respectively	Zhang *et al.*, 2023
Geographical origin of red wines	45	UV-VIS	PCA	OPLS-DA, SVM	• Both models achieved 100% accuracy	Gu *et al.*, 2023
Authenticity of lime juice and its geographic origin	2+	AIJN COP	FFS	SVM, kNN, RF, DT	• SVM with 90%, kNN with 94%, and RF and DT with 100% accuracy • Copper, zinc and iron were vital features	Roozbahani *et al.,* 2024
Discriminated black tea geographically	360	FTIR, NIR	PCA	kNN, SVM, LDA, RF	• kNN and SVM improved GI identification attaining 100% accuracy using FTIR	Li *et al.*, 2024a
Authenticated tea origin	-	ICP-MS	PCA	SVM, ANN	• Both models classified tea samples accurately (100%)	Yang *et al.*, 2024
Authenticated geographical origin of green coffee beans	528	HSI-NIR	SNV, MSC,	PLS-DA,SVM, RBF-SVM, RF	• Feature selection done at regional, national, and continental level • Nonlinear SVM performed better • SVM, highest complexity at regional level	Sim *et al.*, 2024
Predicting and improving complex beer flavor	250	Chemical and sensory properties	-	LR, Lasso, PLSR, ABR, ET, GBR, RF, XGBR, SVR, ANN	• Best-performer, Gradient Boosting, yielded models that outperformed predictions • Tree-based models over -performed linear models	Schreurs *et al.,* 2024
Sub-Regional classification of Barossa wine aged for several years	217	A-TEEM	PCA	XGBDA with PLS	• 100% classification accuracy for vintage year and 98.8% for unknown sample	Wang *et al.*, 2024c
Identified adulterants (5) in cocoa powder	2+	Vis-NIRS	PCA, HCA	RF, SVM, PLS, LASSO, RF, Ridge, ENR	• Both classified with 100% accuracy • PLS combined with Boruta performed best in quantifying	Millatina *et al.,* 2024
Integrated liquor alcohol content prediction and sensory	34	Raman spectra	PCA	SVR	• Expanding alcohol level range improved performance • Alcohol level (±0.15% v/v) precisely measured by quantitative PCA-SVR • 94% of sensory-disqualified samples correctly recognized by qualitative PCA-SVR	Liu *et al.*, 2024c

REVIEW

Table 2. (cont') Machine learning for authenticity of beverages.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/Algorithm for classification/ discrimination	Observation and accuracy	Reference
Geographical authenticity of wine	153	ICP-MS, ICP-OES		ANN, SVM, DA, DT, RF, kNN	• Al, Ba, Ca and Rb in wines depended on variety of grapes • Al, Ba, Rb, Fe, Li, Sr depended on region • Sr, Li and Fe- key predictors • ANN-100%, SVM-98.7%, DA-94.8	Temerdashev *et al.,* 2024
Discriminated quality and quantity of volatiles in beer	9	GC-MS, FTIR	PCA	PLS	• PCA discriminated beer type • Ethyl caproate, ethyl caprylate, and phenylethyl alcohol-key volatiles discriminating different beers	Gao *et al.,* 2024
Monitoring of brewing process of Qingke beer	-	Raman spectra, NIR, UV-Vis	-	NN, PLS, LS-SVM	• PLS-UV-Vis and NN-Raman spectra found suitable for quantifying total phenols • NN-Raman spectra and NN-NIR performed better in identifying reducing sugar in mashing and boiling	Zhou *et al.,* 2024b
Monitored fermentation (fungal) Fuzhuan brick tea	90	E-nose, Spectra, sensory, physicochemical	SPA	MLR	• Achieved 100% accuracy • FBT may be identified at different stages of fermentation using fusion spectra and e-nose	Hu *et al.,* 2024
Identified synthetic colorants in black tea	79	IMS	PCA	PLS-DA, PLSR	• Achieved 100% accuracy on identification of six different colors added at different concentrations • Effective qualitative and quantitative discrimination of adulterated tea	Sobhaninia *et al.,* 2024
Geographical authentication of Chinese wines	90	ICP-MS, HS-SPME-GC–MS, UHPLC-Q-Exactive Orbitrap MS	XCMS, PCA, OPLS-DA	RF, SVM, kNN	• 5 minerals, 13 volatiles and 51 metabolites were found as key factors • Fused datasets based on feature variables, combining mineral elements, volatile components, and metabolites, achieved 100% accuracy across all three algorithms	Chen *et al.,* 2024

https://doi.org/10.26656/lifr.1.e25050

Table 2. (cont') Machine learning for authenticity of beverages.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/Algorithm for classification/ discrimination	Observation and accuracy	Reference
Traced geographical origin and taste of black tea	20+	UV-Vis, colorimetric sensor array	SG, PSO	SVM	• Model SG-PSO-SVM exhibited accuracy of 99.5% • Accurately quantified taste presenters-theaflavin, caffeine, vitexin-2-*O*-rhamnoside, rutin, epigallocatechin gallate, epicatechin gallate, gallic acid	Shen *et al.,* 2024b

distinguish and authenticate beverages, as well as to identify unwanted components (Kamiloglu, 2019). Karabagias *et al*. (2019) conducted research to distinguish prickly pear juice from shelf-grown varieties in Greece. They analyzed minerals employing ICP-OES and volatile components via an optimized HS-SPME/GC–MS method. MANOVA revealed significant variations ($p< 0.05$) in the parameters investigated related to the juice's geographical origin. LDA classified prickly pear juice samples based on geographical origin with 85.7% accuracy using 7 minerals and 88.9% accuracy using 21 volatile compounds. Research by Kaczala *et al*. (2024) utilized smartphone image data to develop models for distinguishing between whole juice (WJ) and nectar/ reconstituted juice (NERJ) in apple beverages, as well as predicting the apple juice content proportion. Their classification models, employing kNN and XGBoost algorithms, successfully identified NERJ (91%) and WJ (87%) samples, showcasing their potential for apple juice authentication. Predictive models using XGBoost and CatBoost algorithms estimated apple juice content (%) in each category with over 96.2% accuracy. These approaches proved to be eco-friendly, non-invasive, economical, and quicker than conventional quality control and authentication methods, indicating their value in industrial and regulatory contexts. Such techniques could become crucial for quality assurance and preventing fraudulent practices in apple juice products. In a separate study, Karabagias *et al*. (2021) examined 45 wine samples from eight varieties of dry and demi-sec white wines using HSSPME/GC–MS, applying MANOVA, LDA, FA, and PLS to the data. They also investigated water adulteration and identified potential chemical markers for authenticating grape variety, geographical origin, or water adulteration.

Baijiu, a distinctive Chinese spirit, has grown in popularity due to its unique flavor profile. Zhou *et al*. (2024a) integrated fluorescence hyperspectral technology with machine learning for swift, non-destructive adulteration detection in bulk liquor in Chengdu. They collected fluorescence hyperspectral data from original liquor, water-adulterated liquor, and industrial alcohol-adulterated liquor. The raw data underwent various preprocessing techniques, with PCA and MDS used for spectral data feature extraction. Three models (BP-ANN, SVM, and 1D-CNN) were evaluated. The study revealed that combining fluorescence hyperspectral technology and machine learning, specifically the MSC-PCA-SVM method, offers a novel approach for bulk liquor detection, achieving 100% classification accuracy, recall, precision, and F1-score. These findings have implications for liquor detection and may serve as a reference for adulteration detection across the food industry, potentially improving market regulation and consumer protection.

Jiménez-Carvelo *et al*. (2019) categorized common computer science algorithms as substitute data mining/ ML techniques (SVM, ANN, CART, J48, C5.0, Random Forest). Ruvalcaba *et al*. (2019) employed LDA to categorize four beer styles based on their aromatic profiles, successfully classifying all 30 samples. Giannetti *et al*. (2019) utilized PLS-DA to differentiate Pilsner-style Lager craft beers from industrialized ones, identifying 13 key volatiles using VIP scores, achieving 96.3% accuracy. Fukui *et al*. (2019) distinguished low-malt fresh beers from those stored at −5°C and 50°C for 2 weeks. PCA demonstrated clear separation, while OPLS-DA accurately classified samples stored at different temperatures. Coelho *et al*. (2019) applied PCA to examine barrel aging procedures, revealing patterns among volatile profiles of beers aged in barrels or stainless-steel vats. Karabagias and Badeka (2021) examined herbal teas using physicochemical and volatile compound analyses to identify chemical markers characterizing each product by brand name, utilizing MANOVA and LDA. Significant differences ($p < 0.05$) were observed, and LDA achieved perfect classification (100%) by brand name. Gottstein *et al*. (2024) investigated 603 roasted arabica coffee samples utilizing 1H NMR fingerprinting and PCA-LDA to distinguish samples by geographical origin and cultivation process. Samples were segregated by origin but not by organic or conventional production. In summary, ML models can effectively detect and maintain beverage quality to ensure product authenticity.

3.3 Cereals, pulses, oilseeds and their products

Verifying the exact geographical origin, variety, and organic characteristics of cereals, pulses, and oilseeds is necessary, in addition to detecting adulterants (Karabagias, 2024; Lozano-Castellón *et al*., 2024; Tomar *et al.,* 2025). Table 3 summarized the investigation focused on authenticating different cereals, pulses, oilseeds and their processed products. Among diverse varieties of cereals, wheat is the most popular staple crop consumed around the world and utilized in manufacturing different baked products (Dewan *et al*., 2022). Thus, it is prone to adulteration in terms of variety, origin as well as its organic characteristics. Jin *et al*. (2023) utilized gas chromatography combined with ion mobility spectrometry (GC-IMS) to examine the volatile profile of cooked wheat grains in three colors: black, green, and yellow. The primary volatile compounds identified were aldehydes, ketones, and furans. The researchers employed PCA and CA to differentiate the three cooked wheat grains and developed a discrimination model using OPLS-DA. A comparable approach was used to analyze 132 wheat samples from various German regions to verify authenticity regarding cultivation technique and origin. Among the isotopes studied (O, H, C, N, and S), only δ13C showed statistically significant differences between organic and conventional wheat samples (Gatzert *et al*., 2021). Izquierdo *et al*. (2020) created an imaging model to distinguish five rice types (*Oryza sativa* L.) and their flours using photographs taken with a standard camera. The images, processed using optimized CNN, proved useful for identifying and classifying different rice types or flours. Ch *et al*. (2021) employed HS/GC-MS to characterize volatile compounds in rice samples from different geographical origins. They developed multivariate analysis models such as PLS-DA, which demonstrated good discrimination rates for rice samples based on geographical origin. Data-Driven Soft Independent Modelling of Class Analogy (DD-SIMCA) and k-NN exhibited perfect specificity and accuracy (both 100%) in identifying rice samples' geographical origin. A method utilizing compound-specific δ13C and δ15N analysis of fatty acids and amino acids, combined with OPLS-DA, was developed to differentiate organic, pesticide-free, and conventional rice samples. OPLS-DA models emphasized the importance of δ13C values of tyrosine, isoleucine, and alanine in recognizing organic from pesticide-free rice. Another study examining δ13C and δ15N variations in environment-friendly and conventional rice evaluated discriminant models for organic rice authentication. SVM analysis provided 4.4–14.6% better overall predictability of rice types compared to DA and effectively distinguished organic (95.9%) and conventional (93.6%) from pesticide-free rice (Chung *et al*., 2019; Chung *et al*., 2021). Research has shown that a grouping of stable isotope analysis (δ13C and δ15N) and elemental profiling (27 elements) employing various statistical methods successfully segregated organically grown rice cultivars fertilized with animal manures from conventional rice grown with green compost or synthetic fertilizers. However, relying solely on δ13C and δ15N values was not sufficient to accurately differentiate organic rice from green and conventional rice. Notably, maximum levels of K and Ca were observed in green and conventional rice due to synthetic fertilizer use. The integration of isotopic and elemental signatures with partial LDA modeling achieved an impressive 100% accuracy in classifying organic rice (Liu *et al*., 2020b). Additionally, NIR spectroscopy within a specific absorption range, coupled with multivariate PCA and PLS regression, classify into

Table 3. Machine learning for authentication of cereals, pulses, oilseeds and their products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Identified cultivar of oat	14846	HSI	PCA,	DCNN, LR, SVM	• all DCNN-based models outperformed the contrast models • DCNN trained in an end-to-end manner achieved highest accuracy of 99.19%	Wu *et al*., 2019b
Quality inspection of colored rice	-	NIRS	-	SVM	• accuracy for broken kernels, chalkiness, and damaged and spotted areas was 99.3%, 96.3% and 93.6%, respectively • running time was 0.15 s, with four types of defects detected at one-time	Chen *et al*., 2019a

Table 3. (cont') Machine learning for authentication of cereals, pulses, oilseeds and their products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Traceability of cotton seed variety	7+	HSI	PCA, CNN, ResNet	CNN, PLS-DA, LR, SVM	• CNN-based self-design models outperformed ResNet-based models • Classification model using full spectra performed better than those using effective wavelengths	Zhu *et al.,* 2019a
Authentication of soybean variety	3+	HSI	PCA	DCNN	• Discriminated with a few samples using pixel-wise spectra rather than large scale of samples using average spectra • CNN dealt with large number of pixel-wise spectra	Zhu *et al*., 2019b
Predicted origin of lentil	42	FT-NIR	MSC, SNV, DT	PCA, MPLS, PLS	• NIRS-DPLS exhibited accuracy of 95%	Revilla *et al.,* 2019
Forecasting of rice yield	-	Climate change	-	MLR, BTR, ANN	• ANN outperformed multiple linear regression and boosted tree. • Predicted rice yield based on climate	Zhang *et al*., 2019
Classified maize cultivar	760	FT-NIR	PCA, SG, MSCM GA	KNN, SIMCA, PLS-DA, SVM	• Full range wavelengths accurately categorized two cultivars with 99.6% accuracy • SVM-DA outperformed other models, and all algorithms showed high classification accuracy (97.56-99.59 %)	Qiu *et al.,* 2019
Wheat variety	5	HSI	PCA, LDA, SNV, MSC, WT, SPA	LDA, SVM, ELM	• Highest accuracy of 91.3%, was obtained by ELM based on full wavelengths	Bao *et al.,* 2019
Screened food markers of chia, linseed, and sesame	-	GC-MS	-	RF	• Potential markers were tyrosol or myo-inositol • Recommended hyperparameter tuning of RF model	Erban *et al.,* 2019
Discriminated silage and common maize seeds	40800	NIR-HSI	PCA	RBFNN, SVM	• SVM and RBFNN exhibited accuracies of 88.2% and 88.4%, respectively in classifying silage and common • To classify variety of common maize-SVM: 97.2 and RBFNN: 98.1 %	Bai *et al.,* 2020
Origin traceability of millet	480	Vis-NIR	MSC, DC, MC, SNV, FD, SD	kNN, LDA, LR, RF, SVM	• SVM and RF slightly overperformed other models (99.5%) • F-score values were: kNN- 99.1% and LR at 98.8%	Kabir *et al.,* 2021
Traceability of origin of Chinese GI rice	131	ICP-MS	PCA	RF, SVM	• Both models showed 100% accuracy • Cd was the best predictor to trace origin of rice	Xu *et al.,* 2021

REVIEW

Table 3. (cont') Machine learning for authentication of cereals, pulses, oilseeds and their products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Estimated shelf life of Chia seeds	3	NIR-HSI, FFA, FA, AV	SNV	Multivariate accelerated shelf life testing modelling	• Within same category (da y x temp), all samples wer e accurately predicted • Chia seeds stored at 25, 35, and 45 °C were estimated to have a shelf life of 1,300, 798, and 90 days, respectively.	Cruz-Tirado *et al.,* 2021
Detection of Aflatoxin-B1 in maize	240	Vis-NIR HSI	PCA	PLS-DA, kNN	• Maximum classification accuracy of 94.7% was obtained using PLS-DA with SNV pre-processed data • The best efficiency (98.2%) was exhibited by k-NN model with raw data	Chakraborty *et al.,* 2021
Identified origin of rice	354	Stable isotope, elemental fingerprints	-	Explainable deep learning	• Explainable deep learning system attained 97% accuracy in origin tracing • Rb, B, Li, Cd, and Mn as key predictors • Surpassing decision tree and SVM	Chu *et al.,* 2024
Traceability of origin of grain maize	200	DART-MS	PCA	RF	• Accuracy of 84.4% was attained • RF identified crucial markers which are classes of mono-, di- and triacylglycerols	Schmauder *et al.,* 2024
Geographical traceability of soybean	6	E-nose	AKCA	CNN	• Developed AKCA-Net to process soybean gas information with reduced complexity	Sun *et al.,* 2024a
Identified geographical origin of peanut	161	Raman spectra	ANOVA	PCA, kNN, SLDA, SVM	• kNN showed highest accuracy	Sun *et al.,* 2024b
Identified sulphite treated sprouted beans	168	FTIR	MCC	RF, CNN, RBFNN	• Models' accuracy were as follows: RF (82%) > CNN (91%) > RBFNN (95%) • MCC was most effective preprocessing	Li *et al.,* 2024b
Classified fungal contamination in rice grains	5	HSI, GC-MS	PCA	DFA, SVM	• Pentamethyl-heptane, decane, dodecane, 3-octanone, and 1-octen-3-ol observed in infected grain • HSI-SVM with 93.4% accuracy was found best	Siripatrawan and Makino, 2024
Classified Iranian wheat flour varieties	200	FT-MIR	D2, SNV	LDA, SVM, ANN	• LDA, SVM and ANN exhibited best results with 100% accuracy • D2 + SNV preprocessing approach produced best results and with all supervised models	Fattahi *et al.,* 2024a
Traceability of rice mildew	3	NIRS	MSC. MMN, SNV, SG, SG-FD	PLSR, SVR, BPNN	• Optimized models, SPXY -SG-BPNN, KS-SG-BPNN and SPXY-SG-BPNN showed good accuracy • recommended the use of BPNN	Wang *et al.,* 2024a

Table 3. (cont') Machine learning for authentication of cereals, pulses, oilseeds and their products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Authenticated Iranian rice varieties	70	FTIR, Vis/ NIR	PCA	SVM, QDA, DT, MDA, MLP	• FTIR-SVM achieved 100% accuracy • PCA-FTIR- 99% and PCA-Vis/NIR-100%	Zaresani *et al.,* 2024
Discriminated apricot kernel in ground almond	120	FT-NIR	-	SIMCA, CE, PLSR,	• Accuracy of 100% was achieved • PLSR performed the best	Menevseoglu *et al.,* 2024
Cookie composition traceability-	120	FT-NIR and nutrition	MSC, SNV	PLS-DA, PLS-kNN, PLS-NB	• Identified cookie type and type of cereal present in cookie • Data preprocessing did not lead to significant improvement • PLS-kNN as best (90%)	Quintelas *et al.,* 2024
Aroma quality of Pixian broad bean paste fermentation	-	E-nose	-	PLSR, PLS-DA, RF, SVM, ANN	• PLS-DA achieved an accuracy of 99% in identifying fermentation stage • ANN overperformed other models in quantifying aroma compounds	Xu *et al.,*2024

organic and conventional rice samples with 87.5% rediction accuracy (Xiao *et al*., 2019). Brigante *et al*. (2020) employed targeted metabolomics using HPLC-DADESI-qTOF (MS/MS) to detect polyphenols in chia, flax, and sesame seeds as potential authenticity markers for raw materials and bakery products. The study identified 44 polyphenols as potential biomarkers for seed differentiation. Various statistical techniques were applied, resulting in satisfactory identification of the distinct seeds studied based on selected polyphenols. Longobardi *et al*. (2017) assessed lentil samples from Italy and Canada using untargeted 1H NMR fingerprinting combined with ML. Multiple multivariate statistical techniques were applied to the NMR data, with the PCA-LDA model yielding the best results. Campmajo *et al*. (2019) utilized non-targeted HPLC–UV analysis to distinguish and classify various nuts and seeds using multivariate statistics. PLS-DA achieved 100% correct categorization of nut samples based on its type and processing. von Wuthenau *et al*. (2022) examined a large number of almond samples from various botanical and geographical origins over four consecutive harvesting years. The study used ICPMS for minerals determination and chemometrics analysis. According to the researchers, employing SVM on certain minerals allowed for the reliable identification of almonds' geographical origin with high prediction rates.

3.4 Spices and edible oils

Significant adulteration of spices involves incorporating cheaper alternatives like almonds, peanuts, tree nuts, and other seeds into cumin powder, or using papaya seeds in black pepper due to their similar appearance and low cost. Saffron, the most expensive spice, is often adulterated with synthetic additives such as glycerin, barium sulfate, borax, sandalwood dust, or tartrazine (Grace, 2019). Another form of fraud is contaminating oregano with impurities from other plant sources, including olive, myrtle, sumac, or hazelnut leaves. These inclusions not only constitute fraud but may pose health risks to consumers, particularly if pesticide residues are present (Drabova *et al*., 2019). Additionally, starch has been identified as an adulterant in condiments like garlic, ginger, and powdered onion (de Lima *et al*., 2020). Garlic powder is especially susceptible to adulteration with starch due to similar coloration. Talc and chalk powders have also been stated as adulterants (Galvin-King *et al*., 2021). The diverse chemical composition of spices and aromatic herbs allows for identifying specific biomarkers useful for authentication. Pages-Rebull *et al*. (2023) established an HPLC method coupled with UV-VIS to characterize, identify, and authenticate cinnamon, oregano, thyme, sesame, bay leaf, clove, cumin, and vanilla. This method was established on determining sesamol, eugenol, thymol, carvacrol, salicylaldehyde, and vanillin integration with chemometrics. The researchers initially applied PCA and SIMCA, followed by PLSDA, to classify as per the biomarkers. They concluded that PLSDA provided better classification of the studied spices and aromatic herbs compared to SIMCA. Table 4 summarizes a few additional studies conducted in the past five years.

REVIEW

Table 4. Machine learning for the authenticity of spices and edible oils.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Classified adulterated and unadulterated olive oil	160	Sensor system (smartphone)	PCA	KNN, SVM, RF, PLS-DA, LW-PLSC	• SVM and RF had 95% accuracy in identifying adulterated olive oil while lower than PLS-DA • PLS-DA and LW-PLSC identified adulterant even at 10% level with 96.2% accuracy	Song *et al.,* 2019
Classified different type of edible oils	5	Chemi-resistive sensor	PCA	RF, kNN	• RF performed better with an accuracy of 73%	Schroeder *et al.,* 2019
Authenticity of olive oil	214	RS	SG	PLS	• Purity of extra virgin olive oil was judged successfully (RS-PLS)	Duraipandian *et al.,* 2019
Source authentication of cocoa bean hybrids	5	NIR-HSI	SNV, SG, PCA	SVM, PLS-DA	• Comparable outcomes were found with PLS-DA and SVM in two-class models, but SVM models significantly reduced prediction error in the five -class	Cruz-Tirado *et al.,* 2020
Authentication of geo origin and adulterant in olive oil	4	LIBS	PCA	LDA	• Resulted in excellent classification results, achieving classification accuracies of 100%	Bellou *et al.,* 2020
Geographical origin of Italian olive oil	92	VIS-NIRS	MANOVA	ANN	• ANN showed a correct classification percentage equal to 94.6%	Violino *et al.,* 2020
Identified oil adulteration and mixtures	-	GC-FID	PCA	GMM	• Model has a 50th percentile absolute error between 1.4–1.8% and a 90th percentile error of 4–5.4% for any 3-way mixtures of the ten oil types	Lim *et al.,* 2020
Authenticity and origin of olive oil	139	LIBS	-	LDA, ERTC, RF, XGBoost	• XGBoost was the best model to trace origin of olive oil with accuracy of 99%	Gyftokostas *et al.,* 2021b
Discriminated olive oils based on olive cultivar origin	60	LIBS, AS	-	LDA, Gradient Boosting	• LIBS performed better with accuracy of 96.0% • Both algorithms were found to provide efficient classification of the olive oil spectra with accuracies exceeding 90%	Stefas *et al.* 2021
Identified coriander oil adulteration	288	GC, NIR	SG, FD, PCA	LDA kNN, PLSR	• LDA and kNN classified samples • Concentration of several adulterants in coriander oil was predicted by PLSR	Kaufmann *et al.,* 2022
Detected fraud of nut shells in cumin powder	-	NIR-HSI	SNV, PCA	SIMCA, PLSR	• Detected peanut, pecan, and walnut shells in cumin powder using NIR-HSI and SIMCA qualitatively and PLSR for quantitatively	Florian-Huaman *et al.,* 2022
Traceability of origin, cultivar, and adulterant in olive oil	-	ATR-FTIR	-	SIMCA,	• Spectral benchmarks in olive oil were extracted by model in the following vibrational modes: 3004, 2952, 2922, 2852, 1742, and 1160 cm-1	Scatigno and Festa, 2022

https://doi.org/10.26656/lifr.1.e25050

Table 4. (cont') Machine learning for the authenticity of spices and edible oils.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Identified type of oil and adulteration in pure oil	47+	GC, RS	-	PCA, RF	• ML-RS showed accuracy of 96.7% for detecting oil type and 98.4% for detecting adulterated oils • PCA-RF was best	Zhao *et al.,* 2022
Identification of cottonseed, canola and soybean oils in olive oil	21	GC	RM	SVM, k-NN, DT	• using SVM model, accuracies of 94.6, 96.4 and 98.2 % were achieved for OO-CO, OO-SO, and OO-CSO mixtures, respectively •	Yakar and Karadag, 2022
Predicted adulterant and fatty acid distributions in flaxseed oil	76+	FT-NIR, FT-MIR, RS	-	SVM, SIMCA, PLSR	• Conditional entropy demonstrated that ML distinguished oils with over 95% accuracy utilizing two or three wavenumbers • oils with an interclass distance greater than 1.1 were identified by SIMCA	Aykas *et al.,* 2022
Identified three Salmonella serotypes at the single-cell level	-	RS	SG, MSC, SNV, HT	CNN	• SG combined with SNV was the most suitable preprocessing achieving 98.7% accuracy • Preprocessing improved performance	Sun *et al.,* 2023
Real-time detection of olive oil adulteration	368	LIBS	-	PCA, LDA, SVMs, LR, GBoost	• all models performed well in terms of separating olive oil from the contaminated ones (100%) • accuracy for adulterant's identification was 92% to 99%	Nanou *et al.,* 2023
Detected adulteration of nutmeg with nutmeg shell	276	FT-NIR	SNV, MSC, PCA	SVM, PLS	• FT-NIR-SVM was found suitable to detect and quantify adulterant • model's projected limit of detection was 1.5% of shell	Drees *et al.,* 2023
Detected adulterants in ground black pepper and dried oregano	39, 24	TD-DART-HRMS		PLS-DA, LASSO	• Black pepper's dual-modality combination distinguished contaminated samples using TD-DART-HRMS-LASSO • simple analysis by TD-DART-HRMS in +ve ion mode was recommended for dried oregano	Zacometti *et al.,* 2023
Detection of soybean oil in olive oil		FT-IR, Vis-NIR, EEM	PCA	PLS-DA	• accuracy of FTIR and Vis-NIR based on PLS-DA was 100% • accuracy of EEMs based on PLS-DA was only 73%	Meng *et al.,* 2023
Adulteration in camellia oil	558	E-nose, FTIR	PCA	SVM, RF, XGBoost, kNN, BPNN, TES	• TES model outperformed other techniques in identifying adulteration ratios • FTIR validated E-nose findings	Wang *et al.,* 2024b
Detected adulteration in saffron	240	FT-MIR	CST, MRMR	PCA, SVM	• SVM-MRMR exhibited accuracy of 98.8 in detecting safflower in saffron	Fattahi *et al.,* 2024b

REVIEW

Table 4. (cont') Machine learning for the authenticity of spices and edible oils.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Authenticity assessment of ground black pepper	99	HS-GC-IMS	-	PLS-DA, SVM	• PLS-DA did not perform well • HS-GC-IMS-SVM classified samples as authentic, exogenously-adulterated or endogenously -adulterated with an overall accuracy of 90 % and 96 %	Zacometti *et al.,* 2024
Identified and classified pure and adulterate oils and margarines	7+	ATR-FTIR	PCA	kNN, LR, SVM, LightGBM, DT	• Detected 1% adulteration • all models correctly classified pure margarines • SVM and DT underperformed while kNN was the best	Tachie *et al.,* 2024
Assessing pungency intensity and origin of Sichuan pepper	210	DPV	-	ANN, kNN	• Origin identification (100%) and adulteration detection (97.9%) • Alkylamides and polyphenols had maximum contribution in pungency • Accuracy was achieved with use of as little as 20% of origin-adulterant	Zhang *et al.,* 2024a
Authenticity of source of edible oils	11500	UV-VIS	-	Orange Data Mining	• Accuracy of 96% was achieved in detecting 6 different types of oils added with olive oil	Deng *et al.,* 2024a
Identified category, origin and grade of Flue-cured tobaccos, Green tea, Purple rice	3+	RS	-	SVM, RF, LDA, kNN, 1D-CNN	• With remarkable accuracy, 1D-CNN was able to identify a variety of agricultural items and simulate contaminated samples, achieving 97.7% and 94.8%, respectively.	Wang *et al.,* 2024d
Geographical origin of chili powder	240	NIRS	PCA	SIMCA, kNN, DT, RF, SVM	• SVM exhibited highest accuracy of 98.4%	Meena *et al.,* 2024
Predicted freeze-drying curve (milk, egg, honey, soymilk)	9	FTIR, DSC	PCA	ANN, RF (18 models)	• ANN as optimum model for predicting temp and time of pre-freezing and desorption stages • RF as best model for parameters of sublimation stage	Liu *et al.,* 2024a
Quality of wheat and pomegranate seeds	3+	NIRS, UV-VIS	PCA, LDA	CNN, PLS	• Protein in wheat and TSS in pomegranate were predictors	Ricci *et al.,* 2024
Detected adulteration in saffron	-	HSI-NIR	MSC, SNV, PCA, PLS	LDA, PLS-DA, SVM, MLP (quantification by PLS, PCA, SVM, MLP)	• PLS outperformed PCA • Recommended SNV-PLS-MLP • While differentiating pure and adulterated saffron, proposed system attained 100% accuracy	Malavi *et al.,* 2024
Authentication of sesame oil	-	Raman spectra	PCA	1D CNN	• 100% accuracy was achieved in identifying multi-component adulteration • Better than GC and colorimetric	Teng *et al.,* 2024

Table 4. (cont') Machine learning for the authenticity of spices and edible oils.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Classified Algerian olive oil cultivar and origin	45	Physicochemical properties, polyphenols, fatty acids	GA	RF, GBoost, LR, DT, NBayes	• K232 and linolenic acid as indicative parameters • DT performed the best	Issaad *et al.,* 2024
Detected variety of edible oils	30	OIRD		RF, kNN, XGBoost, LR	• all models accurately (>95%) identified oil samples • OIRD, DC and fundamental frequency signals contributed 45.7%, 34.1%, and 20.2% to classification, respectively	Sun *et al.,* 2024c
Quantified sorbitol in food packaging	-	ATR-FTIR	SG	PCA, PLS, NN, SVR	• SVR showed superior performance outperforming PCR and NN	Hernández-Fernández *et al.,* 2024
Authenticity of Hainan camellia oil	120	NIR, GC		CNN, PLS-DA, RF, SVM	• SVM-GC, CNN-NIR and CNN based on data fusion exhibited 97.08%, 97.92%, and 98.75% accuracy • CNN overperformed other models	Deng *et al.,* 2024b
Authenticity and quality of Moroccan saffron	120	UV-Vis	PCA	LDA, SVM	• SVM performed the best with an accuracy of 97.9%	Elhamdaoui *et al.,* 2024
Identified adulterated turmeric	168	FT-NIR	MSC, SNV, MMN, SG,	PCA, ENTR, SVR, PCR, PLSR	• PLSR was the best method (>97%) • Successfully detected Sudan dye I qualitatively and quantitatively in turmeric	Kar *et al.,* 2024
Authentication of turmeric essential oil	3+	ATR-FT-MIR	-	PLS, PCR	• PLS efficiently identified sunflower and soybean oil adulterants • PLS overperformed PCR	Cobbinah *et al.,* 2024
Quantified carbonyl in frying oil	116	LF-NMR, AV, PV, IV, TPC, VISC, GC	SNV, MSC, Dt	PLS, PCR, SVR, MLR	• SVR model outperformed the MLR, PCR, and PLS models	Peng *et al.,* 2024

In recent years, consumer interest has grown significantly in cold-pressed oils extracted from various sources such as olive, sesame, hemp, almonds, sunflower, perilla seed, walnut, linseed, and sea buckthorn. This escalated attention owing to their high nutritional value and potential health advantages. From a nutritional standpoint, two essential polyunsaturated fatty acids, linoleic and linolenic, are crucial for human health and must be obtained from external sources like plant materials (Nelson and Cox, 2004). The composition of fatty acids in these oils is also nutritionally significant because of their impact on blood cholesterol levels (Della, 2006). Many of these oil varieties are considered premium products and are priced higher than conventional edible oils, such as refined sunflower oil. This price difference creates an opportunity for unethical producers or vendors to substitute expensive oils with cheaper alternatives for unlawful financial gain. Extra virgin and virgin olive oils are particularly vulnerable to adulteration with lower-quality oils, including soybean, sunflower, sesame, canola, and corn oil. Furthermore, pomace oil, which is extracted from olive fruit residue, is frequently used as an adulterant (Nanou *et al*., 2023).

The development of swift and cost-effective analysis techniques is essential for control laboratories to prevent undesirable practices. Edible oils are intricate food products comprising various components such as glycerides, free fatty acids, lipids, proteins, pigments, sterols, metals, and oxidized substances (Warner and Eskin, 1995), with composition varying based on oil type and extraction method. Numerous analytical approaches have been employed for edible oil examination, including titration, colorimetric or potentiometric studies, GC, GLC, HPLC, and non-destructive methods

https://doi.org/10.26656/lifr.1.e25050

like spectrophotometric techniques (IR, NIR), LIF spectroscopy, raman spectroscopy, and NMR (Endo, 2018). ML techniques have shown effectiveness in categorizing high-dimensional analytical data in oil studies (Houston *et al*., 2020; Gazeli *et al*., 2020). Gazeli *et al*. (2020) utilized LIBS with ML algorithms to extract information from LIBS spectra, offering a straightforward, dependable, and ultra-rapid method for classifying olive oil by acidity level and geographical origin. Classification accuracies ranged from 90 to 99.2% using various statistical models, with LDA yielding the most precise predictions. Data visualization indicated that LIBS combined with these algorithms is a potent tool for swift, on-site, and remote food quality control.

Gyftokostas *et al*. (2020) examined 36 olive oils from different locations in Crete, Greece using LIBS with machine learning algorithms to classify them by geographical origin. Classification accuracy surpassed 90%, showcasing the approach's potential. Pre-processing LIBS spectroscopic data with PCA enhanced some predictive models' accuracy while not impacting k-NN and SVC models. Linear machine learning models were determined to be most appropriate for LIBS spectroscopic datasets. Shiv *et al*. (2024) assessed regression algorithms for identifying argemone in mustard and canola oils. The spectral raw data derived from fluorescence spectrometer assessment of pure and adulterated oils, including local and commercial samples. XGBR, CBR, and RF demonstrated potential for predicting adulteration levels in both oils with high R2 values. In a study by Gyftokostas *et al*. (2021a), researchers analyzed absorption spectra of Greek olive oil samples and mixtures employing Laser-Induced Breakdown Spectroscopy (LIBS) and Atomic Absorption Spectroscopy (AAS). The study incorporated machine learning algorithms to differentiate and categorize olive oils based on their geographical origins. LIBS spectra provided information on elemental composition, while absorption bands (350–750 nm) offered insights into chlorophylls and carotenoids. The spectra underwent preprocessing using Principal Component Analysis (PCA) and were subsequently used to develop predictive models employing Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). These methodologies were validated through "k-fold" cross-validation and external validation, achieving classification accuracy rates of up to 100%. The research examined 170 olive oil samples, comprising 143 from Lesvos, Crete, and Peloponnese, along with 27 binary mixtures of varying ratios. The findings highlighted the benefits of machine learning in enhancing spectroscopic techniques for monitoring olive oil quality. Notably, no negatively impact discrimination performance could be observed on reducing the dimensionality of spectroscopic data, thus, allowing smaller datasets to retain initial information. This feature enables quicker training and the use of less sophisticated hardware. Researchers concluded that ML can effectively assist in authenticating edible oils.

3.5 Honey

Honey, being regarded as a comprehensive food item owing to its composition which includes carbohydrates, proteins, organic acids, vitamins, minerals, and other functional components (Ali *et al*., 2020). It is the sole natural sweetener in human food. Its molecular structure, with three primary sugars (glucose, fructose, and sucrose), allows rapid absorption by the body (Albaridi, 2019). Honey contains abundant natural antioxidants that shield against oxidative stress (Zawawi *et al*., 2021). It offers nutraceutical and therapeutic advantages, such as anti-diabetic, anti-cancer, and anti-microbial properties. Globally, honey is amongst the most popular food commodity (Cianciosi *et al*., 2018).

Economically, honey is consistently classified as one of the top three adulterated foods, alongside milk and olive oil (García, 2018). Honey's composition and price are influenced by its floral and botanical origins, climatic and environmental conditions, and beekeeping practices. Common economic frauds include mislabeling of floral and geographic origins and adding low-cost sweeteners (Arroyo-Manzanares *et al*., 2019). Bee honey can be adulterated using commercially available syrups and inexpensive sweeteners. Recognized adulterants include cane sugar, beet sugar, glucose syrup, fructose syrup, corn syrup, inverted syrup, and high fructose inulin syrup (Morariu *et al*., 2024).

Recent improvements in adulterating techniques have made it challenging to detect honey tampering (Fakhlaei *et al*., 2020). Researchers have developed various analytical methods to identify honey adulteration, including isotopic, chromatographic, and spectroscopic techniques (such as NIR, FTIR, Raman spectroscopy, and NMR), as well as trace element analysis and thermal examination. However, many of these approaches require costly equipment, ongoing maintenance, and skilled operators. Furthermore, sample preparation is often necessary (Gerhardt *et al*., 2018), which is time-consuming and may involve chemical use. Spectroscopy offers a quick, simple, dependable, and cost-effective method that has been underutilized in honey adulteration analysis. Limited research has explored such techniques for identifying corn syrup, high fructose corn syrup, agave syrup, and cane molasses as honey adulterants (Guellis *et al*., 2020; de Souza *et al*., 2021; Valinger *et al*., 2021; Dimakopoulou-Papazoglou *et al*., 2023).

https://doi.org/10.26656/lifr.1.e25050

Table 5. Machine learning in authentication of honey.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observations and accuracy	Reference
Predicted botanical origin of honey (4 floral varieties	119	Wet chemistry, NIR	Boruta	PLS-DA, SVM	• PLS-DA classified acacia and polyfloral • SVM classified acacia, polyfloral and chesnut • linden remained unclassified by both	Bisutti *et al.,* 2019
Classified honey origin	104	GE	-	RF, ANN, SVM	• RF had the highest accuracy of 95.2%	Martinez-Castillo *et al.,* 2020
Authentication of honey origin	103	Isotope, ICP-MS	AVOVA	ANN	• attained accuracy of 96.3% • Nb, δ^2H, As, δ^{18}O, Ir were best markers	Hategan *et al.,* 2021
Botanical classification of honey	-	HSI	LDA	SVM, kNN	• maximum classification accuracy of 95.13% for classifying HSI and 92.80% for classifying hyperspectral instances.	Deshmukh, 2021
Geographical origin of refined sugar	7^+	UPLC-QTof-MS	ANOVA	SVM	• indicate the high potential of UPLC-QTof-MS method for discriminating the refined sugar samples, with an average accuracy of 83.3%	Li *et al.,* 2021b
Identified rice, corn and jaggery syrups in rapeseed honey	59	NMR	-	LRC, DNN, LGBM,	• accuracy of 100% was achieved • DNN performed the best	Rachineni *et al.,* 2022
Botanical recognition of honey	4	Isotopes, ICP-MS	PCA	PLS-DA	• PLS-DA performed well • accuracy was 82-100 % • applied variable selection combined with PLS-DA performed the best	Hategan *et al.,* 2023
Authenticated honey's origin (orange and sunflower)	22	VIS-NIR	Savitzky–Golay smoothing	LDA, SVM, RF	• SVM and RF revealed 100% accuracy	Calle *et al.,* 2023
Detected sugar adulteration in honey	6	HSI	PCA	kNN, SVM (linear and RBF)	• For multi-class classification across various sugar conc and binary adulteration detection, >95% accuracy was attained	Phillips and Abdulla, 2023
Identified geographical origin of rape honey	-	ICP-MS, LC-MS	-	LDA, SVM, RF	• combination of the δ^{13}C, δ^2H, and δ^{18}O values and its extracted protein and saccharides improved accuracy • SVM had highest accuracy of 93.2%	Shuai *et al.,* 2023
Authenticated sucrose in honey	40	UV-VIS	MLR	PLSR, SVR	• SVR performed better than PLSR • correlation between sucrose and honey spectral absorbance was seen	Razavi and Kenari, 2023
Botanical and geo classification of honey (unifloral and multifloral)	247	ICP-MS	Compositional data analysis	PCA, LDA, RF	• RF performed better • Na, Mg, Mn, Sr, Zn, Ce, Nd, Eu, and Tb were best predictors • accuracy was higher in classifying diff botanical origins than same botanical varieties	Mara *et al.,* 2024

https://doi.org/10.26656/lifr.1.e25050

Table 5. (cont') Machine learning in authentication of honey.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observations and accuracy	Reference
Geographical origin of honey	36	ICP-OES	t-test	ML	• Sodium varied significantly among fallow forest and mangrove • Copper found in only 3 locations	Silva *et al.,* 2024
Origin and botanical classification of honey	-	FTIR	PCA, HCA	MLP	• Predicted geographical origin of honey with an accuracy of 96.2%	Avcu, 2024
Detected sugar syrup in honey	5	Logitech web camera	-	CNN	• Accuracy of 94% was achieved	Brar *et al.,* 2024
Identified glucose syrup in honey	181	UV-Vis	SG, PCA	LDA	• Recommended 280-300 nm spectra • LDA achieved accuracy of 99.8%	Nunes *et al.,* 2024
Identification of origin and adulteration in honey	41	Colorimetric sensor/ Optical tongue	-	PLS-DA, PCA	• Classification on the basis of origin achieved an accuracy of 100% • PCA distinguished between different ratios of adulterated honey and syrup	Masoomi *et al.,* 2024

Table 5 summarizes articles published in the past five years, detailing study objectives, sample sizes, methods for determining target honey properties, data preprocessing techniques, and both unsupervised and supervised ML models used to classify and differentiate honey samples based on origin and composition.

Sobrino-Gregorio *et al*. (2019) employed conventional and real-time PCR techniques to detect and measure rice molasses in honey. They simulated various adulteration levels, including 1, 2, 5, 10, 20, and 50% w/w. Real-time PCR proved more effective than conventional methods in assessing adulteration levels by distinguishing different rice DNA concentrations. Wu *et al*. (2022b) utilized Raman spectroscopy combined with convolutional neural networks and chemometrics to identify and quantify adulterants (high fructose corn syrup, rice syrup, maple syrup, maltose syrup, and blended syrup) in honey. Their general CNN model achieved 94.79% accuracy in detecting honey adulterated with any syrup type, while shallow CNNs analyzing honey mixed with single-variety syrup classified samples into four groups by adulteration component with over 97% accuracy. Partial least square regression (PLS) accurately predicted honey purity when combined with single syrup. Raypah *et al*. (2022) integrated NIRS with machine learning to assess water and apple cider vinegar levels added as adulterants in stingless bee honey. NIRS with PCA-LDA identified adulterated honey with 100% accuracy. NIRS and global PLSR together quantified adulterants. Their findings validated the capacity of ML and NIRS in the 700–1100 nm range to accurately identify and measure adulterated samples.

A study by Boateng *et al*. (2022) focused on identifying syrup adulteration in honey. The researchers examined 1525 samples of pure and adulterated honey using five different machine learning algorithms. To classify the samples, they employed Fourier transform infrared spectroscopy with Horizontal attenuated total reflectance. The highest classification prediction accuracies were achieved by gradient boost and support vector machine discriminant analysis. For detecting syrup adulterant concentrations, gradient boost regression yielded superior results. The researchers found that combining spectroscopy with ML provides a quick, simple, and non-destructive method for identifying adulterants. They concluded that utilizing machine learning-based algorithms for honey classification, authentication, and adulteration control could offer a fast and non-destructive approach.

3.6 Milk and milk products

A major hurdle in milk authentication is pinpointing geographical origins, as products from various regions can exhibit significant differences. Milk, also being a perishable product and prone to microbial spoilage, is needed to be authenticated for its freshness or shelf life (Malik *et al.,* 2025). Milk fraud through adulteration is a widespread issue globally, spurring the creation of innovative authentication methods (Kabir *et al*., 2021; Orche *et al*., 2021). Table 6 presents the application of ML algorithms in solving such issues related to milk and milk products. Researchers traced the geographical

Table 6. Machine learning for authentication of milk and milk products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Geographical authentication of cow milk for adulteration	63	FT-NIR, UV-Vis/NIR	t-test	PCA, 20 ANN models (MLP-ANN)	• PCA did not classify samples • 100% accuracy by MLP -ANN • Promising tool for milk adulteration problems irrespective of the spectra	Behkami *et al*., 2019
Discriminated foreign fats and oils in cream and yogurt	12	GC, Raman spectroscopy	SG	PCA	• RS-PCA-a promising system to identify adulteration	Karacaglar *et al.,* 2019
Labelling authentication of milk as per fat	138	Sensor system (smartphone)	PCA	KNN, SVM, RF, PLS-DA, LW-PLSC	• Effective sensor system built in a smartphone • PLS-DA and LW-PLSC correctly classified 100% samples	Song *et al.,* 2019
Classified different type of cheese	5	Chemiresistive sensor	PCA	RF, kNN	• RF was better with accuracy of 91%	Schroeder *et al.,* 2019
Predicted lactoferrin content in bovine milk from mid-infrared spectra	6619	Mid-infrared	RMSE	PLSR. PLS-SVR, PLS-ANN	• PLS-ANN was more suitable for predicting lactoferrin level • If the prediction threshold was set to 500 mg/L, 82% of samples from the validation having a content of LF higher than 600 mg/L were detected	Soyeurt *et al.,* 2020
Classified milk on the basis of heat-loads	1806	FTIR	PCA	kNN, SVM, RF and LDA	• RF performed the best with an accuracy of 97%	Wang *et al.,* 2021
discriminated grass-fed vs nongrass-fed milk	4320	MIRS	PCA	RR, LASSO, EN, LDA MB-DA, etc.	• PLS-DA offered highest accuracy in predicting cow diet	Frizzarin *et al.,* 2021
Predicted total bacterial count in milk		E-nose	Paired t-test	PCA-ANN	• Exact predictions were made for bacteria as low as 1 log cfu/mL.	Yang and Wei, 2021
Identified origin of cow's milk	150	FT-MIR	PCA	LDA, PLS-DA, SVM	• PLS-DA and PCA-LDA showed 100% while PCA-SVM exhibited 94.9% accuracy • SVM without feature extraction showed only 66.7% accuracy	Orche *et al.,* 2021
Authenticated milk of Holstein dairy cattle	1508	FTIR	PLSR	GBM, RF, EN	• Superior ability of GBM to predict difficult-to-measure milk traits shows that this approach can be a promising technique for improving the robustness and accuracy of predictions for selective breeding and dairy cattle industry	Mota *et al.,* 2021
Identified adulterated milk	65,547	MilkoScan, FTIR	-	MD, XGBoost, Extra Trees	• Proposed model had 99.2% accuracy	Chung *et al.,* 2022

Table 6. (cont') Machine learning for authentication of milk and milk products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
"Grass fed" and country of origin of milk	-	SIRA, ICP-MS	-	ML	• More commonly used light elements were H, C, N, O, S • Less used elements were Sr, Pb, Ca	O'Sullivan *et al.,* 2022
Discriminated pure and adulterated milk	1611	MIRS	-	PLS-DA, PLSR, SVM, PPR, BRNN, etc.	• casein as common component • SVM performed better in the classification while BRNN performed better in the quantitative model	Chu *et al.,* 2023
Monitoring of yoghurt fermentation	-	Ultrasonic sensors	Coarse and convolutional	FCNN, LSTM, CNN+LSTM	• Accurate method was non-transmission based and used convolutional neural networks with long short-term memory layers, and a combination of all three	Bowler *et al.,* 2023
Classified Swiss cheese varieties	241	GC, Total acidity, FVCA (8)	Shapley Additive exPlanations	LR, kNN, NB, DT, SVM, Ridge, RF, QDA, ADA, GBC, etc.	• ET and RF had accuracy of about 90% • C1 asmost significant characteristic, followed by C3, C6, and iso-C4. • Iso-C6 as least significant after C2 and C4.	Frohlich-Wyder *et al.,* 2023
Authentication of metabolic impairment in cattle through milk testing	~2700	FTIR	PLSR	EN, RF, GBM, ANN, SE	• Compared with partial least squares regression, ML algorithms achieved more accurate performance • EN and SE exhibited the best predictive capacity for most of the blood parameters	Giannuzzi *et al.,* 2023
Differentiated liquid dairy products	240	Raman spectroscopy	CARS	LightGBM, SVM, RF, XGBoost	• Synergistic combination of all models (fusion model) demonstrated 99% accuracy • Accuracy of only 90% was observed by individual models	Feng *et al.,* 2024
Predicted proteolysis for mozzarella and cheddar cheese	-	Protein, fat, m.c., salt, pH, enzyme conc., storage time	-	XGBoost, SVM, EF,	• XGBoost performed better • Storage time as most crucial • Coagulating enzyme conc and Ca in mozzarella cheese were also imp • Fat or m.c. for cheddar cheese was imp	Golzarijalal *et al.,* 2024
Identified adulterated Podolica milk	309	FTIR	jmpSAS	ML	• Accurately detected presence of Pezzata Rossa milk as low as 5%	Spina *et al.,* 2024
Quality of Parmigiano Reggiano hard cheese	60	LC-MS, UHPLC	Gini impurity	OPLS-DA, RF	• Ripening was important factor than altimetric zone and rind inclusion level • Low performance of OPLS-DA rind inclusion	Becchi *et al.,* 2024

Table 6. (cont') Machine learning for authentication of milk and milk products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/ Algorithm for classification/ discrimination	Observation and accuracy	Reference
Classified milk with subclinical mastitis	20	NIRS	IF	PCA, PLS-DA, RF, SVM	• IF increased accuracy by 10-25 % • in detecting mastitis milk, accuracy of PLS-DA, RF and SVM were 78, 62 and 62 %	Pereira *et al.*, 2024
Categorized yoghurt into fermented, cheesy, fruity, milky	20	QDA and GC -O-MS, sensory	PCA	kNN	• Achieved accuracy of 80% • Identified 10 indicator compounds	Qiu *et al.*, 2024
Detected adulteration in camel's milk	100	Sensory, FT-MIR	PCA, HCA	PLSR, PCR, SVR	• PCA and HCA discriminated against pure and adulterated milk samples • PLSR and PCR effectively quantified cow milk in camel milk	Alia *et al.*, 2024

source of Saanen goat milk from three Chinese provinces using mass spectrometric data and OPLS-DA modeling (Shang *et al*., 2022). Behkami *et al*. (2019) employed FTIRS with machine learning models to detect milk adulteration as per geographical origin. They optimized the model by decreasing input variables, achieving 100% classification accuracy for milk's geographical source. Xue and Zhao (2024) created a semi-supervised deep learning approach that combined autoencoder feature extraction and multilayer perceptron regression networks. This method rapidly detected goat milk adulteration on-site employing voltammetric fingerprints. A ten-electrode microsensor array distinguished among pure goat and cow milk profiles with 2 to 5% precision. Training the model on 88 labeled spiked samples, supplemented by unlabeled surrogates, enhanced test prediction accuracy by 4.3% compared to only SL. It achieved 2.21 RMSE lower to 1% adulteration. Independent validation across 0-100% cow milk ratios demonstrated under 2% relative error up to 90% blending before declining fidelity, detecting adulterants up to legal limits.

For organic milk authentication, methods based on organic feed fingerprints, including stable isotopic ratio, elemental distribution, and fatty acid profile, are most commonly explored (Malik and Sharma, 2021; Lozano-Castellón *et al*., 2024). Scientists differentiated organic from conventional milk by analyzing fatty acid profiles, vitamin E, stable isotopes (δ13C and δ15N), and stable isotopic ratio of fatty acids and amino acids. A multivariate OPLS-DA model utilizing stable isotope ratios, elements, and fatty acids was also successfully employed (Chung *et al*., 2019; Xu *et al*., 2021b).

In an additional study, a PLS-DA model was developed leveraging the entire 1H NMR metabolite profile, and it effectively differentiated between conventional and organic milk (Phuenpong *et al*., 2021). A recent investigation used an OPLS-DA model and the fatty acid profile identified by GC-MS/FID to accurately classify conventional and organic milk (both raw and retail) (Hou *et al*., 2023). This model to distinguish between milk from grass-fed cows in non-organic systems and milk produced by organic cows would be intriguing to investigate, though, as this was not examined. Later, it proved that the highest level of accuracy could be obtained by using the combination of linear discriminant analysis and partial least squares discriminant analysis (PLS-DA) to forecast the cow diet from MIR spectra and distinguish between grass-fed and nongrass-fed milks obtained from cows on pasture or indoor total mixed ration-based feeding systems over a three-year period (Frizzarin *et al*., 2021). Both hydrogen and oxygen isotopes in Slovenian milk, as well as the amounts of water, casein, and lactose, were examined by Hamzić Gregorčič *et al*. (2020). Additionally, stable oxygen isotope ratios of goat, sheep, and cow milk were compared by the authors. Based on the isotopic data collected, the scientists came to the conclusion that, if validated internationally, the technology used might form a protocol for determining if milk has been blended with water.

A significant dairy product scam was documented in China as a result of melamine contamination of powdered newborn milk, which sickened and killed young children. In order to improve the protein content of food and make money, melamine (2,4,6-triamino-1,3,5-triazine), a rich nitrogen molecule, can be incorporated illegally (Visciano and Schirone, 2021). Marchetti *et al*. (2021) explored at the adulteration of buffalo yogurt using bovine (Bos taurus) milk in order to identify the risk variables that could render the product susceptible to fraud, given the popular demand and popularity for buffalo (*Bubalus bubalis*) dairy products. In 25% of the samples examined, the scientists' use of RTPCR revealed the undisclosed existence of bovine DNA in instead of buffalo DNA. The application of machine learning driven analysis of fatty acid profiles presents a promising approach for ensuring the market authenticity of Labaneh (a traditional middle eastern dairy product) and detecting non-milk fat adulterants. By leveraging machine learning algorithms such as SVM and ANN, we have successfully trained models to discriminate between authentic and adulterated Labaneh samples with high accuracy and reliability (Raje, 2024).

In the dairy industry, cheese-making characteristics of cattle are crucial but challenging to assess individually due to constraints in gathering phenotypic data. The AfiLab system (Afimilk) offers advantages by enabling phenotype collection from each cow during milking. This research aimed to evaluate the integration of the AfiLab real-time milk analyzer. Information and specimens were obtained from 499 Holstein cows on two farms equipped with AfiLab. The study examined 16 traits: 9 milk coagulation traits (3 milk coagulation properties [MCP], 6 curd firming traits [CFt]) and 7 cheese-making traits (3 cheese yield [CY] traits, 4 milk nutrient recovery in the curd [REC] traits). ANN method outperformed other base learners (EN, GBM, XGBoost), enhancing accuracy across cross-validation scenarios by 7% for MCP, 5% for CFt, 8% for CY, and 7% for REC. The findings demonstrate that combining on-farm in-line data with stacking ensemble ML effectively yields reliable daily predictions of milk cheese-making traits (Mota *et al*., 2022). Prato and mozzarella cheeses share similar production processes and end-product features. Prato cheese requires a 25-day maturation period, making it susceptible to fraud. Mid-infrared spectroscopy and ML methods were employed to validate the authenticity of commercial prato cheese. Forty samples each of prato and mozzarella cheese were produced and analyzed over a 60-day maturation period. Twenty commercial prato cheese samples from 13 brands were attained. PCA identified two distinct groups (prato and mozzarella). Sixteen commercial samples aligned with mozzarella cheese, while four aligned with prato cheese. Discriminant methods achieved classification rates exceeding 84%. Market samples were categorized as mozzarella cheese with an 82.5% probability. These techniques effectively distinguished and grouped samples, indicating that the market samples were not genuine prato cheese (Tolentino *et al*., 2023).

3.7 Meat, fish, egg and seafood products

Meat and other animal-derived products are susceptible to adulteration, necessitating authentication of these high-protein items. Fraudulent practices in meat authentication can be classified into three categories: (i) meat origin (including breed, geographical source,

Table 7. Machine learning for the authenticity of meat, fish, egg and seafood products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/Algorithm for classification/ discrimination	Observation and accuracy	Reference
Total volatile basic nitrogen content in shrimp	-	HSI	SPA, SAE	LS-SVM, PLSR, MLR	• SAEs-based prediction models performed better than either full wavelengths-based or SPA-based • SAE-LS-SVM was the best model	Yu *et al*., 2019
Identified duck meat in lamb meat	20	Vis-NIR-HIS	SDC	PLSR	• PLSR model achieved an accuracy of 98%	Zheng *et al*., 2019
Detected minced lamb and beef fraud	4	NIR	PCA	PLS-DA	• Addition of 1% or higher minced beef in foal was identified • Adulteration of pork, meat of Lidia breed or foal in lamb and beef was better identified than adulteration with chicken	López-Maestresalas *et al*., 2019
Identified rainbow trout adulteration in salmon	516	RS	FD, SD, MSC, SNV	PLSR, kNN, Glmboost, Enet, Rdge, LASSO elc.	• Based on MSC preprocessing, GA–KM–Cubist model produced satisfactory results (87%)	Chen *et al.*, 2019b
Detection of irradiated dry fermented sausages	50	NIRS	PCA	OPLS-DA	• Models provided a correct classification rate of 100% for both irradiated and non-irradiated specimens	Varra *et al.*, 2020

Table 7. (cont') Machine learning for the authenticity of meat, fish, egg and seafood products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/Algorithm for classification/ discrimination	Observation and accuracy	Reference
Microbial load of beef	104	Visible- and NIR - HSI	PCA	PLS-R	• Use of appropriate spectral pretreatments and band selection methods was key for robust model development • Potential method for real time monitoring of microbial growth along the meat supply chain	Achata *et al.,* 2020
Minerals assessment and sodium control in hamburger	63	FAAS, NIRS	PCA	PLS-DA, PLS-R	• Classification model was efficient with high selectivity and sensitivity • Successfully used to control essential minerals Fe, Ca and K besides of quantify and detect hamburger with high content of Na	Rebellato *et al.,* 2020
Detected pork and duck meat in minced beef	82	NIRS	-	DA, PLS	• DA model with selected wavelength achieved the best results • Optimal PLS models with full-wavelength gained correlation coefficient Rp of 95.80% and 95.69%,	Leng *et al.,* 2020
Authentication of geographical origin of pork	70	ICP-MS	ANOVA, PCA, Duncan, CDA, CA	RF, SVM, FNN	• FNN was found best in discriminating origin of Chinese pork (95.7%) • SVM overperformed RF	Qi *et al.,* 2021
Geographical traceability of seafood, Shandong Scallop	153	ICP-MS	ANOVA, PCA	LDA, kNN, RF, SVM	• Achieved an overall accuracy of 99% • LDA performed the best (100% • 13 elements significantly affected the traceability oforigin	Kang *et al.,* 2022b
Authenticated animal blood food	150	ICP-MS, AAS	PCA, FLDA	ELM	• FLDA was better than PCA • Provided identification accuracy of 96 percent or higher.	Han *et al.,* 2022b
Traceability of origin of seafood	4	Stable isotopes	-	PML	• Achieved an accuracy of 100% • Stable isotopes ratios of carbon, nitrogen and sulfur could be used	Tsirogiannis *et al.,* 2022
Quantitative detection of beef adulterated with pork	8	E-nose	Stable value	SVR, RFR, BPNN, 1DCNN, 1DCNN-RFR	• 1DCNN-RFR achieved the highest accuracy of 99.8%	Huang and Gu, 2022
Authenticated source of meat floss products	300	E-nose, FTIR, GC -MS	PCA	LDA, QDA, kNN, RF	• LDA exhibited the highest accuracy of more than 99.0% • Pork predominantly included aldehydes (dodecanal and 9-octadecanal)	Putri *et al.,* 2023
Authentication of beef cuts	-	NAA	ANOVA	CART, MLP, NB, RF, SMO	• MLP-96%, AMO-95%, RF-91%, NB-86% and CART-70% • Br, Fe, K, Na, Se and Zn were assessed • MLP provided the best classification	Mazola *et al.,* 2023

Table 7. (cont') Machine learning for the authenticity of meat, fish, egg and seafood products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/Algorithm for classification/ discrimination	Observation and accuracy	Reference
Geographical traceability of marine bivalves	120	Stable isotope, content of C, O, H	-	XGBoost	• achieved an accuracy of 93.7% • Stable isotope and content of C, N, O and H were significantly different • $\delta^{13}C$, $\delta^{15}N$, C and H made higher contributions	Kang *et al.*, 2023
Identified provenance of chicken eggs	263	LC-MS/ MS	R software	OPLS-DA, SVM	• Differentiated cage, barn, and free-range eggs with accuracy of 94, 82 and 82 %, respectively	Chin *et al.*, 2023
Detected lard in sausage	-	FTIR and GC-MS	PCA	DA	• FTIR spectroscopy is simpler in handling samples than GC-MS • halal food analysis using GC-MS confirmed and clarified the products adulterated by pork	Ahda *et al.*, 2023
Authentication of pork meat	61	NIRS	PCA	DD-SIMCA	• Excellent results were obtained with SNV pre-treated data and 4 PCs that allowed to reach 100 % sensitivity and specificity in calibration and validation	Totaro *et al.*, 2023
Authenticity of eggs (housing system)	290	^{1}H NMR	-	LDA, QDA, PLS-DA, RF, kNN, ANN	• SVM was the best model • predicting housing system needs research in terms of factors like breed, feeding	Bischof *et al.*, 2024
Screened cooked beef adulterated samples	-	MALDI-TOF MS	-	XGBoost, PLS-DA	• XGBoost was used classify raw and heat-processed meats based on 36 heat-stable distinctive protein • PLS-DA achieved 97.4% accuracy in determining adulteration ratio of simulated adulterated beef	Pu *et al.*, 2024
Traceability of frozen thawed salmon in fresh ones	-	Bioimpeda-nce flexible sensing	-	PCA-BOA-SVM	• optimized model worked with accuracy of 97.1%	Zhang *et al.*, 2024b
Authentication of Mediterranean anchovies	-	FT-NIR	SNV, SG	L-SVM, P-SVM, RF, kNN	• L-SVM and kNN could not successfully classified • P-SVM and RF showed 95% accuracy • SG+SVM pre-processing worked best	Dalal *et al.*, 2024
Identified freshness and adulteration in meat	-	µPD-OES	-	LDA	• traced meat freshness with accuracies of 96.0, 98.7, and 94.7% in beef, pork, and chicken meat • Identified beef adulterated with duck meat with an accuracy of 99.5%	Ren *et al.*, 2024
Discriminated wild and farmed mussels	366	FT-NIR, FT-MIR	-	PLS-DA	• best discrimination (98% accuracy) done using MIR spectra of mussel's interior • Accuracy using shell's spectra was slightly lower (92%)	Ayvaz *et al.*, 2024

Table 7. (cont') Machine learning for the authenticity of meat, fish, egg and seafood products.

Objective	Sample size	Analytical method	Method of feature extraction/ data process	Model/Algorithm for classification/ discrimination	Observation and accuracy	Reference
Predicted shelf life of marine fish	5	14 features	GA, BP	RBF	• Developed a real time shelf-life prediction system • Maximum absolute error was only 0.48 • A radial basis function model performed the best	Cui *et al.,* 2024
Authenticated fresh retail beef cuts	125	iKnife-REIMS	-	DA, SVM, kNN	• DA, SVM, and KNN had mislabeling discrimination rates of 95.54%, 99.91%, and 100%, respectively • 11 fatty acids and 44 phospholipids were identified	Song *et al.,* 2024
Monitored quality and freshness of eggs	-	Haugh Unit, pH, color, air cell height, UV–Vis, FS	PCA	PLS	• PLS models correctly predicted the ages of the eggs • When egg samples were kept at 35°C, the most appropriate age assessment models for eggs were found for their UV-Vis and fluorescence spectra (for example, 320 nm)	Ciftci *et al.,* 2024
Assessment of microbial spoilage of broiler breast meat	197	organoleptic, plate count,molecular methods	Deep learning, PCA	LDA, SVM	• SVM-ResNeXt101 from amplitude component images at 0.25 cycles mm^{-1} achieved classification accuracy of 76% • SIRI technique was found effective	Olaniyi *et al.,* 2024
Authenticity of adulteration in collagen powder	-	HSI	PCA	GA-kNN, PSO-SVM, GBoost, RF	• GA-kNN-99%, PSO-SVM-94%, GBoost-98% in qualitative detecting adulterant • RF quantified adulterant	Lin *et al.,* 2024
Geographical origin of chicken (breast and drumstick)	120	ICP-OES, ICP-MS	-	OPLS-DA, CDA	• The elements identified in OPLS-DA were verified by receiver operating characteristic (ROC), and the accuracy was 100 % • The canonical discriminant analysis (CDA) also showed 100 % accuracy	An *et al.,* 2024

farming methods, wild or farmed animals, organic or conventional production, feed, and meat proportions), (ii) meat preservation techniques, particularly distinguishing fresh from thawed meat, and (iii) non-meat components such as water or added foreign substances like cereals (Esteki *et al*., 2018; Chaudhary *et al.,* 2022). Horse meat is sometimes fraudulently incorporated into processed foods labeled as beef (Visciano and Schirone, 2021). Cereals (barley, oat, rye, maize, rice, and wheat) are common plant-based substitutes for meat protein due to their high protein content and good bioavailability. Since younger animals typically yield more expensive meat than older ones, various analytical methods are needed to verify an animal's age before slaughter (Vishnuraj *et al*., 2021). Some studies focusing on detecting these problems have been summarized in Table 7. Robert *et al*. (2021) showcased Raman spectroscopy as a viable alternative for intact meat discrimination. By combining Raman spectroscopy with three chemometric techniques (PCA, PLS-DA, and SVM), they successfully distinguished beef, venison, and lamb meat samples despite their similar chemical compositions. All three chemometric tools performed adequately in differentiating the three red meat types. With its brief 15-second analysis time, minimal sample preparation requirements, and potential for inline integration, Raman spectroscopy presents an effective method for meat industry implementation in discriminating against and sorting red meats.

Keshavarzi *et al*. (2020) used ATR-FTIR and transmission FTIR spectroscopy with PCA to discriminate beef from chicken meat. PCA can cluster beef from chicken meat when applied without preprocessing on whole spectra, while ATR fails. Transmission FTIR (KBr plate) is preferable to ATR-FTIR for PCA on raw spectra. For quantification, ANN improved regression models compared to PLS-R due to non-linear modeling. Windarsih *et al*. (2023) found pork sausage and beef sausage containing pork could be classified from authentic beef sausages using PLS-DA with high accuracy. Metabolites of 2-arachidonyl-sn-glycero-3-phosphoethanolamine, 3-hydroxyoctanoylcarnitine, 8Z,11Z,14Z-eicosatrienoic acid, D-(+)-galactose, oleamide, 3-hydroxyhexadecanoylcarnitine, arachidonic acid, and α-eleostearic acid increased pork levels in beef sausage, providing indicators to detect pork. Nunes *et al*. (2020) developed a multivariate classification method using Mid-IRS and PLS-DA to detect carrageenan, sodium chloride, and tripolyphosphate in bovine meat. Meat pieces were injected with adulterant mixtures. Multiclass PLS-DA models detected only tripolyphosphate, but a two-class model discriminated adulterated and non-adulterated meat with high success. Derz *et al*. (2021) studied purity and adulteration of sausages and ham from chamois, red deer, or roe deer, from various countries. They developed a duplex probe real-time PCR to determine chamois meat in sausages. Prolonged freezing negatively affects meat's nutritional value and flavor, potentially introducing harmful microorganisms and carcinogens. Microbial load, represented by total viable count, is essential for food freshness evaluation (Liang *et al*., 2022).

A data analysis-enhanced electronic nose was employed to create and evaluate a technique for authenticating sausages and identifying soy protein adulteration. The proposed method reduces sensor array data dimensionality and differentiates four sausage varieties. Another approach combined electronic nose features with multivariate analysis, utilizing PCA for feature examination. Clustering outcomes indicated that the regularities of four sausage types require investigation through non-linear methods like ANN (Kalinichenko and Arseniyeva, 2020). Song *et al*. (2021) identified 41 pork meat compounds using HS-SPME-GC-MS, including alcohols, aldehydes, ketones, and hydrocarbons. PCA and OPLS-DA were used to identify volatile compound profiles. Migration testing and spoilage in glass vials revealed that aromatic hydrocarbons originated from packaging rather than pork thermal decomposition. Among the identified volatiles, only ethanol, 2,3-butanediol, and 2-ethyl-1-hexanol were recognized as potential minced pork meat spoilage markers and possible quality indicators. Scientists detected wild boar in beef meatballs using FTIR with chemometrics. FTIR spectra were analyzed using PLS and PCA, revealing distinct characteristics between wild boar and beef in meatballs. The PLS calibration method achieved a determination coefficient (R2) of 0.9991 and RMSEC of 1.028%. Validation methods yielded R2 and RMSECV values of 0.9999 and 0.300% respectively (Ahda *et al*., 2020). Shi *et al*. (2019) investigated meat freshness using IRS information for viable count and sensory score. Organic pork authentication was accomplished by blending SNNs multi-element and isotopic data from defatted meat and developing PCA and OPLS-DA models (Zhao *et al*., 2020). Organic and conventional beef were distinguished using HRMS lipidomic analysis and PCA-LDA models, achieving 84% accuracy (Robson *et al*., 2022). RSDE was utilized as a rapid machine learning algorithm for authenticating chicken fillets' growth conditions and freshness within a validated chemometric workflow. RSDE outperformed other classification models such as PLS-DA, CPANN, and SVM. Combining spectra further upgraded categorization performance. Implementing minimal categorization probabilities can prevent categorized meat of recognized origin into group. However, this approach is not recommended for classifying meats of unknown origin (Parastar *et al*., 2020).

Food fraud also extends to fishery products. Fish substitution is a global issue, with instances reported worldwide. For example, cheaper species are often substituted for Atlantic cod (*Gadus morhua*), or the cod designation is illegally used for unauthorized species. Additionally, some fish products fail to adhere to proper salting or drying protocols. In Malaysia, other reported substitutions include replacing lizardfish with threadfin, salmon with trout and tuna, and fish with prawn (Visciano and Schirone, 2021). Eggs, being a crucial component of diets globally, require a non-destructive technique for validating native egg authenticity. Research has explored the potential of merging near-infrared spectroscopy (NIRS) with data driven-based class-modeling (DDCM) and model-independent variable choice. A study involving 122 eggs of three types employed PCA for exploratory analysis. The findings suggest that NIR spectroscopy blended with class-modeling could be an effective tool for detecting native egg authenticity (Chen *et al*., 2019c). In another study, a heat transfer model was used to predict the impact of wet heating at 95-100°C on myosin and actin denaturation in false abalone (*Volutharpa ampullaceal perryi*). Using 3D finite element heat transfer analysis and reaction kinetics, it was found that muscle proteins were fully denatured within 60-80 seconds of wet heating. Low field 1H nuclear magnetic resonance (LF-NMR) and magnetic resonance imaging analyses showed a decrease in immobilized water content with increased cooking time, along with reduced shear force. Partial least squares (PLS) analysis revealed a strong correlation among free water, color, taste, and overall acceptability (He *et al*., 2019).

4. Challenges and future prospects

While machine learning presents a number of opportunities for the food industry, it is important to recognize and address some of the relevant obstacles to the industry's use of this technology. The problem with these approaches is that a food product's physical, chemical, and biological characteristics are influenced by a wide range of variables, including location, type, environment, water, fertilizers, harvesting time and technique, pre-processing and processing conditions, packaging, storage, transportation, and consumption. Furthermore, there are many commonalities among many commodities. To guarantee the model's predictability, a variety of samples ought to be included in its design. Furthermore, it is essential to carefully address data privacy and security issues to ensure that sensitive data is protected at every stage. Food, health, and agriculture are only a few of the fields where many food quality and safety records are widely scattered and sometimes not digitalized. Modeling performance may be enhanced by combining this information in a digitalized and computer-based format and integrating other data sources to create a larger dataset. One can cover a wider range of traits and variables that affect food authenticity by combining data from several sources. A deeper comprehension of the intricate relationships and interactions within the food realm is made possible by this all-encompassing approach. High-quality, labeled datasets are essential for training machine learning models. In food authentication, obtaining such datasets can be difficult due to the diversity in food products, variations in ingredients, and processing methods. Moreover, food fraud cases are often rare and hard to detect, making it difficult to gather sufficient authentic and fraudulent examples. Food items come in many forms, compositions, and origins, resulting in complex data for ML models to process. Variations in texture, color, composition, and even the physical environment in which the food is produced can complicate feature extraction and classification. ML models need to generalize across these variations, which is difficult with high diversity in food types and processing techniques. Extracting relevant features from raw data (e.g., chemical compositions, images, spectroscopy readings) is a critical step. The authenticity of food may be identified through subtle differences in molecular structures, textures, or chemical compositions that are not always easily captured with traditional sensing technologies. This requires advanced feature engineering or multi-modal data fusion. Food authentication often deals with imbalanced datasets, where genuine samples far outnumber fraudulent or adulterated ones. This imbalance can lead to biased models that fail to accurately detect fraud or adulteration. Addressing this requires techniques such as oversampling, under sampling, or cost-sensitive learning to better train the model on rare fraudulent examples. Food fraud detection models should be interpretable, especially in critical regulatory contexts. Decision-makers (e.g., food safety authorities, regulatory bodies) need to understand why a model classified a particular food item as authentic or fraudulent. The "black-box" nature of many ML algorithms, such as deep learning models, can make it challenging to provide clear and trustworthy explanations. ML models used for food authentication may need to operate in real-time, especially in supply chains, production lines, or point-of-sale systems. This requires models to process large volumes of data quickly while maintaining accuracy. Moreover, integrating real-time data sources (e.g., sensor data, IoT devices) into the authentication process can be technically challenging. Food fraud methods can evolve, and new adulterants or fraudulent practices may emerge. ML models need to continuously adapt and generalize new, unseen types of fraud. This can be difficult because retraining models with new data may require extensive resources and data collection. The implementation of ML in food authentication must adhere to regulatory standards, and the models must meet the legal requirements for safety, accuracy, and fairness. Moreover, ethical concerns around privacy, bias, and the impact of false positives/negatives in food safety decisions must be considered. Food authentication can involve various types of data such as chemical, spectral, image, and sensor data. Integrating these different data modalities into a unified machine learning model is a significant challenge but can improve model accuracy and robustness.

When integrating various data sources, care must be taken to consider the quality, relevancy, and potential biases of the data. Food companies, governments bodies, regulatory agencies, academicians, consumers, and other stakeholders along the food distribution network must come up with new ideas and collaborating strategically in view to promote ML development and integration in the industry. If the food industry intends on

https://doi.org/10.26656/lifr.1.e25050

implementing these new models, ethical issues, regulatory compliance, and security measures must be given top priority to ensure responsible and secure implementations. The way food experts and industry participants manage risks and make decisions will be significantly disrupted and modified as amalgamation of machine learning and food integrity fosters. This might lead to the rise of potentially new operational risks. As a result, anticipating these risks and developing preparations for emergency mitigation are essential.

ML can detect subtle patterns and anomalies in food products that may indicate adulteration or mislabeling. By analyzing data from various sources (e.g., chemical profiles, spectroscopy, DNA testing, or images), ML models can help identify fraudulent food products, such as mislabeled organic goods or diluted expensive ingredients, thereby protecting consumers and maintaining food integrity. ML algorithms can analyze large, complex datasets more efficiently than traditional methods. By automating the analysis of food authenticity, ML can reduce human error, speed up testing processes, and increase throughput, which is particularly beneficial in food production, processing, and retail. ML, when paired with techniques such as hyperspectral imaging, near-infrared spectroscopy (NIR), or other sensor technologies, enables non-invasive and non-destructive food authentication. This minimizes the need for costly, time-consuming, or waste-generating testing methods (e.g., chemical analysis or physical destruction of samples), making it easier to authenticate food at various stages in the supply chain. ML models can be integrated into real-time systems for on-the-spot food testing in production lines, warehouses, or retail environments. For example, using sensors and cameras, ML algorithms can analyze products as they pass through conveyor belts or on store shelves, offering instantaneous feedback on product authenticity. This enables rapid decision-making and enhances food safety monitoring. ML can facilitate more transparent and traceable food supply chains. By combining ML with blockchain or other data storage technologies, it is possible to track the journey of a food product from farm to table, verifying its authenticity at each stage. This transparency can help prevent fraud and improve consumer trust in food products. ML can be used to develop customized authentication methods tailored to specific food categories. For instance, different authentication models can be applied to high-value or high-risk food products such as olive oil, honey, or seafood, where fraud is more prevalent. Additionally, personalized solutions for specific regional or local food types can improve the accuracy of fraud detection. ML models, once trained, can be deployed at scale with relatively low cost compared to traditional laboratory methods. This reduces the need for expensive, specialized equipment and skilled labor, making food authentication more accessible to small and medium-sized enterprises (SMEs), regulatory agencies, and even developing countries where resources are limited. ML can help identify early warning signs of potential fraud before it becomes widespread. By analyzing historical data and detecting emerging trends, machine learning models can predict and mitigate risks of fraud, allowing food producers, retailers, and regulators to act preemptively and reduce potential harm. ML can play a role in verifying the sustainability and ethical sourcing of food products. By analyzing data such as production practices, environmental impacts, or geographical origin, ML models can help authenticate whether products are genuinely organic, fair trade, or sustainably sourced, fostering consumer confidence and supporting ethical practices in the food industry. ML-powered food authentication systems can support regulatory agencies in enforcing food safety standards and ensuring compliance with labeling laws. Automated tools can analyze trends and patterns in food products, providing insights for policymakers to establish better regulations, monitor food standards, and improve enforcement strategies. ML can help consumers make informed decisions by offering transparency in food products. For instance, smartphone apps that use machine learning could scan food labels, ingredients, or barcodes to verify the authenticity of food, empowering consumers to verify claims such as “organic” or “locally sourced” and reduce reliance on trust alone.

Conclusion

Modern analytical devices and multivariate data analysis techniques aid in the extraction of vital data from signals generated by numerous equipment, which is helpful in the complex problem of recognizing food fraud or verifying authenticity. Application of ML algorithms in food integrity is crucial for the competitive and quality-focused food industry. These modern tools offer distinct benefits by anticipating and detecting risks to food safety, thereby, streamlining production procedures, reducing food waste, and boosting product monitoring. Customized food safety solutions that are suited to various food industry segments will continue to benefit from the ongoing development of machine learning and its predictive capabilities. ML blended with analytical techniques on food metrics data can help identify adulterated food samples and ensure proper food authentication. Innovative methodologies employed for food authentication include DNA-based methods, lateral flow assays, paper-based optical tongue, e-tongue and more. Modern techniques to distinguish between organic and conventional foods have provided valuable perceptions into the origins and authenticity of food products. In order to determine the differences between organic and conventional food, methods such as elemental analysis, stable isotope analysis, molecular assessment, rDNA fingerprinting of microbes, and the combination of techniques like FTIR spectroscopy, ICP-MS, e-Nose, and VIS/NIR spectroscopy with sophisticated statistical treatment of data are very effective. Preserving food integrity and quality during production operations is a challenging issue for the food business. The food and beverage business may actively lower risks, enhance product safety, and boost consumer trust by utilizing ML in conjunction with technologies like exhaustive data, predictive analytics, IoT, lockchain and sensor-based approaches.

Conflict of interest

The authors declare no conflict of interest.

References

Abdulhafedh, A. (2022). Comparison between common statistical modeling techniques used in research, including Discriminant analysis vs logistic regression, ridge regression vs LASSO, and decision tree vs random forest. *Open Access Library Journal*, 9(2), 1-19. https://doi.org/10.4236/oalib.1108414

Achata, E.M., Oliveira, M., Esquerre, C.A., Tiwari, B.K. and O'Donnell, C.P. (2020). Visible and NIR hyperspectral imaging and chemometrics for prediction of microbial quality of beef Longissimus dorsi muscle under simulated normal and abuse storage conditions. *LWT-Food and Science Technology*, 128, 109463. https://doi.org/10.1016/j.lwt.2020.109463

Aggarwal, C.C. (2018). Neural networks and deep learning. Cham: Springer. https://doi.org/10.1007/978-3-319-94463-0

Ahda, M., Guntarti, A., Kusbandari, A. and Melianto, Y. (2020). Authenticity analysis of beef meatball adulteration with wild boar using FTIR spectroscopy combined with chemometrics. *Journal of Microbiology, Biotechnology and Food Sciences*, 9(5), 937-940. https://doi.org/10.15414/jmbfs.2020.9.5.937-940

Ahda, M., Guntarti, A., Kusbandari, A. and Nugroho, H.A. (2023). Identification of Adulterated Sausage Products by Pork using FTIR and GC-MS Combined with Chemometrics. *Journal of Chemical Health Risks*, 13(2(, 325-332. Retrieved from https://jchr.org/index.php/JCHR/article/view/200/201

Ahn, D., Choi, J.-Y., Kim, H.-C., Cho, J.-S., Moon, K.-D. and Park, T. (2019). Estimating the Composition of Food Nutrients from Hyperspectral Signals Based on Deep Neural Networks. *Sensors*, 19, 1560. https://doi.org/10.3390/s19071560

Alam, S., Cui, S., Alam, Y. and Akujuobi, C.M. (2023). Unsupervised Network Intrusion Detection Using Convolutional Neural Networks. 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), *Las Vegas, NV, USA*, 2023, 0712–0717. https://doi.org/10.1109/ccwc57344.2023.10099151

Al-Awadhi, M.A. and Deshmukh, R.R. (2021). Honey Classification using Hyperspectral Imaging and Machine Learning. In *2021 Smart Technologies, Communication and Robotics (STCR)*, Sathyamangalam, India, 2021, pp. 1-5. https://doi.org/10.1109/STCR51658.2021.9588907

Albaridi, N.A. (2019). Antibacterial potency of honey. *International Journal of Microbiology*, 2019(1), 2464507. https://doi.org/10.1155/2019/2464507

Ali, A., Sajood, S., Taban, Q., Mumtaz, P.T., Rather, M.A., Paray, B.A. and Ganie, S.A. (2020). Honey as component of diet: importance and scope. *Therapeutic Applications of Honey and its Phytochemicals,* 1, 215-236. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-15-6799-5_11

Alia, O.A.E., Zine-Eddine, Y., Souhassou, S., Chaji, S., Ansari, I.A.A., Mohammed, Z.R., Kzaiber, F., Oussama, A., Ansari, S.A. and Boutoial, K. (2024). Investigating the potential of Fourier transform mid-infrared spectroscopy combined with chemometrics for detecting camel’s milk adulteration. *Italian Journal of Food Science*, 36(2), 150–162. https://doi.org/10.15586/ijfs.v36i2.2500

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Fadhel, S.M.A., Al Amidie, M. and Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8, 1-74. https://doi.org/10.1186/s40537-021-00444-8

Amini, M. and Rahmani, A. (2023). Machine learning process evaluating damage classification of composites. *International Journal of Science and Advanced Technology*, 9(2023), 240-250. Retrieved from https://ssrn.com/abstract=4331923

Amoghin, M.L., Abbaspour-Gilandeh, Y., Tahmasebi, M. and Arribas, J.I. (2024). Automatic non-destructive estimation of polyphenol oxidase and peroxidase enzyme activity levels in three bell pepper varieties by Vis/NIR spectroscopy imaging data based on machine learning methods. *Chemometrics and Intelligent Laboratory Systems*, 250, 105137. https://doi.org/10.1016/j.chemolab.2024.105137

An, J.M., Hur, S.H., Kim, H., Lee, J.H., Kim, Y.K., Sim, K.S., Lee, S.E. and Kim, H.J. (2024). Determination of the geographical origin of chicken (breast and drumstick) using ICP-OES and ICP-MS: Chemometric analysis. *Food Chemistry*, 437, 137836. https://doi.org/10.1016/j.foodchem.2023.137836

Araujo, E.M., de Lima, M.D., Barbosa, R. and Alleoni, L.R.F. (2019). Using machine learning and multi-element analysis to evaluate the authenticity of organic and conventional vegetables. *Food Analytical Methods*, 12 (11), 2542-2554. https://doi.org/10.1007/s12161-019-01597-2

Arroyo-Manzanares, N., García-Nicolás, M., Castell, A., Campillo, N., Viñas, P., López-García, I. and Hernández-

REVIEW

Córdoba, M. (2019). Untargeted headspace gas chromatography–Ion mobility spectrometry analysis for detection of adulterated honey. *Talanta*, 205, 120123. https://doi.org/10.1016/j.talanta.2019.120123

Avcu, F.M. (2024). Clustering honey samples with unsupervised machine learning methods using FTIR data. *Anais da Academia Brasileira de Ciências*, 96(1), e20230409. https://doi.org/10.1590/0001-3765202420230409

Aykas, D.P., Menevseoglu, A. and Gunes, N. (2022). Information theory and machine learning based authentication of flaxseed oil using portable and handheld vibrational spectroscopy sensors. *Chinese Journal of Analytical Chemistry*, 50(4), 100064. https://doi.org/10.1016/j.cjac.2022.100064

Ayvaz, H., Temizkan, R., Kaya, B., Salman, M., Menevseoglu, A., Ayvaz, Z., Gunes, N., Dogan, M.A. and Mortas, M. (2024). Machine Learning-Assisted Near- and Mid-Infrared spectroscopy for rapid discrimination of wild and farmed Mediterranean mussels (*Mytilus galloprovincialis*). *Microchemical Journal*, 196, 109669. https://doi.org/10.1016/j.microc.2023.109669

Babellahi, F., Paliwal, J., Erkinbaev, C., Amodio, M.L., Chaudhry, M.M.A. and Colelli, G. (2020). Early detection of chilling injury in green bell peppers by hyperspectral imaging and chemometrics. *Postharvest Biology and Technology*, 162, 111100. https://doi.org/10.1016/j.postharvbio.2019.111100

Bai, X., Zhang, C., Xiao, Q., He, Y. and Bao, Y. (2020). Application of near-infrared hyperspectral imaging to identify a variety of silage maize seeds and common maize seeds. *RSC Advances*, 10(20), 11707-11715. https://doi.org/10.1039/c9ra11047j

Bai, Y., Xiong, Y., Huang, J., Zhou, J. and Zhang, B. (2019). Accurate prediction of soluble solid content of apples from multiple geographical regions by combining deep learning with spectral fingerprint features. *Postharvest Biology and Technology*, 156, 110943. https://doi.org/10.1016/j.postharvbio.2019.110943

Bao, Y., Mi, C., Wu, N., Liu, F. and He, Y. (2019). Rapid classification of wheat grain varieties using hyperspectral imaging and chemometrics. *Applied Sciences*, 9(19), 4119. https://doi.org/10.3390/app9194119

Barbon Jr, S., Costa Barbon, A.P.A.D., Mantovani, R.G. and Barbin, D.F. (2018). Machine learning applied to near-infrared spectra for chicken meat classification. *Journal of Spectroscopy*, 2018(1), 8949741. https://doi.org/10.1155/2018/8949741

Becchi, P.P., Rocchetti, G., García-Pérez, P., Michelini, S., Pizzamiglio, V. and Lucini, L. (2024). Untargeted metabolomics and machine learning unveil quality and authenticity interactions in grated Parmigiano Reggiano PDO cheese. *Food Chemistry*, 447, 138938. https://doi.org/10.1016/j.foodchem.2024.138938

Behkami, S., Zain, S.M., Gholami, M. and Khir, M.F.A. (2019). Classification of cow milk using artificial neural network developed from the spectral data of single-and three-detector spectrophotometers. *Food Chemistry*, 294, 309-315. https://doi.org/10.1016/j.foodchem.2019.05.060

Bellou, E., Gyftokostas, N., Stefas, D., Gazeli, O. and Couris, S. (2020). Laser-induced breakdown spectroscopy assisted by machine learning for olive oils classification: The effect of the experimental parameters. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 163, 105746. https://doi.org/10.1016/j.sab.2019.105746

Berahmand, K., Daneshfar, F., Salehi, E.S., Li, Y. and Xu, Y. (2024). Autoencoders and their applications in machine learning: a survey. *Artificial Intelligence Review*, 57(2). https://doi.org/10.1007/s10462-023-10662-6

Bhat, A.A., Chen, Y.W., Liang, Y.S., Chen, S.Y., Hsieh, C.W., Lin, C.M., Chen, M.H., Huang, P.H., Shih, M.K. and Hou, C.Y. (2025). Exploring the Potential of Microbubble Innovation in Revolutionizing Food Industry: A Comprehensive Review. *Food and Bioprocess Technology*, 18(5), 4175-4194. https://doi.org/10.1007/s11947-024-03697-2

Bhat, M.A., Rather, M.Y., Singh, P., Hassan, S. and Hussain, N. (2024). Advances in Smart Food Authentication for Enhanced Safety and Quality. *Trends in Food Science and Technology*, 155, 104800. https://doi.org/10.1016/j.tifs.2024.104800

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

Bisutti, V., Merlanti, R., Serva, L., Lucatello, L., Mirisola, M., Balzan, S., Tenti, Fontana, F., Trevisan, G., Montanucci, L., Contiero, B., Segato, S. and Capolongo, F. (2019). Multivariate and machine learning approaches for honey botanical origin authentication using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 27 (1), 65-74. https://doi.org/10.1177/0967033518824765

Boateng, A.A., Sumaila, S., Lartey, M., Oppong, M.B., Opuni, K.F. and Adutwum, L.A. (2022). Evaluation of chemometric classification and regression models for the detection of syrup adulteration in honey. *LWT-Food Science and Technology*, 163, 113498. https://doi.org/10.1016/j.lwt.2022.113498

Botoran, O.R., Ionete, R.E., Miricioiu, M.G., Costinel, D., Radu, G.L. and Popescu, R. (2019). Amino acid profile of fruits as potential fingerprints of varietal origin. *Molecules*, 24(24), 4500. https://doi.org/10.3390/molecules24244500

Bowler, A., Ozturk, S., Bari, V., Glover, Z.J. and Watson, N.J. (2023). Machine learning and domain adaptation to monitor yoghurt fermentation using ultrasonic measurements. *Food Control*, 147, 109622. https://doi.org/10.1016/j.foodcont.2023.109622

Brar, D.S., Aggarwal, A.K., Nanda, V., Kaur, S., Saxena, S. and Gautam, S. (2024). Detection of sugar syrup adulteration in unifloral honey using deep learning framework: An effective quality analysis technique. *Food and Humanity*, 2, 100190. https://doi.org/10.1016/

j.foohum.2023.11.017

Brendel, R., Schwolow, S., Rohn, S. and Weller, P. (2021). Volatilomic Profiling of Citrus Juices by Dual-Detection HS-GC-MSIMS and Machine Learning-An Alternative Authentication Approach. *Journal of Agricultural and Food Chemistry*, 69, 1727-1738. https://dx.doi.org/10.1021/acs.jafc.0c07447?ref=pdf

Brigante, F.I., Mas, A.L., Pigni, N.B., Wunderlin, D.A. and Baroni, M.V. (2020). Targeted metabolomics to assess the authenticity of bakery products containing chia, sesame and flax seeds. *Food Chemistry*, 312, 126059. https://doi.org/10.1016/j.foodchem.2019.126059

Brooks, C., Parr, L., Smith, J.M., Buchanan, D., Snioch, D. and Hebishy, E. (2021). A review of food fraud and food authenticity across the food supply chain, with an examination of the impact of the COVID-19 pandemic and Brexit on food industry. *Food Control*, 130, 108171. https://doi.org/10.1016/j.foodcont.2021.108171

Calle, J.L.P., Ferreiro-González, M., Ruiz-Rodríguez, A., Fernández, D. and Palma, M. (2022) Detection of Adulterations in Fruit Juices Using Machine Learning Methods over FT-IR Spectroscopic Data. *Agronomy*, 12, 683. https://doi.org/10.3390/agronomy12030683

Calle, J.L.P., Punta-Sánchez, I., González-de-Peredo, A.V., Ruiz-Rodríguez, A., Ferreiro-González, M. and Palma, M. (2023) Rapid and Automated Method for Detecting and Quantifying Adulterations in High-Quality Honey Using Vis-NIRs in Combination with Machine Learning. *Foods*, 12, 2491. https://doi.org/10.3390/foods12132491

Campmajó, G., Navarro, G.J., Núñez, N., Puignou, L., Saurina, J. and Núñez, O. (2019). Non-targeted HPLC-UV fingerprinting as chemical descriptors for the classification and authentication of nuts by multivariate chemometric methods. *Sensors*, 19(6), 1388. https://doi.org/10.3390/s19061388

Centonze, V., Lippolis, V., Cervellieri, S., Damascelli, A., Casiello, G., Pascale, M., Logrieco, A.F. and Longobardi, F. (2019). Discrimination of geographical origin of oranges (*Citrus sinensis* L. Osbeck) by mass spectrometry -based electronic nose and characterization of volatile compounds. *Food Chemistry*, 277, 25-30. https://doi.org/10.1016/j.foodchem.2018.10.105

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. and Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215. https://doi.org/10.1016/j.neucom.2019.10.118

Ch, R., Chevallier, O., McCarron, P., McGrath, T.F., Wu, D., Duy, L.N.D., Kapil, A.P., McBride, M. and Elliott, C.T. (2021). Metabolomic fingerprinting of volatile organic compounds for the geographical discrimination of rice samples from China, Vietnam and India. *Food Chemistry*, 334, 127553. https://doi.org/10.1016/j.foodchem.2020.127553

Chakraborty, S.K., Mahanti, N.K., Mansuri, S.M., Tripathi, M.K., Kotwaliwale, N. and Jayas, D. S. (2021). Non-destructive classification and prediction of aflatoxin-B1 concentration in maize kernels using Vis–NIR (400–1000 nm) hyperspectral imaging. *Journal of Food Science and Technology*, 58, 437-450. https://doi.org/10.1007/s13197-020-04552-w

Chaplot, N., Pandey, D., Kumar, Y. and Sisodia, P.S. (2023). A comprehensive analysis of artificial intelligence techniques for the prediction and prognosis of genetic disorders using various gene disorders. *Archives of Computational Methods in Engineering*, 30(5), 3301-3323. https://doi.org/10.1007/s11831-023-09904-1

Chaudhary, V., Kajla, P., Dewan, A., Pandiselvam, R., Socol, C.T. and Maerescu, C.M. (2022). Spectroscopic techniques for authentication of animal origin foods. *Frontiers in Nutrition*, 9, 979205. https://doi.org/10.3389/fnut.2022.979205

Cheiran, K.P., Raimundo, V.P., Manfroi, V., Anzanello, M.J., Kahmann, A., Rodrigues, E. and Frazzon, J. (2019). Simultaneous identification of low-molecular weight phenolic and nitrogen compounds in craft beers by HPLC-ESI-MS/MS. *Food Chemistry*, 286, 113-122. https://doi.org/10.1016/j.foodchem.2019.01.198

Chen, H., Tan, C. and Lin, Z. (2019c). Non-destructive identification of native egg by near-infrared spectroscopy and data driven-based class-modeling. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 206, 484-490. https://doi.org/10.1016/j.saa.2018.08.041

Chen, K., Xue, H., Shi, Q., Zhang, F., Ma, Q., Sun, J., Liu, Y., Tang, Y. and Wang, W. (2024). Geographical identification of Chinese wine based on chemometrics combined with mineral elements, volatile components and untargeted metabonomics. *Food Chemistry: X*, 22, 101412. https://doi.org/10.1016/j.fochx.2024.101412

Chen, L., Zhang, H.W., Zhang, X.M., Yu, F.Q.H., Zhang, F., Xue, C.H., Xue, Y., Tang, Q.J. and Li, Z.J. (2020). Identification of potential peptide markers for the shelf-life of Pacific oysters (*Crassostrea gigas*) during anhydrous preservation via mass spectrometry-based peptidomics. *LWT-Food Science and Technology*, 134, 109922. https://doi.org/10.1016/j.lwt.2020.109922.

Chen, S., Xiong, J., Guo, W., Bu, R., Zheng, Z., Chen, Y., Yang, Z. and Lin, R. (2019a). Colored rice quality inspection system using machine vision. *Journal of Cereal Science*, 88, 87-95. https://doi.org/10.1016/j.jcs.2019.05.010

Chen, W., Krainak, D., Sahiner, B. and Petrick, N. (2023). A regulatory science perspective on performance assessment of machine learning algorithms in imaging. *Machine Learning for Brain Disorders*, 705-752. https://doi.org/10.1007/978-1-0716-3195-9_23

Chen, Z., Wu, T., Xiang, C., Xu, X. and Tian, X. (2019b). Rapid identification of rainbow trout adulteration in Atlantic salmon by Raman spectroscopy combined with machine learning. *Molecules*, 24(15), 2851. https://doi.org/10.3390/molecules24152851

Chhetri, K.B. (2024). Applications of Artificial Intelligence

REVIEW

and Machine Learning in Food Quality Control and Safety Assessment. *Food Engineering Reviews*, 16(1), 1-21. https://doi.org/10.1007/s12393-023-09363-1

Chin, S.T., Hoerlendsberger, G., Wong, K.W., Li, S., Bong, S.H., Whiley, L., Wist, J., Masuda, R., Greeff, J., Holmes, E., Nicholson, J.K. and Loo, R.L. (2023). Targeted lipidomics coupled with machine learning for authenticating the provenance of chicken eggs, *Food Chemistry*, 410, 135366. https://doi.org/10.1016/j.foodchem.2022.135366

Choudhary, G. and Sethi, D. (2023). From conventional approach to machine learning and deep learning approach: an experimental and comprehensive review of image fusion techniques. *Archives of Computational Methods in Engineering*, 30(2), 1267-1304. https://doi.org/10.1007/s11831-022-09833-5

Chu, C., Wang, H., Luo, X., Wen, P., Nan, L., Du, C., Fan, Y., Gao, D., Wang, D., Yang, Z. and Yang, G. (2023). Possible alternatives: Identifying and quantifying adulteration in buffalo, goat, and camel milk using mid-infrared spectroscopy combined with modern statistical machine learning methods. *Foods*, 12(20), 3856. https://doi.org/10.3390/foods12203856

Chu, Y., Wu, J., Yan, Z., Zhao, Z., Xu, D. and Wu, H. (2024). Towards generalizable food source identification: An explainable deep learning approach to rice authentication employing stable isotope and elemental marker analysis. *Food Research International*, 179, 113967. https://doi.org/10.1016/j.foodres.2024.113967

Chung, I.M., Kim, J.K., An, Y.J., Kwon, C., Kim, S.Y., Yang, Y.J., Yarnes, C., Chi, H.Y. and Kim, S.H. (2019). Compound-specific δ13C and δ15N analyses of fatty acids and amino acids for discrimination of organic, pesticide-free, and conventional rice (*Oryza sativa* L.). *Food Chemistry*, 283, 305-314. https://doi.org/10.1016/j.foodchem.2018.12.129

Chung, I.M., Kim, J.K., Yarnes, C.T., An, Y.J., Kwon, C., Kim, S.Y., Yang, Y.J., Chi, H.Y. and Kim, S.H. (2018). Fatty acid-and amino acid-specific isotope analysis for accurate authentication and traceability in organic milk. *Journal of Agricultural and Food Chemistry*, 67(2), 711-722. https://doi.org/10.1021/acs.jafc.8b05063

Chung, I.M., Kim, Y.J., Moon, H.S., Chi, H.Y. and Kim, S.H. (2021). Long-term isotopic model study for ecofriendly rice (*Oryza sativa* L.) authentication: updating a case study in South Korea. *Food Chemistry*, 362, 130215. https://doi.org/10.1016/j.foodchem.2021.130215

Chung, T., Tam, I.Y.S., Lam, N.Y.Y., Yang, Y., Liu, B., He, B., Li, W., Xu, J., Yang, Z., Zhang, L., Cao, J.N. and Lau, L.T. (2022). Non-targeted detection of food adulteration using an ensemble machine-learning model. *Scientific Reports*, 12(1), 20956. https://doi.org/10.1038/s41598-022-25452-3

Cianciosi, D., Forbes-Hernández, T.Y., Afrin, S., Gasparrini, M., Reboredo-Rodriguez, P., Manna, P.P., Zhang, J., Bravo Lamas, L., Martínez Flórez, S., Agudo Toyos, P. and Quiles, J.L. (2018). Phenolic compounds in honey and their associated health benefits: A review. *Molecules*, 23 (9), 2322. https://doi.org/10.3390/molecules23092322

Ciftci, H.G. and Aytac, E. (2024). Monitoring the quality changes and freshness assessment of eggs based on chemometrics method. *Journal of Food Process Engineering*, 47(3), e14591. https://doi.org/10.1111/jfpe.14591

Cobbinah, E., Generalao, O.B., Ke, G., Malaluan, R., Lubguban, A. and Dumancas, G.G. (2024). A rapid analytical method for turmeric essential oil authentication using mid-infrared spectroscopy and chemometrics. *Journal of Food Composition and Analysis*, 129, 106102. https://doi.org/10.1016/j.jfca.2024.106102

Coelho, E., Magalhães, J., Pereira, F.B., Macieira, F., Domingues, L. and Oliveira, J.M. (2019). Volatile fingerprinting differentiates diverse-aged craft beers. *LWT - Food Science and Technology*, 108, 129-136. https://doi.org/10.1016/j.lwt.2019.03.044

Costa, N.L.D., Valentin, L.A., Castro, I.A. and Barbosa, R.M. (2021). Predictive modeling for wine authenticity using a machine learning approach. *Artificial Intelligence in Agriculture*, 5, 157-162. https://doi.org/10.1016/j.aiia.2021.07.001

Costa, V.G. and Pedreira, C.E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, 56(5), 4765-4800. https://doi.org/10.1007/s10462-022-10275-5

Cruz-Tirado, J.P., Fernandez Pierna, J.A., Rogez, H., Barbin, D.F. and Baeten, V. (2020) Authentication of Cocoa (*Theobroma cacao*) Bean Hybrids by NIR-Hyperspectral Imaging and Chemometrics. *Food Control*. 118, 107445. https://doi.org/10.1016/j.foodcont.2020.107445

Cruz-Tirado, J.P., Oliveira, M., de Jesus Filho, M., Godoy, H.T., Amigo, J.M. and Barbin, D.F. (2021). Shelf-Life Estimation and Kinetic Degradation Modeling of Chia Seeds (*Salvia hispanica*) Using Principal Component Analysis Based on NIR-Hyperspectral Imaging. *Food Control.* 123, 107777. https://doi.org/10.1016/j.foodcont.2020.107777

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

Cui, F., Zheng, S., Wang, D., Ren, L., Meng, Y., Ma, R., Wang, S., Li, X., Li, T. and Li, J. (2024). Development of machine learning-based shelf-life prediction models for multiple marine fish species and construction of a real-time prediction platform. *Food Chemistry*, 450, 139230. https://doi.org/10.1016/j.foodchem.2024.139230

Dais, D., Bal, I.E., Smyrou, E. and Sarhosis, V. (2021). Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. *Automation in Construction*, 125, 103606. https://doi.org/10.1016/j.autcon.2021.103606

Dalal, N., Sáiz, M.J., Caporale, A.G., Baldini, F., Babayan, S.A. and Adamo, P. (2024). Fishy forensics: FT-NIR and machine learning based authentication of Mediterranean anchovies (*Engraulis encrasicolus*). *Journal of Food Composition and Analysis*, 136, 106847. http://dx.doi.org/10.2139/ssrn.4761744

Dargan, S., Kumar, M., Ayyagari, M.R. and Kumar, G. (2020). A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 27, 1071-1092. https://doi.org/10.1007/s11831-019-09344-w

de Lima, A.B.S., Batista, A.S., de Jesus, J.C., de Jesus Silva, J., de Araújo, A.C.M. and Santos, L.S. (2020). Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling. *Food Control*, 107, 106802. https://doi.org/10.1016/j.foodcont.2019.106802

de Souza, R.R., de Sousa Fernandes, D.D. and Diniz, P.H.G.D. (2021). Honey authentication in terms of its adulteration with sugar syrups using UV–Vis spectroscopy and one-class classifiers. *Food Chemistry*, 365, 130467. https://doi.org/10.1016/j.foodchem.2021.130467

Della Porta, R.A. (2006). AOCS/SFA Edible Oils Manual. AOCS Publishing. Retrieved from www.amazon.com/Edible-Manual-Second-Richard-Paperback/dp/B011YTXNHC

Deng, P., Lin, X., Yu, Z., Huang, Y., Yuan, S., Jiang, X., Niu, M. and Peng, W.K. (2024a). Machine learning-enabled high-throughput industry screening of edible oils, *Food Chemistry*, 447, 139017. https://doi.org/10.1016/j.foodchem.2024.139017

Deng, Z., Fu, J., Yang, M., Zhang, W., Yun, Y.H. and Zhang, L. (2024b). Geographical origin identification of *Hainan camellia* oil based on fatty acid composition and near infrared spectroscopy combined with chemometrics. *Journal of Food Composition and Analysis*, 125, 105730. https://doi.org/10.1016/j.jfca.2023.105730

Derz, W., Pavlovic, M., Huber, I., Schalch, B. and Gerdes, L. (2021). Food fraud in the Alps?-Detection of chamois (*Rupicapra rupicapra*) in firm raw sausages, ham, and meat via qualitative duplex real-time PCR. *Food Control*, 123, 107764. https://doi.org/10.1016/j.foodcont.2020.107764

Dewan, A., Chhikara, N. and Khatkar, B.S. (2024b). Evaluating dough and gluten strength in wheat varieties: a combined qualitative and quantitative approach. *Agricultural Research Journal*, 61(3), 420-431. https://doi.org/10.5958/2395-146X.2024.00054.5

Dewan, A., Khatkar, B.S., Bangar, S.P., Chaudhary, V. and Lorenzo, J.M. (2022). Infrared spectroscopy and microstructural assessment of dough with varying wheat gluten fractions. *Food Analytical Methods*, 15(10), 2821-2829. https://doi.org/10.1007/s12161-022-02331-1

Dewan, A., Malik, M., Chaudhary, V. and Sharma, A. (2024a). Packaging Applications of Cocoa Pod Husk. In Agro-Wastes for Packaging Applications (pp. 264-287). CRC Press. https://doi.org/10.1201/9781003453277-12

Dewan, A., Nagaraja, S.K., Yadav, S., Bishnoi, P., Malik, M., Chhikara, N., Luthra, L., Singh. S., Davis, C. and Poonam. (2025). Advances in Wine Processing: Current Insights, Prospects, and Technological Interventions. *Food and Bioprocess Technology*, 18(6), 5058-5093. https://doi.org/10.1007/s11947-025-03774-0

Dimakopoulou-Papazoglou, D., Ploskas, N., Serrano, S., Silva, C.S., Valdramidis, V., Koutsoumanis, K. and Katsanidis, E. (2023). Application of UV–Vis spectroscopy for the detection of adulteration in Mediterranean honeys. *European Food Research and Technology*, 249 (12), 3043-3053. https://doi.org/10.1007/s00217-023-04347-1

Drabova, L., Alvarez-Rivera, G., Suchanova, M., Schusterova, D., Pulkrabova, J., Tomaniova, M., Kocourek, V., Chevallier, O., Elliott, C. and Hajslova, J. (2019). Food fraud in oregano: Pesticide residues as adulteration markers. *Food Chemistry*, 276, 726-734. https://doi.org/10.1016/j.foodchem.2018.09.143

Drees, A., Bockmayr, B., Bockmayr, M. and Fischer, M. (2023). Rapid Determination of Nutmeg Shell Content in Ground Nutmeg Using FT-NIR Spectroscopy and Machine Learning. *Foods*, 12, 2939. https://doi.org/10.3390/foods12152939

Dubey, A.K., Gupta, U. and Jain, S. (2016). Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. *International Journal of Computer Assisted Radiology and Surgery*, 11, 2033-2047. https://doi.org/10.1007/s11548-016-1437-9

Duraipandian, S., Petersen, J.C. and Lassen, M. (2019). Authenticity and concentration analysis of extra virgin olive oil using spontaneous Raman spectroscopy and multivariate data analysis. *Applied Sciences*, 9(12), 2433. https://doi.org/10.3390/app9122433

Ebrahimi, I., Castro, R.D., Ehsani, R., Brillante, L. and Feng, S. (2024). Advancing grape chemical analysis through machine learning and multi-sensor spectroscopy. *Journal of Agriculture and Food Research*, 16, 101085. https://doi.org/10.1016/j.jafr.2024.101085

El Orche, A., Mamad, A., Elhamdaoui, O., Cheikh, A., El Karbane, M. and Bouatia, M. (2021). Comparison of machine learning classification methods for determining the geographical origin of raw milk using vibrational spectroscopy. *Journal of Spectroscopy*, 2021(1), 5845422. https://doi.org/10.1155/2021/5845422

ElHaj, K., Alshamsi, D. and Aldahan, A. (2023). GeoZ: a region-based visualization of clustering algorithms. *Journal of Geovisualization and Spatial Analysis*, 7(1), 15. https://doi.org/10.1007/s41651-023-00146-0

Elhamdaoui, O., El Orche, A., Laouni, A., Alami, L., El Karbane, M., Cheikh, A., Karrouchi, K., Laarej, K., El Otmani, I.S. and Bouatia, M. (2024). Spectral insights: advancing the authentication and quality assessment of Moroccan saffron through UV–visible spectroscopy and multivariate chemometric methods. *Food Measure* 18, 2428–2437. https://doi.org/10.1007/s11694-023-02319-3

Endo, Y. (2018). Analytical methods to evaluate the quality of edible fats and oils: the JOCS standard methods for analysis of fats, oils and related materials (2013) and advanced methods. *Journal of Oleo Science*, 67(1), 1-10. https://doi.org/10.5650/jos.ess17130

Erban, A., Fehrle, I., Martinez-Seidel, F., Brigante, F., Más, A.L., Baroni, V., Wunderlin, D. and Kopka, J. (2019) Discovery of food identity markers by metabolomics and machine learning technology. *Scientific Reports,* 9(1), 9697. https://doi.org/10.1038/s41598-019-46113-y

Esteki, M., Simal-Gandara, J., Shahsavari, Z., Zandbaaf, S., Dashtaki, E. and Vander Heyden, Y. (2018). A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. *Food Control*, 93, 165-182. https://doi.org/10.1016/j.foodcont.2018.06.015

Fakhlaei, R., Selamat, J., Khatib, A., Razis, A.F.A., Sukor, R., Ahmad, S. and Babadi, A.A. (2020). The toxic impact of honey adulteration: A review. *Foods*, 9(11), 1538. https://doi.org/10.3390/foods9111538

Faqe Ibrahim, G.R., Rasul, A. and Abdullah, H. (2023). Improving crop classification accuracy with integrated Sentinel-1 and Sentinel-2 data: a case study of barley and wheat. *Journal of Geovisualization and Spatial Analysis*, 7(2), 22. https://doi.org/10.1007/s41651-023-00152-2

Farooq, F., Ahmed, W., Akbar, A., Aslam, F. and Alyousef, R. (2021). Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *Journal of Cleaner Production*, 292, 126032. https://doi.org/10.1016/j.jclepro.2021.126032

Fattahi, S.H., Kazemi, A. and Seyfari, Y. (2024b). A New Detection Method of Safflower Adulteration in Saffron (Red Gold) Using a Combination of Feature Selection Algorithm and Machine Learning. Retrieved from SSRN: https://ssrn.com/abstract=4806906 or http://dx.doi.org/10.2139/ssrn.4806906

Fattahi, S.H., Kazemi, A., Khojastehnazhand, M., Roostaei, M. and Mahmoudi, A. (2024a). The classification of Iranian wheat flour varieties using FT-MIR spectroscopy and chemometrics methods. *Expert Systems with Applications*, 239, 122175. https://doi.org/10.1016/j.eswa.2023.122175

Feng, L., Zhu, S., Zhou, L., Zhao, Y., Bao, Y., Zhang, C. and He, Y. (2019). Detection of subtle bruises on winter jujube using hyperspectral imaging with pixel-wise deep learning method. *IEEE Access*, *7*, 64494-64505. https://doi.org/10.1109/ACCESS.2019.2917267

Feng, X., Ma, G., Su, S.F., Huang, C., Boswell, M.K. and Xue, P. (2020). A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan. *Ocean Engineering*, 211, 107526. https://doi.org/10.1016/j.oceaneng.2020.107526

Feng, Z., Liu, D., Gu, J. and Zheng, L. (2024). Raman spectroscopy and fusion machine learning algorithm: A novel approach to identify dairy fraud. *Journal of Food Composition and Analysis*, 129, 106090. https://doi.org/10.1016/j.jfca.2024.106090

Fernando, I., Fei, J., Cahoon, S. and Close, D.C. (2024). A review of the emerging technologies and systems to mitigate food fraud in supply chains. *Critical Reviews in Food Science and Nutrition*, 1-28. https://doi.org/10.1080/10408398.2024.2405840

Florian-Huaman, J., Cruz-Tirado, J.P., Fernandes Barbin, D. and Siche, R. (2022). Detection of Nutshells in Cumin Powder Using NIR Hyperspectral Imaging and Chemometrics Tools. *Journal of Food Composition and Analysis*, 108, 104407. https://doi.org/10.1016/j.jfca.2022.104407

Frizzarin, M., O'Callaghan, T.F., Murphy, T.B., Hennessy, D. and Casa, A. (2021). Application of machine-learning methods to milk mid-infrared spectra for discrimination of cow milk from pasture or total mixed ration diets. *Journal of Dairy Science*, 104(12), 12394-12402. https://doi.org/10.3168/jds.2021-20812

Fröhlich-Wyder, M.T., Bachmann, H.P. and Schmidt, R.S. (2023). Classification of cheese varieties from Switzerland using machine learning methods: Free volatile carboxylic acids. *LWT-Food Science and Technology*, 184, 115095. https://doi.org/10.1016/j.lwt.2023.115095

Fukui, S., Takayama, T., Toyo'oka, T., Mizuno, H. and Todoroki, K. (2019). An accurate differential analysis of carboxylic acids in beer using ultra high-performance liquid chromatography–quadrupole time-of-flight mass spectrometry based on chiral derivatization combining three isotopic reagents. *Talanta*, 205, 120146. https://doi.org/10.1016/j.talanta.2019.120146

Galvin-King, P., Haughey, S.A. and Elliott, C.T. (2021). Garlic adulteration detection using NIR and FTIR spectroscopy and chemometrics. *Journal of Food Composition and Analysis*, 96, 103757. https://doi.org/10.1016/j.jfca.2020.103757

Gao, Y.F., Li, X.Y., Wang, Q.L., Li, Z.H., Chi, S.X., Dong, Y., Guo, L. and Zhang, Y.H. (2024). Discrimination and quantification of volatile compounds in beer by FTIR combined with machine learning approaches. *Food Chemistry: X*, 22, 101300. https://doi.org/10.1016/j.fochx.2024.101300

García, N.L. (2018). The current situation on the international honey market. *Bee World*, 95(3), 89–94. https://doi.org/10.1080/0005772x.2018.1483814

Gatzert, X., Chun, K.P., Boner, M., Hermanowski, R., Mäder, R., Breuer, L., Gattinger, A. and Orlowski, N. (2021). Assessment of multiple stable isotopes for tracking regional and organic authenticity of plant products in Hesse, Germany. *Isotopes in Environmental and Health Studies*, 57(3), 281-300. https://doi.org/10.1080/10256016.2021.1905635

Gazeli, O., Bellou, E., Stefas, D. and Couris, S. (2020). Laser-based classification of olive oils assisted by machine learning. *Food Chemistry*, 302, 125329. https://doi.org/10.1016/j.foodchem.2019.125329

Ge, T., Luo, X., Wang, Y., Sedlmair, M., Cheng, Z., Zhao, Y., Liu, X., Deussen, O. and Chen, B. (2023). Optimally ordered orthogonal neighbor joining trees for hierarchical cluster analysis. *IEEE Transactions on Visualization and Computer Graphics*, 30(8), 5034-5046. https://doi.org/10.1109/TVCG.2023.3284499

Gerhardt, N., Schwolow, S., Rohn, S., Pérez-Cacho, P.R., Galán-Soldevilla, H., Arce, L. and Weller, P. (2019). Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: Comparison of different processing approaches by LDA, kNN, and SVM. *Food Chemistry*, 278, 720-728. https://doi.org/10.1016/j.foodchem.2018.11.095

Giannetti, V., Mariani, M.B., Torrelli, P. and Marini, F. (2019). Flavour component analysis by HS-SPME/GC–MS and chemometric modeling to characterize Pilsner-style Lager craft beers. *Microchemical Journal*, 149, 103991. https://doi.org/10.1016/j.microc.2019.103991

Giannuzzi, D., Mota, L.F.M., Pegolo, S., Tagliapietra, F., Schiavon, S., Gallo, L., Marsan, P.A., Trevisi, E. and Cecchinato, A. (2023). Prediction of detailed blood metabolic profile using milk infrared spectra and machine learning methods in dairy cattle. *Journal of Dairy Science*, 106(5), 3321-3344. https://doi.org/10.3168/jds.2022-22454

Golzarijalal, M., Ong, L., Neoh, C.R., Harvie, D.J.E. and Gras, S.L. (2024). Machine learning for the prediction of proteolysis in Mozzarella and Cheddar cheese. *Food and Bioproducts Processing*, 144, 32-144. https://doi.org/10.1016/j.fbp.2024.01.009

Gomes, C., Yang, H. and Jin, Z. (2021). Insurance fraud detection with unsupervised deep learning. *Journal of Risk and Insurance*, 88(3), 591–624. https://doi.org/10.1111/jori.12359

Gottstein, V., Lachenmeier, D.W., Kuballa, T. and Bunzel, M. (2024). 1H NMR-based approach to determine the geographical origin and cultivation method of roasted coffee. *Food Chemistry*, 433, 137278. https://doi.org/10.1016/j.foodchem.2023.137278

Goyal, K., Kumar, P. and Verma, K. (2022). Food adulteration detection using artificial intelligence: A systematic review. *Archives of Computational Methods in Engineering*, 29(1), 397-426. https://doi.org/10.1007/s11831-021-09600-y

Grace, D. (2019). Food fraud. In: Ferranti, P., Berry, E.M. and Anderson, J.R. (eds), *Encyclopedia Food Security and Sustainability,*1, 238–248. https://doi.org/10.1016/B978-0-08-100596-5.21577-1

Granato, D., Santos, J.S., Escher, G.B., Ferreira, B.L. and Maggio, R.M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends in Food Science and Technology*, 72, 83-90. https://doi.org/10.1016/j.tifs.2017.12.006

Grocery Manufacturers Association (GMA). (2017). Five Guidelines of the Global Food Supply Chain. Retrieved from https://www.supplychain247.com/search/results/search?keywords=food+fraud+cost (Data retrieved on 30 March 2025)

Gu, H.W., Zhou, H.H., Lv, Y., Wu, Q., Pan, Y., Peng, Z.X., Zhang, X.H. and Yin, X.L. (2023). Geographical origin identification of Chinese red wines using ultraviolet-visible spectroscopy coupled with machine learning techniques. *Journal of Food Composition and Analysis*, 119, 105265. https://doi.org/10.1016/j.jfca.2023.105265

Guellis, C., Valério, D.C., Bessegato, G.G., Boroski, M., Dragunski, J.C. and Lindino, C.A. (2020). Non-targeted method to detect honey adulteration: Combination of electrochemical and spectrophotometric responses with principal component analysis. *Journal of Food Composition and Analysis*, 89, 103466. https://doi.org/10.1016/j.jfca.2020.103466

Gul, N., Ahmad Shah, S.Z., Dar, B.N., Makroo, H., Assad, A. and Muzaffar, K. (2024). Deep learning hyperspectral imaging: a rapid and reliable alternative to conventional techniques in the testing of food quality and safety. *Quality Assurance and Safety of Crops and Foods*, 16(1), 78–97. https://doi.org/10.15586/qas.v16i1.1392

Gyftokostas, N., Nanou, E., Stefas, D., Kokkinos, V., Bouras, C. and Couris, S. (2021a). Classification of Greek olive oils from different regions by machine learning-aided laser-induced breakdown spectroscopy and absorption spectroscopy. *Molecules*, 26(5), 1241. https://doi.org/10.3390/molecules26051241

Gyftokostas, N., Stefas, D. and Couris, S. (2020). Olive oils classification via laser-induced breakdown spectroscopy. *Applied Sciences*, 10(10), 3462. https://doi.org/10.3390/app10103462

Gyftokostas, N., Stefas, D. Kokkinos, V. (2021b). Laser-induced breakdown spectroscopy coupled with machine learning as a tool for olive oil authenticity and geographic discrimination. *Scientific Reports,* 11, 5360. https://doi.org/10.1038/s41598-021-84941-z

Hajikhani, M., Hegde, A., Snyder, J., Cheng, J. and Lin, M. (2024). Integrating transformer-based machine learning with SERS technology for the analysis of hazardous pesticides in spinach. *Journal of Hazardous Materials*, 470, 134208. https://doi.org/10.1016/j.jhazmat.2024.134208

Hamzić Gregorčič, S., Potočnik, D., Camin, F. and Ogrinc, N. (2020). Milk authentication: Stable isotope composition of hydrogen and oxygen in milks and their constituents. *Molecules*, 25(17), 4000. https://doi.org/10.3390/molecules25174000

Han, F., Aheto, J.H., Rashed, M.M.A. and Zhang, X. (2022b). Machine-learning assisted modelling of multiple elements for authenticating edible animal blood food, *Food Chemistry: X*, 14, 100280. https://doi.org/10.1016/j.fochx.2022.100280

Han, J., Li, T., He, Y. and Gao, Q. (2022a). Using Machine Learning Approaches for Food Quality Detection. *Mathematical Problems in Engineering,* Volume 2022, Article ID 6852022. https://doi.org/10.1155/2022/6852022

REVIEW

REVIEW

Hategan, A.R., David, M., Dehelean, A., Cristea, G., Puscas, R., Molnar, A.J. and Magdas, D.A. (2023). Impact of Pre-Processing Methods for the Identification of the Botanical Origin of Honey Based Upon Isotopic and Elemental Profiles. *Analytical Letters*, 56(2), 231–243. https://doi.org/10.1080/00032719.2022.2044347

Hategan, A.R., Puscas, R., Cristea, G., Dehelean, A., Guyon, F., Molnar, A.J., Mirel, V. and Magdas, D.A. (2021). Opportunities and Constraints in Applying Artificial Neural Networks (ANNs) in Food Authentication. Honey—A Case Study. *Applied Sciences*, 11(15), 6723. https://doi.org/ 10.3390/app11156723

He, G., Yang, S.B. and Wang, Y.Z. (2024). A rapid method for identification of *Lanxangia tsaoko* origin and fruit shape: FT-NIR combined with chemometrics and image recognition. *Journal of Food Science*, 89(4), 2316-2331. https://doi.org/10.1111/1750-3841.16989

He, S., Sun, X., Du, M., Chen, H., Tan, M., Sun, H. and Zhu, B. (2019). Effects of muscle protein denaturation and water distribution on the quality of false abalone (*Volutharpa ampullaceal perryi*) during wet heating. *Journal of Food Process Engineering*, 42(1), e12932. https://doi.org/10.1111/jfpe.12932

Hemmati, A., Mahmoudi, A., Jamshidi, B. and Ghaffari, H. (2024). Assessment of Persian export pomegranate quality: A reliable non-destructive method based on spectroscopy and chemometrics. *Journal of Food Composition and Analysis*, 131, 106202. https://doi.org/10.1016/j.jfca.2024.106202

Heo, S., Choi, J.Y., Kim, J. and Moon, K.D. (2021). Prediction of moisture content in steamed and dried purple sweet potato using hyperspectral imaging analysis. *Food Science and Biotechnology*, 30(6), 783-791. https://doi.org/10.1007/s10068-021-00921-z

Hernández-Fernández, J., Martinez-Trespalacios, J. and Marquez, E. (224). Development of a Measurement System Using Infrared Spectroscopy-Attenuated Total Reflectance, Principal Component Analysis and Artificial Intelligence for the Safe Quantification of the Nucleating Agent Sorbitol in Food Packaging. *Foods*, 13, 1200. https://doi.org/10.3390/foods13081200

Hou, H., Tang, Y., Zhao, J., Debrah, A.A., Shen, Z., Li, C. and Du, Z. (2023). Authentication of organically produced cow milk by fatty acid profile combined with chemometrics: A case study in China. *Journal of Food Composition and Analysis*, 120, 105297. https://doi.org/10.1016/j.jfca.2023.105297

Houssein, E.H., Hammad, A. and Ali, A.A. (2022). Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review. *Neural Computing and Applications*, 34(15), 12527-12557. https://doi.org/10.1007/s00521-022-07292-4

Houston, J., Glavin, F.G. and Madden, M.G. (2020). Robust classification of high-dimensional spectroscopy data using deep learning and data synthesis. *Journal of Chemical Information and Modeling*, 60(4), 1936-1954. https://doi.org/10.1021/acs.jcim.9b01037

Hu, X.Z., Liu, S.Q., Li, X.H., Wang, C.X., Ni, X.L., Liu, X., Wang, Y., Liu, Y. and Xu, C.H. (2019). Geographical origin traceability of Cabernet Sauvignon wines based on Infrared fingerprint technology combined with chemometrics. *Scientific Reports*, 9(1), 8256. https://doi.org/10.1038/s41598-019-44521-8

Hu, Y., Chen, W., Gouda, M., Yao, H., Zuo, X., Yu, H., Zhang, Y., Ding, L., Zhu, F., Wang, Y., Li, X., Zhou, J. and He, Y. (2024). Fungal fermentation of Fuzhuan brick tea: A comprehensive evaluation of sensory properties using chemometrics, visible near-infrared spectroscopy, and electronic nose. *Food Research International*, 186, 114401. https://doi.org/10.1016/j.foodres.2024.114401

Hu, Z., Luo, Z.; Wang, Y., Zhou, Q., Liu, S. and Wang, Q. (2023). Texture Feature Extraction from ^{1}H NMR Spectra for the Geographical Origin Traceability of Chinese Yam. *Foods*, 12, 2476. https://doi.org/10.3390/foods12132476

Huang, C. and Gu, Y.A. (2022). Machine Learning Method for the Quantitative Detection of Adulterated Meat Using a MOS-Based E-Nose. *Foods*, 11, 602. https://doi.org/10.3390/foods11040602

Huang, L., Guo, S.Y., Wang, Y., Wang, S., Chu, Q.B., Li, L. and Bai, T. (2019). Attention based residual network for medicinal fungi near infrared spectroscopy analysis. *Mathematical Biosciences and Engineering*, 16 (4), 3003-3017. https://doi.org/10.3934/mbe.2019149

Imdoukh, M., Ahmad, I. and Alfailakawi, M.G. (2020). Machine learning-based auto-scaling for containerized applications. *Neural Computing and Applications*, 32(13), 9745-9760. https://doi.org/10.1007/s00521-019-04507-z

Issaad, F.Z., Abdessemed, A., Bouhedjar, K., Bouyahmed, H., Derdour, M., Ouffroukh, K., Fellak, A., Dems, M.A.S., Chihoub, S., Bechlem, R., Mahrouk, A., Houasnia, M., Belaidi, A., Moumed, K., Sebai, Z., Saidani, F. and Akmouche, H. (2024). Classification of Algerian olive oils: Physicochemical properties, polyphenols and fatty acid composition combined with machine learning models. *Journal of Food Composition and Analysis*, 125, 105812. https://doi.org/10.1016/j.jfca.2023.105812

Izquierdo, M., Lastra-Mejías, M., González-Flores, E., Pradana-López, S., Cancilla, J.C. and Torrecilla, J.S. (2020). Visible imaging to convolutionally discern and authenticate varieties of rice and their derived flours. *Food Control*, 110, 106971. https://doi.org/10.1016/j.foodcont.2019.106971

Jalal, N., Mehmood, A., Choi, G.S. and Ashraf, I. (2022). A novel improved random forest for text classification using feature ranking and optimal number of trees. *Journal of King Saud University-Computer and Information Sciences*, 34(6), 2733-2742. https://doi.org/10.1016/j.jksuci.2022.03.012

Jiménez-Carvelo, A.M., González-Casado, A., Bagur-González, M.G. and 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

Jin, W., Zhao, S., Sun, H., Pei, J., Gao, R. and Jiang, P. (2023). Characterization and discrimination of flavor volatiles of different colored wheat grains after cooking based on GC-IMS and chemometrics. *Current Research in Food Science*, 7, 100583. https://doi.org/10.1016/j.crfs.2023.100583

Jing, Q., Yingying, L., Chen, Z., Cheng, W., Juanqiang, W., Wenping, G. and Shouwei, W. (2021). Geographic origin discrimination of pork from different Chinese regions using mineral elements analysis assisted by machine learning techniques. *Food Chemistry*, 337, 127779. https://doi.org/10.1016/j.foodchem.2020.127779

Kabir, M.H., Guindo, M.L., Chen, R. and Liu, F. (2021). Geographic origin discrimination of millet using Vis-NIR spectroscopy combined with machine learning techniques. *Foods*, 10(11), 2767. https://doi.org/10.3390/foods10112767

Kaczala, S., De Lima, V.A. and Felsner, M.L. (2024). Authentication of Apple Juices Using Digital Image Processing (DIP) Combined with Machine Learning (ML). *Research Square*, 1-24. https://doi.org/10.21203/rs.3.rs-5373364/v1

Kalinichenko, A. and Arseniyeva, L. (2020). Electronic nose combined with chemometric approaches to assess authenticity and adulteration of sausages by soy protein. *Sensors and Actuators B: Chemical*, 303, 127250. https://doi.org/10.1016/j.snb.2019.127250

Kamiloglu, S. (2019). Authenticity and traceability in beverages. *Food Chemistry*, 277, 12-24. https://doi.org/10.1016/j.foodchem.2018.10.091

Kang, X., Zhao, Y. and Peng, J. (2022b). Authentication of the Geographical Origin of Shandong Scallop *Chlamys farreri* Using Mineral Elements Combined with Multivariate Data Analysis and Machine Learning Algorithm. *Food Analytical Methods,* 15, 2984–2993. https://doi.org/10.1007/s12161-022-02346-8

Kang, X., Zhao, Y. and Tan, Z. (2023). An explainable machine learning for geographical origin traceability of mussels *Mytilus edulis* based on stable isotope ratio and compositions of C, N, O and H. *Journal of Food Composition and Analysis*, 123, 105508. https://doi.org/10.1016/j.jfca.2023.105508

Kang, Z., Zhao, Y., Chen, L., Guo, Y., Mu, Q. and Wang, S. (2022a). Advances in machine learning and hyperspectral imaging in the food supply chain. *Food Engineering Reviews*, 14(4), 596-616. https://doi.org/10.1007/s12393-022-09322-2

Kar, S., Tudu, B. and Bandyopadhyay, R. (2024). Statistical machine learning techniques applied to NIR spectral data for rapid detection of sudan dye-I in turmeric powders with optimized pre-processing and wavelength selection. *Journal of Food Science and Technology,* 61, 1955-1964. https://doi.org/10.1007/s13197-024-05971-9

Karabagias, I.K. (2024). Food authentication and adulteration control based on metrics data of foods and chemometrics. *European Food Research and Technology*, 250(5), 1269-1283. https://doi.org/10.1007/s00217-024-04477-0

Karabagias, I.K. and Badeka, A.V. (2021). Physicochemical parameters and volatile compounds of herbal teas as indicators of products' brand name using chemometrics. *European Food Research and Technology*, 247(4), 961-974. https://doi.org/10.1007/s00217-020-03677-8

Karabagias, I.K., Karabagias, V.K. and Badeka, A.V. (2021). Volatilome of white wines as an indicator of authenticity and adulteration control using statistical analysis. *Australian Journal of Grape and Wine Research*, 27(3), 269-279. https://doi.org/10.1111/ajgw.12486

Karabagias, V.K., Karabagias, I.K., Louppis, A., Badeka, A., Kontominas, M.G. and Papastephanou, C. (2019). Valorization of prickly pear juice geographical origin based on mineral and volatile compound contents using LDA. *Foods*, 8(4), 123. https://doi.org/10.3390/foods8040123

Karacaglar, N.N.Y., Bulat, T., Boyaci, I.H. and Topcu, A. (2019). Raman spectroscopy coupled with chemometric methods for the discrimination of foreign fats and oils in cream and yogurt. *Journal of Food and Drug Analysis*, 27 (1), 101-110. https://doi.org/10.1016/j.jfda.2018.06.008

Kasampalis, D.S., Tsouvaltzis, P., Ntouros, K., Gertsis, A., Moshou, D. and Siomos, A.S. (2021). Rapid nondestructive postharvest potato freshness and cultivar discrimination assessment. *Applied Sciences*, 11(6), 2630. https://doi.org/10.3390/app11062630

Kastrati, G., Sopaj, F., Tašev, K., Stafilov, T., Šajn, R. and Paçarizi, M. (2023). Analysis of chemical elements in honey samples in the territory of Kosovo. *Journal of Food Composition and Analysis,* 124, 105505. https://doi.org/10.1016/j.jfca.2023.105505

Kaufmann, K.C., Sampaio, K.A., Garcia-Martin, J.F. and Barbin, D.F. (2022). Identification of coriander oil adulteration using a portable NIR spectrometer. *Food Control*, 132, 108536. https://doi.org/10.1016/j.foodcont.2021.108536

Keshavarzi, Z., Barzegari Banadkoki, S., Faizi, M., Zolghadri, Y. and Shirazi, F.H. (2020). Comparison of transmission FTIR and ATR spectra for discrimination between beef and chicken meat and quantification of chicken in beef meat mixture using ATR-FTIR combined with chemometrics. *Journal of Food Science and Technology*, 57, 1430-1438. https://doi.org/10.1007/s13197-019-04178-7

Khatodiya, N. and Malik, M. (2022). Effects of edible coating on fresh-cut fruits. *Journal of Pharmacognosy and Phytochemistry*, 11(1), 192-199. https://doi.org/10.22271/phyto.2022.v11.i1c.14342

Lan, H., Shuyu, G., Ye, W., Shang, W., Qiubo, C., Lu, L. and Tian, B. (2019). Attention based residual network for medicinal fungi near infrared spectroscopy analysis. *Mathematical Biosciences and Engineering*, 16(4), 3003-

REVIEW

3017. https://doi.org/10.3934/mbe.2019149

Leal, W., Llanos, E.J., Restrepo, G., Suárez, C.F. and Patarroyo, M.E. (2016). How frequently do clusters occur in hierarchical clustering analysis? A graph theoretical approach to studying ties in proximity. *Journal of Cheminformatics*, 8, 1-16. https://doi.org/10.1186/s13321-016-0114-x

Lee, J., Nazki, H., Baek, J., Hong, Y. and Lee, M. (2020). Artificial intelligence approach for tomato detection and mass estimation in precision agriculture. *Sustainability*, 12 (21), 9138. https://doi.org/10.3390/su12219138

Leng, T., Li, F., Xiong, L., Xiong, Q., Zhu, M. and Chen, Y. (2020). Quantitative detection of binary and ternary adulteration of minced beef meat with pork and duck meat by NIR combined with chemometrics. *Food Control*, 113, 107203. https://doi.org/10.1016/j.foodcont.2020.107203

Li, D., Zhang, F., Yu, J., Chen, X., Liu, B. and Meng, X. (2021a). A rapid and non-destructive detection of *Escherichia coli* on the surface of fresh-cut potato slices and application using hyperspectral imaging. *Postharvest Biology and Technology*, 171, 111352. https://doi.org/10.1016/j.postharvbio.2020.111352

Li, S., Yu, X., Zhen, Z., Huang, M., Lu, J., Pang, Y., Wang, X. and Gao, Y. (2021b). Geographical origin traceability and identification of refined sugar using UPLC-QTof-MS analysis. *Food Chemistry*, 348, 128701. https://doi.org/10.1016/j.foodchem.2020.128701

Li, Y., Chen, B., Ye, S., Wu, Q., Zhu, L. and Ding, Y. (2024b). Discrimination of untreated and sodium sulphite treated bean sprouts by Fourier transform infrared spectroscopy and chemometrics. *Food Additives and Contaminants*: *Part A*, 41(6), 587-600. https://doi.org/10.1080/19440049.2024.2341104

Li, Y., Logan, N., Quinn, B., Hong, Y., Birse, N., Zhu, H., Haughey, S., Elliott, C.T. and Wu, D. (2024a). Fingerprinting black tea: When spectroscopy meets machine learning a novel workflow for geographical origin identification. *Food Chemistry*, 438, 138029. https://doi.org/10.1016/j.foodchem.2023.138029

Liang, N., Sun, S., Zhang, C., He, Y. and Qiu, Z. (2022). Advances in infrared spectroscopy combined with artificial neural network for the authentication and traceability of food. *Critical Reviews in Food Science and Nutrition*, 62(11), 2963-2984. https://doi.org/10.1080/10408398.2020.1862045

Ligar, B., Madenda, S., Mardjan, S. and Kusuma, T. (2024). Design of a traceability system for a coffee supply chain based on blockchain and machine learning. *Journal of Industrial Engineering and Management*, 17(1), 151-167. https://doi.org/10.3926/jiem.6256

Lim, K., Pan, K., Yu, Z. and Xiao, R.H. (2020). Pattern recognition based on machine learning identifies oil adulteration and edible oil mixtures. *Nature Communications*, 11(1), 5353. https://doi.org/10.1038/s41467-020-19137-6

Lin, Y., Wu, Y., Fan, R., Zhan, C. and Kang, Z. (2024). Identification and quantification of adulterated collagen powder by fluorescence hyperspectral technology. *Food Measure*, 18, 5423–5436. https://doi.org/10.1007/s11694-024-02577-9

Liu, H. and Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences*, 9(20), 4396. https://doi.org/10.3390/app9204396

Liu, J., Meng, C., Zhang, K., Gong, S., Wang, F., Guo, L., Zou, N., Wu, M., Peng, C. and Xiong, L. (2024b). Origin traceability of Yimucao (*Chinese motherwort*) in China using stable isotopes and extracts assisted by machine learning techniques, *Journal of Food Composition and Analysis*, 126, 105900. https://doi.org/10.1016/j.jfca.2023.105900

Liu, L., Zuo, Z.T., Wang, Y.Z. and Xu, F.R. (2020a). A fast multi-source information fusion strategy based on FTIR spectroscopy for geographical authentication of wild *Gentiana rigescens*. *Microchemical Journal*, 159, 105360. https://doi.org/10.1016/j.microc.2020. 105360

Liu, W., Liang, X., He, S., Shi, Z., Cen, B., Chen, W., Bi, H. and Wang, C. (2024c). Rapid and simultaneous quantitative and discriminative analyses of liquor quality parameters with machine learning-assisted batch Raman spectroscopy: Synergistic instrumental upgrade and chemometric optimization. *Food Control*, 158, 110242. https://doi.org/10.1016/j.foodcont.2023.110242

Liu, X., Feng, B., Liu, H., Wang, Y., Luo, B., Yang, Y., Zhang, Q., Wang, Z., Xu, Z., Li, B. and Wang, F. (2024a). Machine learning-assisted FTIR spectra to predict freeze-drying curve of food. *LWT-Food Science and Technology*, 197, 115894. https://doi.org/10.1016/j.lwt.2024.115894

Liu, Y., Zhou, S., Han, W., Liu, W., Qiu, Z. and Li, C. (2019). Convolutional neural network for hyperspectral data analysis and effective wavelengths selection. *Analytica Chimica Acta*, 1086, 46-54. https://doi.org/10.1016/j.aca.2019.08.026

Liu, Z., Yuan, Y., Xie, T., Zhang, Y., Shao, S., Nie, J., Xia, W., Rogers, K.M. and Zhang, W. (2020b). Long-term agricultural effects on the authentication accuracy of organic, green, and conventional rice using isotopic and elemental chemometric analyses. *Journal of Agricultural and Food Chemistry*, 68(5), 1213-1225. https://doi.org/10.1021/acs.jafc.9b06847

Longobardi, F., Innamorato, V., Di Gioia, A., Ventrella, A., Lippolis, V., Logrieco, A.F., Catucci, L. and Agostiano, A. (2017). Geographical origin discrimination of lentils (*Lens culinaris* Medik.) using 1H NMR fingerprinting and multivariate statistical analyses. *Food Chemistry*, 237, 743 -748. https://doi.org/10.1016/j.foodchem.2017.05.159

Lozano-Castellón, J., Laveriano-Santos, E.P., Abuhahib, M.M., Pozzoli, C., Pérez, M., Vallverdú-Queralt, A. and Lamuela-Raventós, R. (2024). Proven traceability strategies using chemometrics for organic food authenticity. *Trends in Food Science and Technology*, 147, 104430. https://doi.org/10.1016/j.tifs.2024.104430

Luna, A.S., Da Silva, A.P., Da Silva, C.S., Lima, I.C.A. and De Gois, J.S. (2019). Chemometric methods for classification of clonal varieties of green coffee using Raman spectroscopy and direct sample analysis. *Journal of Food Composition and Analysis*, 76, 44-50. https://doi.org/10.1016/j.jfca.2018.12.001

Lyu, C., Yang, J., Wang, T., Kang, C., Wang, S., Wang, H., Wan, X., Zhou, L., Zhang, W., Huang, L. and Guo, L. (2020). A field trials-based authentication study of conventionally and organically grown Chinese yams using light stable isotopes and multi-elemental analysis combined with machine learning algorithms. *Food Chemistry,* 1-10. https://doi.org/10.1016/j.foodchem.2020.128506

Malavi, D., Nikkhah, A., Alighaleh, P., Einafshar, S., Raes, K. and Haute, S.V. (2024). Detection of saffron adulteration with *Crocus sativus* style using NIR-hyperspectral imaging and chemometrics. *Food Control*, 157, 110189. https://doi.org/10.1016/j.foodcont.2023.110189

Malik, M. and Sharma, A. (2021). Physicochemical and functional properties of yoghurt powder. *International Journal of Chemical Studies*, 9(4), 298-303. https://doi.org/10.22271/chemi.2021.v9.i4d.12072

Malik, M., Dewan, A., Sethi, P. and Khatkar, B.S. (2025). Food and Dairy Product Shelf-Life. Engineering Solutions for Sustainable Food and Dairy Production: Innovations and Techniques in Food Processing and Dairy Engineering, 403-462. https://doi.org/10.1007/978-3-031-75834-8_15

Mara, A., Migliorini, M., Ciulu, M., Chignola, R., Egido, C., Núñez, O., Sentellas, S., Saurina, J., Caredda, M., Deroma, M.A. and Deidda, S. (2024). Elemental fingerprinting combined with machine learning techniques as a powerful tool for geographical discrimination of honeys from nearby regions. *Foods*, 13(2), 243. https://doi.org/10.3390/foods13020243

Marchetti, P., Mottola, A., Tantillo, G., Castrica, M. and Di Pinto, A. (2021). Detection of undeclared presence of bovine milk in buffalo yogurt. *Journal of Dairy Science*, 104(4), 4056-4061. https://doi.org/10.3168/jds.2020-19194

Martinez-Castillo, C., Astray, G., Mejuto, J.C. and Simal-Gandara, J. (2020). Random Forest, Artificial Neural Network, and Support Vector Machine Models for Honey Classification. *Food*, 1(1), 69-76. https://doi.org/10.2991/efood.k.191004.001

Masoomi, S., Sharifi, H. and Hemmateenejad, B. (2024). A paper-based optical tongue for characterization of Iranian honey: Identification of geographical/botanical origins and adulteration detection. *Food Control*, 155, 110052. https://doi.org/10.1016/j.foodcont.2023.110052

Mazola, Y.T., Fernandes, E.A.D.N., Sarriés, G.A., Bacchi, M.A. and Gonzaga, C.L. (2023). Authentication of beef cuts by multielement and machine learning approaches, *Journal of Trace Elements in Medicine and Biology*, 78, 127164. https://doi.org/10.1016/j.jtemb.2023.127164.

Meena, D., Chakraborty, S. and Mitra, J. (2024). Geographical Origin Identification of Red Chili Powder Using NIR Spectroscopy Combined with SIMCA and Machine Learning Algorithms. *Food Analytical Methods,* 17, 1005-1023. https://doi.org/10.1007/s12161-024-02625-6

Meenakshi, M. (2020). Machine learning algorithms and their real-life applications: A survey. *Proceedings of the International Conference on Innovative Computing and Communications (ICICC)*. https://doi.org/10.2139/ssrn.3595299

Menegatti, D., Giuseppi, A., Delli Priscoli, F., Pietrabissa, A., Di Giorgio, A., Baldisseri, F., Mattioni, M., Monaco, S., Lanari, L., Panfili, M. and Suraci, V. (2023, August). CADUCEO: a platform to support federated healthcare facilities through artificial intelligence. *In Healthcare,* 11 (15), 2199. https://doi.org/10.3390/healthcare11152199

Menevseoglu, A., Entrenas, J.A., Gunes, N., Dogan, M.A. and Pérez-Marín, D. (2024). Machine learning-assisted near-infrared spectroscopy for rapid discrimination of apricot kernels in ground almond. *Food Control*, 159, 110272. https://doi.org/10.1016/j.foodcont.2023.110272

Meng, X., Yin, C., Yuan, L., Zhang, Y., Ju, Y., Xin, K., Chen, W., Lv, K. and Hu, L. (2023). Rapid detection of adulteration of olive oil with soybean oil combined with chemometrics by Fourier transform infrared, visible-near-infrared and excitation-emission matrix fluorescence spectroscopy: A comparative study. *Food Chemistry*, 405, 134828. https://doi.org/10.1016/j.foodchem.2022.134828

Millatina, N.R.N., Calle, J.L.P., Barea-Sepúlveda, M., Setyaningsih, W. and Palma, M. (2024). Detection and quantification of cocoa powder adulteration using Vis-NIR spectroscopy with chemometrics approach. *Food Chemistry*, 449, 139212. https://doi.org/10.1016/j.foodchem.2024.139212

Minaei, S., Shafiee, S., Polder, G., Moghadam-Charkari, N., Van Ruth, S., Barzegar, M., Zahiri, J., Alewijn, M. and Kus, P.M. (2017). VIS/NIR imaging application for honey floral origin determination. *Infrared Physics and Technology,* 86, 218-225. http://dx.doi.org/http://dx.doi.org/10.1016/j.infrared.2017.09.001

Montesinos López, O.A., Montesinos López, A. and Crossa, J. (2022). Overfitting, model tuning, and evaluation of prediction performance. In Multivariate statistical machine learning methods for genomic prediction (pp. 109-139). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_4

Morariu, I.D., Avasilcai, L., Vieriu, M., Lupu, V., Ioniuc, I., Morariu, B.A., Lupu, A., Morariu, P., Pop, O.L., Burduloi, V., Starcea, I. and Trandafir, L. (2024). A Comprehensive Narrative Review on the Hazards of Bee Honey Adulteration and Contamination. *Journal of Food Quality,* Volume 2024. Article ID3512676. https://doi.org/10.1155/2024/3512676

Mota, L.F., Giannuzzi, D., Bisutti, V., Pegolo, S., Trevisi, E., Schiavon, S., Gallo, L., Fineboym, D., Katz, G. and Cecchinato, A. (2022). Real-time milk analysis integrated with stacking ensemble learning as a tool for the daily prediction of cheese-making traits in Holstein cattle. *Journal of Dairy Science*, 105(5), 4237-4255. https://doi.org/10.3168/jds.2021-21426

REVIEW

Mota, L.F., Pegolo, S., Baba, T., Peñagaricano, F., Morota, G., Bittante, G. and Cecchinato, A. (2021). Evaluating the performance of machine learning methods and variable selection methods for predicting difficult-to-measure traits in Holstein dairy cattle using milk infrared spectral data. *Journal of Dairy Science*, 104(7), 8107-8121. https://doi.org/10.3168/jds.2020-19861

Muna, E.R., Loh, J.Z., Lim, Z.L. and Ahmad, F.O. (2022). Near-infrared spectroscopy with chemometrics for identification and quantification of adulteration in high-quality stingless bee honey. *Chemometrics and Intelligent Laboratory Systems*, 224, 104540. https://doi.org/10.1016/j.chemolab.2022.104540

Murru, C., Chimeno-Trinchet, C., Díaz-García, M.E., Badía-Laíño, R. and Fernández-González, A. (2019). Artificial Neural Network and Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy to identify the chemical variables related to ripeness and variety classification of grapes for Protected. Designation of Origin wine production. *Computers and Electronics in Agriculture*, 164, 104922. https://doi.org/10.1016/j.compag.2019.104922

Mustapha, A., Ishak, I., Zaki, N.N.M., Ismail-Fitry, M.R., Arshad, S. and Sazili, A.Q. (2024). Application of machine learning approach on halal meat authentication principle, challenges, and prospects: A review. *Heliyon*, 10(12), e32198. https://doi.org/10.1016/j.heliyon.2024.e32189

Nanou, E., Pliatsika, N. and Couris, S. (2023). Rapid Authentication and Detection of Olive Oil Adulteration Using Laser-Induced Breakdown Spectroscopy. *Molecules*, 28, 7960. https://doi.org/10.3390/molecules28247960

Nantongo, J.S., Serunkuma, E., Burgos, G., Nakitto, M., Davrieux, F. and Ssali, R. (2024). Machine learning methods in near infrared spectroscopy for predicting sensory traits in sweet potatoes. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 318, 124406. https://doi.org/10.1016/j.saa.2024.124406

Nazir, A., Hussain, A. and Assad, A. (2025) CNN in Food Industry: Current Practices and Future Trends. In Artificial Intelligence in the Food Industry (pp. 329-354). CRC Press. https://doi.org/10.1201/9781032633602-17

Neelakandan, S. and Paulraj, D. (2020). A gradient boosted decision tree-based sentiment classification of twitter data. *International Journal of Wavelets, Multiresolution and Information Processing*, 18(04), 2050027. https://doi.org/10.1142/S0219691320500277

Nelson, D.L. and Cox, M.M. (2004). Principles of bioenergetics. Lehninger, Principles of Biochemistry (4th ed., pp. 489-520). New York, USA: WH Freeman and company.

Nie, P., Zhang, J., Feng, X., Yu, C. and He, Y. (2019). Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning. *Sensors and Actuators B: Chemical*, 296, 126630. https://doi.org/10.1016/j.snb.2019.126630

Nunes, A., Azevedo, G.Z., Santos, B.R.D., Lima, G.P.P., Moura, S. and Maraschin, M. (2024). Application of UV–vis spectrophotometry and chemometrics to investigate adulteration by glucose syrup in Brazilian polyfloral honey. *Food and Humanity*, 2, 100194. https://doi.org/10.1016/j.foohum.2023.12.002

Nunes, K.M., Andrade, M.V.O., Almeida, M.R. and Sena, M.M. (2020). A soft discriminant model based on mid-infrared spectra of bovine meat purges to detect economic motivated adulteration by the addition of non-meat ingredients. *Food Analytical Methods*, 13, 1699-1709. https://doi.org/10.1007/s12161-020-01795-3

Olaniyi, E.O., Lu, Y., Zhang, X., Sukumaran, A.T., Thames, H.T. and Pokhrel, D. (2024). Non-Destructive Assessment of Microbial Spoilage of Broiler Breast Meat Using Structured Illumination Reflectance Imaging with Machine Learning. *Food Analytical. Methods* 17, 652–663. https://doi.org/10.1007/s12161-024-02605-w

Oliveira, A.N.D., Bolognini, S.R.F., Navarro, L.C., Delafiori, J., Sales, G.M., Oliveira, D.N.D. and Catharino, R.R. (2023). Tomato classification using mass spectrometry-machine learning technique: A food safety-enhancing platform, *Food Chemistry*, 398, 133870. https://doi.org/10.1016/j.foodchem.2022.133870

Orche, A.E., Mamad, A., Elhamdaoul, O., Cheikh, A., Karbane, M.E. and Bouatia, M. (2021). Comparison of Machine Learning Classification Methods for Determining the Geographical Origin of Raw Milk Using Vibrational Spectroscopy. *Journal of Spectroscopy*, Volume 2021, Article ID 5845422. https://doi.org/10.1155/2021/5845422

O'Sullivan, R., Schmidt, O. and Monahan, F.J. (2022). Stable isotope ratio analysis for the authentication of milk and dairy ingredients: A review. *International Dairy Journal*, 126, 105268. https://doi.org/10.1016/j.idairyj.2021.105268

Pages-Rebull, J., Pérez-Ràfols, C., Serrano, N., del Valle, M, and Díaz-Cruz, J.M. (2023). Classification and authentication of spices and aromatic herbs by means of HPLC-UV and chemometrics. *Food Bioscience*, 52, 102401. https://doi.org/10.1016/j.fbio.2023.102401

Parastar, H., van Kollenburg, G., Weesepoel, Y., van den Doel, A., Buydens, L. and Jansen, J. (2020). Integration of handheld NIR and machine learning to “Measure and Monitor” chicken meat authenticity. *Food Control*, 112, 107149. https://doi.org/10.1016/j.foodcont.2020.107149

Peng, D., Zhou, Q., Su, M., Zheng, S., Xie, S. and Li, J. (2024). Quantitative determination of the carbonyl value in frying oils based on LF-NMR combined with chemometrics. *LWT-Food Science and Technology*, 198, 116067. https://doi.org/10.1016/j.lwt.2024.116067

Pereira, E.D.S., Cruz-Tirado, J.P., Crippa, B.L., Morasi, R.M., Almeida, J.M.D., Barbin, D.F., Junior, S.B. and Silva, N.C.C. (2024). Portable near infrared (NIR) spectrometer coupled with machine learning to classify milk with subclinical mastitis. *Food Control*, 163, 110527. https://://

doi.org/10.1016/j.foodcont.2024.110527

Phillips, T. and Abdulla, W. (2023). A new honey adulteration detection approach using hyperspectral imaging and machine learning. *European Food Research and Technology*, 249(2), 259-272. https://doi.org/10.1007/s00217-022-04113-9

Phuenpong, T., Kongboonkird, M., Duangmal, K., Lerdvorasap, W., Suksawwawimon, M., Mekboonsonglarp, W., Nuamchit, J., Chantaprasarn, N. and Settachaimongkon, S. (2021). Molecular Discrimination between Organic and Conventional Liquid Milk Products in Thailand Using ¹H-NMR Metabolomics Approach. *Tropical Animal Science Journal*, 44(4), 478-488. https://doi.org/10.5398/tasj.2021.44.4.478

Pragasam, T.T.N., Thomas, J.V.J., Vensuslaus, M.A. and Radhakrishnan, S. (2023). Ceat: Categorising ethereum addresses' transaction behaviour with ensemble machine learning algorithms. *Computation*, 11(8), 156. https://doi.org/10.3390/computation11080156

Przybył, K., Gawrysiak-Witulska, M., Bielska, P., Rusinek, R., Gancarz, M., Dobrzański, B. Jr. and Siger, A. (2023). Application of Machine Learning to Assess the Quality of Food Products—Case Study: Coffee Bean. *Applied Sciences,* 13, 10786. https://doi.org/10.3390/app131910786

Pu, K., Qiu, J., Tong, Y., Liu, B., Cheng, Z., Chen, S., Ni, W.X., Lin, Y. and Ng, K.M. (2022). Integration of non-targeted proteomics mass spectrometry with machine learning for screening cooked beef adulterated samples. *Journal of Agricultural and Food Chemistry*, 71 (4), 2173-2182. https://doi.org/10.1021/acs.jafc.2c06266

Putri, L.A., Rahman, I., Puspita, M., Hidayat, S.N., Dharmawan, A.B., Rianjanu, A., Wibirama, S., Roto, R., Triyana, K. and Wasisto, H.S. (2023). Rapid analysis of meat floss origin using a supervised machine learning-based electronic nose towards food authentication. *npj Science of Food,* 7(1), 31. https://doi.org/10.1038/s41538-023-00205-2

Qiu, G., Lü, E., Wang, N., Lu, H., Wang, F. and Zeng, F. (2019). Cultivar classification of single sweet corn seed using fourier transform near-infrared spectroscopy combined with discriminant analysis. *Applied Sciences*, 9 (8), 1530. https://doi.org/10.3390/app9081530

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

Quintelas, C., Rodrigues, C., Sousa, C., Ferreira, E.C. and Amaral, A.L. (2024). Cookie composition analysis by Fourier transform near infrared spectroscopy coupled to chemometric analysis. *Food Chemistry*, 435, 137607. https://doi.org/10.1016/j.foodchem.2023.137607

Rachineni, K., Kakita, V.M.R., Awasthi, N.P., Shirke, V.S., Hosur, R.V. and Shukla, S.C. (2022). Identifying type of sugar adulterants in honey: Combined application of NMR spectroscopy and supervised machine learning classification. *Current Research in Food Science*, 5, 272-277. https://doi.org/10.1016/j.crfs.2022.01.008

Rady, A.M., Guyer, D.E. and Watson, N.J. (2021). Near-infrared spectroscopy and hyperspectral imaging for sugar content evaluation in potatoes over multiple growing seasons. *Food Analytical Methods*, 14(3), 581-595. https://doi.org/10.1007/s12161-020-01886-1

Raje, V.V. (2024). Ensuring Market Authenticity of Labaneh: Machine Learning-Driven Analysis of Fatty Acid Profiles for Detecting Non-Milk Fat Adulterants in Jordanian Products. *Revista Electronica de Veterinaria*, 25(1), 502-512.

Ranaweera, R.K., Gilmore, A.M., Capone, D.L., Bastian, S.E. and Jeffery, D.W. (2021). Spectrofluorometric analysis combined with machine learning for geographical and varietal authentication, and prediction of phenolic compound concentrations in red wine. *Food Chemistry*, 361, 130149. https://doi.org/10.1016/j.foodchem.2021.130149

Raypah, M.E., Zhi, L.J., Loon, L.Z. and Omar, A.F. (2022). Near-infrared spectroscopy with chemometrics for identification and quantification of adulteration in high-quality stingless bee honey. *Chemometrics and Intelligent Laboratory Systems*, 224, 104540. https://doi.org/10.1016/j.chemolab.2022.104540

Razavi, R. and Kenari, R.E. (2023). Ultraviolet–visible spectroscopy combined with machine learning as a rapid detection method to the predict adulteration of honey. *Heliyon*, 9, e20973. https://doi.org/10.1016/j.heliyon.2023.e20973

Rebellato, A.P., dos Santos Caramês, E.T., de Moraes, P.P. and Pallone, J.A.L. (2020). Minerals assessment and sodium control in hamburger by fast and green method and chemometric tools. *LWT-Food Science and Technology*, 128, 109438. https://doi.org/10.1016/j.lwt.2020.109438

Ren, T., Lin, Y., Su, Y., Ye, S. and Zheng, C. (2024). Machine Learning-Assisted Portable Microplasma Optical Emission Spectrometer for Food Safety Monitoring. *Analytical Chemistry*, 96(13), 5170-5177. https://doi.org/10.1021/acs.analchem.3c05332

Revilla, C., Lastras, M.I., González-Martín, A.M. Vivar-Quintana, R., Morales-Corts, M.A., Gómez-Sánchez, R. and Pérez-Sánchez, R. (2019). Predicting the physicochemical properties and geographical ORIGIN of lentils using near infrared spectroscopy. *Journal of Food Composition and Analysis*, 77, 84-90. https://doi.org/10.1016/j.jfca.2019.01.012

Ricci, C., Gadaleta, A., Gerardino, A., Didonna, A., Ferrara, G. and Bertani, F.R. (2024). Portable optical spectroscopy and machine learning techniques for quantification of the biochemical content of raw food materials. *CABI Agriculture and Bioscience*, 5(1), 39. https://doi.org/10.1186/s43170-024-00244-z

Richter, B., Gurk, S., Wagner, D., Bockmayr, M. and Fischer, M. (2019). Food authentication: Multi-elemental analysis of white asparagus for provenance discrimination. *Food Chemistry*, 286, 475-482. https://doi.org/10.1016/

REVIEW

j.foodchem.2019.01.105

Robert, C., Fraser-Miller, S.J., Jessep, W.T., Bain, W.E., Hicks, T.M., Ward, J.F., Craigie, C.R., Loeffen, M. and Gordon, K.C. (2021). Rapid discrimination of intact beef, venison and lamb meat using Raman spectroscopy. *Food Chemistry*, 343, 128441. https://doi.org/10.1016/j.foodchem.2020.128441

Robson, K., Birse, N., Chevallier, O. and Elliott, C. (2022). Metabolomic profiling to detect different forms of beef fraud using rapid evaporative ionisation mass spectrometry (REIMS). *npj Science of Food*, 6(1), 9. https://doi.org/10.1038/s41538-022-00125-7

Roozbahani, Z.A., Labbafi, M. and Aghakhani, A. (2024). Identification of the authenticity and geographical origin of Iranian lime juice by targeted method and machine learning. *Food Measure,* 18, 40–55. https://doi.org/10.1007/s11694-023-02142-w

Ropelewska, E. (2022). Postharvest authentication of potato cultivars using machine learning to provide high-quality products. *Chemistry Proceedings*, 10(1), 30. https://doi.org/10.3390/IOCAG2022-12285

Roshani, M., Phan, G.T., Ali, P.J.M., Roshani, G.H., Hanus, R., Duong, Corniani, E., Nazemi, E. and Kalmoun, E.M. (2021). Evaluation of flow pattern recognition and void fraction measurement in two phase flow independent of oil pipeline's scale layer thickness. *Alexandria Engineering Journal*, 60(1), 1955-1966. https://doi.org/10.1016/j.aej.2020.11.043

Ru, C., Li, Z. and Tang, R. (2019). A hyperspectral imaging approach for classifying geographical origins of rhizome atractylodis macrocephalae using the fusion of spectrum-image in VNIR and SWIR ranges (VNIR-SWIR-FuSI). *Sensors*, 19(9), 2045. https://doi.org/10.3390/s19092045

Ruvalcaba, J.E., Durán-Guerrero, E., Barroso, C.G. and Castro, R. (2019). Development of a stir bar sorptive extraction method to study different beer styles volatile profiles. *Food Research International*, 126, 108680. https://doi.org/10.1016/j.foodres.2019.108680

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

Scatigno, C. and Festa, G. (2022). FTIR coupled with machine learning to unveil spectroscopic benchmarks in the Italian EVOO. *International Journal of Food Science and Technology*, 57(7), 4156-4162. https://doi.org/10.1111/ijfs.15735

Schmauder, F., Schütz, D., Creydt, M., Riedl, J. and Fischer, M. (2024). Food authentication goes green: Determination of the geographical origin of grain maize (*Zea mays* L.) using direct analysis in real time mass spectrometry (DART-MS). *Food Control*, 163, 110489. https://doi.org/10.1016/j.foodcont.2024.110489

Schreurs, M., Piampongsant, S., Roncoroni, M., Cool, L., Herrera-Malaver, B., Vanderaa, C., Theßeling, F.A., Kreft, Ł., Botzki, A., Malcorps, P. and Verstrepen, K.J. (2024). Predicting and improving complex beer flavor through machine learning. *Nature Communications*, 15(1), 2368. https://doi.org/10.1038/s41467-024-46346-0

Schroeder, V., Evans, E.D., Wu, Y.C.M., Voll, C.C.A., McDonald, B.R., Savagatrup, S. and Swager, T.M. (2019). Chemiresistive sensor array and machine learning classification of food. *ACS Sensors*, 4(8), 2101-2108. http://dx.doi.org/10.1021/acssensors.9b00825

Sen, P.C., Hajra, M. and Ghosh, M. (2020). Supervised classification algorithms in machine learning: A survey and review. In Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*,* 99-111. Springer Singapore. https://doi.org/10.1007/978-981-13-7403-6_11

Shang, J., Liu, N., Cheng, J., Gao, W., Sun, X. and Guo, M. (2022). Analysis and comparison of lipids in Saanen goat milk from different geographic regions in China based on UHPLC-QTOF-MS lipidomics. *Food Research International*, 157, 111441. https://doi.org/10.1016/j.foodres.2022.111441

Sharma, A., Kumar, R., Kumar, N. and Saxena, V. (2024). Machine learning driven portable Vis-SWNIR spectrophotometer for non-destructive classification of raw tomatoes based on lycopene content. *Vibrational Spectroscopy*, 130, 103628. https://doi.org/10.1016/j.vibspec.2023.103628

Shen, C., Wang, R., Nawazish, H., Wang, B., Cai, K. and Xu, B. (2024a). Machine vision combined with deep learning–based approaches for food authentication: An integrative review and new insights. *Comprehensive Reviews in Food Science and Food Safety*, 23(6), e70054. https://doi.org/10.1111/1541-4337.70054

Shen, J., Li, T., Chen, Y., Zhou, H., Dong, S., Wei, Y., Li, F., Ning, J. and Li, L. (2024b). Tracing the geographic origin of CTC black tea based on colorimetric sensor array response to taste substances combined with chemometrics. *Food Control*, 161, 110422. https://doi.org/10.1016/j.foodcont.2024.110422

Shi, Y., Gong, F., Wang, M., Liu, J., Wu, Y. and Men, H. (2019). A deep feature mining method of electronic nose sensor data for identifying beer olfactory information. *Journal of Food Engineering*, 263, 437-445. https://doi.org/10.1016/j.jfoodeng.2019.07.023

Shiv, K., Singh, A., Kumar, S., Prasad, L.B., Gupta, S. and Bharty, M.K. (2024). Evaluation of different regression models for detection of adulteration of mustard and canola oil with argemone oil using fluorescence spectroscopy coupled with chemometrics. *Food Additives and Contaminants: Part A*, 41(2), 105-119. https://doi.org/10.1080/19440049.2023.2297869

Shuai, M., Li, H., She, S., Wu, M. and Chen, L. (2023). C/H/O stable isotope of rape honey and its components combined with machine learning algorithms to identify geographical origin. *Rapid Communications in Mass Spectrometry*, 37 (15), e9539. https://doi.org/10.1002/rcm.9539. PMID: 37184427

Silva, A.D.S., Maciel, M.C., Oliveira, A.A.F. and Oliveira, T.F. (2024). Evaluation of the content of macro and trace elements and the geographic origin of honey in North Brazil through statistical and machine learning techniques. *Journal of Food Composition and Analysis*, 128, 106050. https://doi.org/10.1016/j.jfca.2024.106050

Silva, L.A.D., Flumignan, D.L., Tininis, A.G., Pezza, H.R. and Pezza, L. (2019). Discrimination of Brazilian lager beer by 1H NMR spectroscopy combined with chemometrics. *Food Chemistry*, 272, 488-493. https://doi.org/10.1016/j.foodchem.2018.08.077

Sim, J., Dixit, Y., Mcgoverin, C., Oey, I., Frew, R., Reis, M.M. and Kebede, B. (2024). Machine learning-driven hyperspectral imaging for non-destructive origin verification of green coffee beans across continents, countries, and regions. *Food Control*, 156, 110159. https://doi.org/10.1016/j.foodcont.2023.110159

Siripatrawan, U. and Makino, Y. (2024). Assessment of food safety risk using machine learning-assisted hyperspectral imaging: Classification of fungal contamination levels in rice grain. *Microbial Risk Analysis*, 27–28, 100295. https://doi.org/10.1016/j.mran.2024.100295

Sobhaninia, M., Mani-Varnosfaderani, A., Barzegar, M. and Sahari, M.A. (2024). Combining ion mobility spectrometry and chemometrics for detecting synthetic colorants in black tea: A reliable and fast method. *Food Chemistry: X*, 21, 101213. https://doi.org/10.1016/j.fochx.2024.101213

Sobrino-Gregorio, L., Vilanova, S., Prohens, J. and Escriche, I. (2019). Detection of honey adulteration by conventional and real-time PCR. *Food Control*, 95, 57-62. https://doi.org/10.1016/j. foodcont.2018.07.037

Song, G., Guo, X., Li, Q., Lv, J., Wang, D., Yuan, T., Liu, S., Li, L., Liao, J., Zhang, M., Shen, Q., Zheng, F. and Gong, J. (2024). Mislabeling identification of fresh retail beef cuts using machine learning – guided REIMS lipidomic fingerprints. *Food Control*, 161, 110401. https://doi.org/10.1016/j.foodcont.2024.110401

Song, W. (2019). Spectral pattern recognition and application. Ultser University. Data retrieved on 1 August, 2025.

Song, W., Jiang, N., Wang, H. and Guo, G. (2020b). Evaluation of machine learning methods for organic apple authentication based on diffraction grating and image processing. *Journal of Food Composition and Analysis*, 88, 103437. https://doi.org/10.1016/j.jfca.2020.103437

Song, W., Jiang, N., Wang, H. and Vincent, J. (2020a). Use of smartphone videos and pattern recognition for food authentication. *Sensors and Actuators B: Chemical*, 304, 127247. https://doi.org/10.1016/j.snb.2019.127247

Song, X., Canellas, E. and Nerin, C. (2021). Screening of volatile decay markers of minced pork by headspace-solid phase microextraction–gas chromatography–mass spectrometry and chemometrics. *Food Chemistry*, 342, 128341. https://doi.org/10.1016/j.foodchem.2020.128341

Soyeurt, H., Grelet, C., McParland, S., Calmels, M., Coffey, M., Tedde, A., A., Delhez, P., Dehareng, F. and Gengler, N. (2020). A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra. *Journal of Dairy Science*, 103 (12), 11585-11596. https://doi.org/10.3168/jds.2020-18870

Spina, A.A., Ceniti, C., De Fazio, R., Oppedisano, F., Palma, E., Gugliandolo, E., Crupi, R., Raza, S.H.A., Britti, D., Piras, C. and Morittu, V.M. (2024). Spectral Profiling (Fourier Transform Infrared Spectroscopy) and Machine Learning for the Recognition of Milk from Different Bovine Breeds. *Animals*, 14(9), 1271. https://doi.org/10.3390/ani14091271

Stefas, D., Gyftokostas, N., Kourelias, P., Nanou, E., Kokkinos, V., Bouras, C. and Couris, S. (2021). Discrimination of olive oils based on the olive cultivar origin by machine learning employing the fusion of emission and absorption spectroscopic data. *Food Control*, 130, 108318. https://doi.org/10.1016/j.foodcont.2021.108318

Sun, H., Hua, Z., Yin, C., Li, F. and Shi, Y. (2024a). Geographical traceability of soybean: An electronic nose coupled with an effective deep learning method, *Food Chemistry*, 440, 138207. https://doi.org/10.1016/j.foodchem.2023.138207

Sun, J., Xu, X., Feng, S., Zhang, H., Xu, L., Jiang, H., Sun, B., Meng, Y. and Chen, W. (2023). Rapid identification of salmonella serovars by using Raman spectroscopy and machine learning algorithm, *Talanta*, 253, 123807. https://doi.org/10.1016/j.talanta.2022.123807

Sun, T., Yang, Q., Zhang, Y., Guo, B., Guo, Y., Jia, Q. and Zhao, H. (2024b). Identification of the Geographic Origin of Peanut Kernels by Raman Spectroscopy Fingerprinting with Chemometrics. *Analytical Letters*, 57(4), 628–639. https://doi.org/10.1080/00032719.2023.22208434

Sun, X., Hu, Y., Liu, C., Zhang, S., Yan, S., Liu, X. and Zhao, K. (2024c). Characterizing Edible Oils by Oblique-Incidence Reflectivity Difference Combined with Machine Learning Algorithms. *Foods*, 13(9), 1420. https://doi.org/10.3390/foods13091420

Ta, N., Wei, H., Hu, Z., Cao, X. and Xiao, M. (2023). Wine component tracing method based on near infrared spectrum fusion machine learning. *Frontiers in Sustainable Food Systems*, *7*, 1197508. https://doi.org/0.3389/fsufs.2023.1197508

Tachie, C.Y., Obiri-Ananey, D., Alfaro-Cordoba, M., Tawiah, N.A. and Aryee, A.N. (2024). Classification of oils and margarines by FTIR spectroscopy in tandem with machine learning. *Food Chemistry*, 431, 137077. https://doi.org/10.1016/j.foodchem.2023.137077

Takahashi, K., Ichikawa, K., Park, J. and Pao, G.M. (2023). Scalable empirical dynamic modeling with parallel computing and approximate k-NN search. *IEEE Access*, 11, 68171-68183. https://doi.org/10.1109/ACCESS.2023.3289836

Taki, O., Rhazi, K.S. and Mejdoub, Y. (2023). Stirling engine optimization using artificial neural networks algorithm. *ITM Web of Conferences,* 52, 02010. EDP Sciences. https://doi.org/10.1051/itmconf/20235202010

Temerdashev, Z., Khalafyan, A., Abakumov, A., Bolshov, M.,

REVIEW

Akin'shina, V. and Kaunova, A. (2024). Authentication of selected white wines by geographical origin using ICP spectrometric and chemometric analysis. *Heliyon*, 10(8), e29607. https://doi.org/10.1016/j.heliyon.2024.e29607

Teng, Y., Chen, Y., Chen, X., Zuo, S., Li, X., Pan, Z., Shao, K., Du, J. and Li, Z. (2024). Revealing the adulteration of sesame oil products by portable Raman spectrometer and 1D CNN vector regression: A comparative study with chemometrics and colorimetry. *Food Chemistry*, 436, 137694. https://doi.org/10.1016/j.foodchem.2023.137694

Thomas, T., Vijayaraghavan, A.P. and Emmanuel, S. (2020). Nearest neighbor and fingerprint classification. *Machine Learning Approaches in Cyber Security Analytics*, 107-128. https://doi.org/10.1007/978-981-15-1706-8_6

Tolentino, I.C.D.S.A., de Jesus, J.C., Conceição, D.G., Onelli, R.R.V., Reis, L.C.C., Santos, L.S. and Ferrão, S.P.B. (2023). Use of MIR spectroscopy associated with chemometric techniques to verify the authenticity of prato cheese. *International Dairy Journal*, 142, 105643. https://doi.org/10.1016/j.idairyj.2023.105643

Tomar, D., Malik, M. and Khatkar, B.S. (2025). A multi-criteria decision-making (MCDM) framework for selecting wheat varieties to produce superior quality Bulgur. *Food Measure* (2025). https://doi.org/10.1007/s11694-025-03535-9

Totaro, M.P., Squeo, G., De Angelis, D., Pasqualone, A., Caponio, F. and Summo, C. (2023). Application of NIR spectroscopy coupled with DD-SIMCA class modelling for the authentication of pork meat. *Journal of Food Composition and Analysis*, 118, 105211. https://doi.org/10.1016/j.jfca.2023.105211

Tseng, Y.J., Chuang, P.J. and Appell, M. (2023). When machine learning and deep learning come to the big data in food chemistry. *ACS Omega*, 8(18), 15854-15864. https://doi.org/10.1021/acsomega.2c07722

Tsirogiannis, G., Thomatou, A.A., Psarra, E., Mazarakioti, E.C., Katerinopoulou, K., Zotos, A., Kontogeorgos, A., Patakas, A. and Ladavos, A. (2022). Probabilistic Machine Learning for the Authentication of the Protected Designation of Origin of Greek Bottarga from Messolongi: A Generic Methodology to Cope with Very Small Number of Samples. *Applied Sciences*, 12(13), 6335. https://doi.org/10.3390/app12136335

Tsouvaltzis, P., Babellahi, F., Amodio, M.L. and Colelli, G. (2020). Early detection of eggplant fruit stored at chilling temperature using different non-destructive optical techniques and supervised classification algorithms. *Postharvest Biology and Technology*, 159, 111001. https://doi.org/10.1016/j.postharvbio.2019.111001

Tüzemen, E.Ş., Yüksek, A.G., Demir, İ., Horoz, S. and Altuntaş, İ. (2023). Modeling of temperature-dependent photoluminescence of GaN epilayer by artificial neural network. *Journal of the Australian Ceramic Society*, 59 (5), 1145-1159. https://doi.org/10.1007/s41779-023-00911-w

Uddin, K.M.M., Biswas, N., Rikta, S.T. and Dey, S.K. (2023). Machine learning-based diagnosis of breast cancer utilizing feature optimization technique. *Computer Methods and Programs in Biomedicine Update*, 3, 100098. https://doi.org/10.1016/j.cmpbup.2023.100098

Valinger, D., Longin, L., Grbeš, F., Benković, M., Jurina, T., Kljusurić, J.G. and Tušek, A.J. (2021). Detection of honey adulteration–The potential of UV-VIS and NIR spectroscopy coupled with multivariate analysis. *LWT*, 145, 111316. https://doi.org/10.1016/j.lwt.2021.111316

Varrà, M.O., Fasolato, L., Serva, L., Ghidini, S., Novelli, E. and Zanardi, E. (2020). Use of near infrared spectroscopy coupled with chemometrics for fast detection of irradiated dry fermented sausages. *Food Control*, 110, 107009. https://doi.org/10.1016/j.foodcont.2019.107009

Vavoura, M.V., Karabagias, I.K., Kosma, I.S., Badeka, A.V. and Kontominas, M.G. (2022). Characterization and Differentiation of Fresh Orange Juice Variety Based on Conventional Physicochemical Parameters, Flavonoids, and Volatile Compounds Using Chemometrics. *Molecules*, 27(19), 6166. https://doi.org/10.3390/molecules27196166

Viejo, C.G. and Fuentes, S. (2020). Low-Cost Methods to Assess Beer Quality Using Artificial Intelligence Involving Robotics, an Electronic Nose, and Machine Learning. *Fermentation*, 6, 104. https://doi.org/10.3390/fermentation6040104

Viejo, C.G., Fuentes, S., Torrico, D.D., Godbole, A. and Dunshea, F.R. (2019). Chemical characterization of aromas in beer and their effect on consumers liking. *Food Chemistry*, 293, 479-485. https://doi.org/10.1016/j.foodchem.2019.04.114

Visciano, P. and Schirone, M. (2021). Food frauds: Global incidents and misleading situations. *Trends in Food Science and Technology*, 114, 424-442. https://doi.org/10.1016/j.tifs.2021.06.010

Vishnuraj, M.R., Devatkal, S., Vaithiyanathan, S., Kumar, R.U. and Mendiratta, S.K. (2021). Development and validation of miRNA based method for rapid identification of offal meats in processed chicken meat products. *Food Control*, 121, 107593. https://doi.org/10.1016/j.foodcont.2020.107593

von Wuthenau, K., Segelke, T., Müller, M.S., Behlok, H. and Fischer, M. (2022). Food authentication of almonds (*Prunus dulcis* Mill.). Origin analysis with inductively coupled plasma mass spectrometry (ICP-MS) and chemometrics. *Food Control*, 134, 108689. https://doi.org/10.1016/j.foodcont.2021.108689

Wadood, S.A., Guo, B., Zhang, X. and Wei, Y. (2019). Geographical origin discrimination of wheat kernel and white flour using near-infrared reflectance spectroscopy fingerprinting coupled with chemometrics. *International Journal of Food Science and Technology*, 54(6), 2045-2054. https://doi.org/10.1111/ijfs.14105

Wakhid, S., Sarno, R., Sabilla, S.I. and Maghfira, D.B. (2020). Detection and Classification of Indonesian Civet and Non-Civet Coffee Based on Statistical Analysis Comparison

Using E-Nose. *International Journal of Intelligent Engineering and Systems*, 13(4), 56-65. https://doi.org/10.22266/ijies2020.0831.06

Wang, A.X., Chukova, S.S. and Nguyen, B.P. (2023). Ensemble k-nearest neighbors based on centroid displacement. *Information Sciences*, 629, 313-323. https://doi.org/10.1016/j.ins.2023.02.004

Wang, H. and Jeffery, D.W. (2024). Machine Learning Model Stability for Sub-Regional Classification of Barossa Valley Shiraz Wine Using A-TEEM Spectroscopy. *Foods*, 13(9), 1376. https://doi.org/10.3390/foods13091376

Wang, R., Song, J., Liu, J., Ren, Z., Zhu, C., Yu, Y., Li, Z. and Huang, Y. (2024a). The combination of near-infrared spectroscopy with chemometrics in achieving rapid and accurate determination of rice mildew. *Journal of Food Measurement and Characterization*, 18(7), 5437-5449. https://doi.org/10.1007/s11694-024-02578-8

Wang, X., Gu, Y., Lin, W. and Zhang, Q. (2024b). Rapid quantitative authentication and analysis of camellia oil adulterated with edible oils by electronic nose and FTIR spectroscopy. *Current Research in Food Science*, 8, 100732. https://doi.org/10.1016/j.crfs.2024.100732

Wang, X., Li, F., Wei, L., Huang, Y., Wen, X., Wang, D., Cheng, G., Zhao, R., Lin, Y., Yang, H. and Fan, M. (2024c). Rapid and precise differentiation and authentication of agricultural products via deep learning-assisted multiplex SERS fingerprinting. *Analytical Chemistry*, 96(11), 4682-4692. https://doi.org/10.1021/acs.analchem.4c00064

Wang, Y.T., Ren, H.B., Liang, W.Y., Jin, X., Yuan, Q., Liu, Z.R., Chen, D.M. and Zhang, Y.H. (2021). A novel approach to temperature-dependent thermal processing authentication for milk by infrared spectroscopy coupled with machine learning. *Journal of Food Engineering*, 311, 110740. https://doi.org/10.1016/j.jfoodeng.2021.110740

Warner, K. and Eskin, M. (1995). Methods to access quality and stability of oils and fat-containing foods. Aocs Publishing. https://doi.org/10.4324/9781003040958

Windarsih, A., Bakar, N.K.A., Dachriyanus, Yuliana, N.D., Riswanto, F.D.O. and Rohman, A. (2023). Analysis of Pork in Beef Sausages Using LC-Orbitrap HRMS Untargeted Metabolomics Combined with Chemometrics for Halal Authentication Study. *Molecules*, 28(16), 5964. https://doi.org/10.3390/molecules28165964

Wu, M., Sun, J., Lu, B., Ge, X., Zhou, X. and Zou, M. (2019a). Application of deep brief network in transmission spectroscopy detection of pesticide residues in lettuce leaves. *Journal of Food Process Engineering*, 42(3), e13005. https://doi.org/10.1111/jfpe.13005

Wu, N., Y. Zhang, R.S., Na, C.X., Mi, S.S., Zhu, Y., He, L. and C. Zhang, C. (2019b). Variety identification of oat seeds using hyperspectral imaging: Investigating the representation ability of deep convolutional neural network. *RSC Advances* 9(22), 12635–12644. https://doi.org/10.1039/C8RA10335F

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T. and He, L. (2022a). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, 135, 364-381. https://doi.org/10.1016/j.future.2022.05.014

Wu, X., Xu, B., Luo, H., Ma, R., Du, Z., Zhang, X., Liu, H. and Zhang, Y. (2023). Adulteration quantification of cheap honey in high-quality Manuka honey by two-dimensional correlation spectroscopy combined with deep learning. *Food Control*, 154, 110010. https://doi.org/10.1016/j.foodcont.2023.110010

Wu, X., Xu, B., Ma, R., Niu, Y., Gao, S., Liu, H. and Zhang, Y. (2022b). Identification and quantification of adulterated honey by Raman spectroscopy combined with convolutional neural network and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 274, 121133. https://doi.org/10.1016/j.saa.2022.121133

Wu, Y., Li, L., Liu, L. and Liu, Y. (2019c). Nondestructive measurement of internal quality attributes of apple fruit by using NIR spectroscopy. *Multimedia Tools and Applications*, 78(4), 4179-4195. https://doi.org/10.1007/s11042-017-5388-0

Xiao, R., Liu, L., Zhang, D., Ma, Y. and Ngadi, M.O. (2019c). Discrimination of organic and conventional rice by chemometric analysis of NIR spectra: a pilot study. *Journal of Food Measurement and Characterization*, 13, 238-249. https://doi.org/10.1007/s11694-018-9937-7

Xiong, F., Wang, S., Kang, C., Wang, Y., Bai, R., Li, H., Yang, J. and Guo, L. (2024). Geographical verification of *Pleuropterus multiflorus thunb*. by functional compounds, stable isotope ratios, and multielement combined with machine learning methods. *Food Chemistry*, 440, 138209. https://doi.org/10.1016/j.foodchem.2023.138209

Xu, F., Kong, F., Peng, H., Dong, S., Gao, W. and Zhang, G. (2021 a). Combing machine learning and elemental profiling for geographical authentication of Chinese Geographical Indication (GI) rice. *npj Science of Food*, 5 (1), 18. https://doi.org/10.1038/s41538-021-00100-8

Xu, M., Wang, X., Xu, Z., Wang, Y., Jia, P., Ding, W., Dong, S. and Liu, P. (2024). Aroma quality characterization for Pixian broad bean paste fermentation by electronic nose combined with machine learning methods. *Journal of Food Measurement and Characterization*, 18(5), 3363-3377. https://doi.org/10.1007/s11694-024-02410-3

Xu, S., Lu, H., Ference, C. and Zhang, Q. (2019). Visible/near infrared reflection spectrometer and electronic nose data fusion as an accuracy improvement method for portable total soluble solid content detection of orange. *Applied Sciences*, 9(18), 3761. https://doi.org/10.3390/app9183761

Xu, S., Zhao, C., Deng, X., Zhang, R., Qu, L., Wang, M., Ren, S., Wu, H., Yue, Z. and Niu, B. (2021). Multivariate analysis for organic milk authentication. *Journal of Chromatography B*, 1186, 123029. https://doi.org/10.1016/j.jchromb.2021.123029

Xue, D. and Zhao, H. (2024). Machine learning and electrochemistry techniques for detecting adulteration of

REVIEW

goat milk with cow milk. *Journal of Food Measurement and Characterization*, 18(7), 6012-6019. https://doi.org/10.1007/s11694-024-02625-4

Xue, Z., Zhang, Y., Yu, W., Zhang, J., Wang, J., Wan, F., Kim, Y., Liu, Y. and Kou, X. (2019). Recent advances in aflatoxin B1 detection based on nanotechnology and nanomaterials-A review. *Analytica Chimica Acta*, 1069, 1-27. https://doi.org/10.1016/j.aca.2019.04.032

Yakar, Y. and Karadağ, K. (2022). Identifying olive oil fraud and adulteration using machine learning algorithms. *Química Nova*, 45(10), 1245-1250. http://dx.doi.org/10.21577/0100-4042.20170948

Yang, S., Li, C., Li, C. Wang, Z., Huang, L. and Ding, Q. (2019) The Differences in the Dried Tangerine Peels Stored for Different Years Revealed by Terahertz Spectroscopy and Imaging. *Modern Food Science and Technology*, 35(12), 258-266. https://doi.org/10.13982/j.mfst.1673-9078.2019.12.033

Yang, Y. and Wei, L. (2021). Application of E-nose technology combined with artificial neural network to predict total bacterial count in milk. *Journal of Dairy Science*, 104(10), 10558-10565. https://doi.org/10.3168/jds.2020-19987

Yang, Y., Xiong, A., Sun, D. and Jin, P. (2024) Icp-Ms Combined with Machine-Learning for Geographic Authentication of Xinyang Maojian Tea. *Available at SSRN 4700415*.

Yang, Y., Zhao, X., Huang, M., Wang, X. and Zhu, Q. (2021). Multispectral image-based germination detection of potato by using supervised multiple threshold segmentation model and Canny edge detector. *Computers and Electronics in Agriculture*, 182, 106041. https://doi.org/10.1016/j.compag.2021.106041

Yazici, İ., Shayea, I. and Din, J. (2023). A survey of applications of artificial intelligence and machine learning in future mobile networks-enabled systems. *Engineering Science and Technology, an International Journal*, 44, 101455. https://doi.org/10.1016/j.jestch.2023.101455

Ye, W., Yan, T., Zhang, C., Duan, L., Chen, W., Song, H., Zhang, Y., Xu, W. and Gao, P. (2022). Detection of pesticide residue level in grape using hyperspectral imaging with machine learning. *Foods*, 11(11), 1609. https://doi.org/10.3390/foods11111609

Yong, B.X. and Brintrup, A. (2022). Bayesian autoencoders with uncertainty quantification: Towards trustworthy anomaly detection. *Expert Systems with Applications*, 209, 118196. https://doi.org/10.1016/j.eswa.2022.118196

Yu, X., Wang, J., Wen, S., Yang, J. and Zhang, F. (2019). A deep learning-based feature extraction method on hyperspectral images for nondestructive prediction of TVB-N content in Pacific white shrimp (*Litopenaeus vannamei*). *Biosystems Engineering*, 178, 244-255. https://doi.org/10.1016/j.biosystemseng.2018.11.018

Zacometti, C., Massaro, A., di Gioia, T., Lefevre, S., Frégière-Salomon, A., Lafeuille, J.L., Fiordaliso Candalino, I., Suman, M., Piro, R. and Tata, A. (2023). Thermal desorption direct analysis in real-time high-resolution mass spectrometry and machine learning allow the rapid authentication of ground black pepper and dried oregano: A proof-of-concept study. *Journal of Mass Spectrometry*, 58(10), e4953. https://doi.org/10.1002/jms.4953

Zacometti, C., Sammarco, G., Massaro, A., Lefevre, S., Frégière-Salomon, A., Lafeuille, J.L., Candalino, I.F., Piro, R., Tata, A. and Suman, M. (2024). Authenticity assessment of ground black pepper by combining headspace gas-chromatography ion mobility spectrometry and machine learning. *Food Research International*, 179, 114023. https://doi.org/10.1016/j.foodres.2024.114023

Zaresani, H., Afkari Sayyah, A.H., Zareiforoush, H., Khorramifar, A., Gancarz, M., Tabor, S. and Karami, H. (2024). Vis/NIR and FTIR spectroscopy supported by machine learning techniques to distinguish pure from impure Iranian rice varieties. *International Agrophysics*, 38(2), 203-211. https://doi.org/10.31545/intagr/185392

Zaukuu, J.L.Z., Adams, Z.S., Donkor-Boateng, N.A., Mensah, E.T., Bimpong, D. and Amponsah, L.A. (2024). Non-invasive prediction of maca powder adulteration using a pocket-sized spectrophotometer and machine learning techniques. *Scientific Reports*, 14(1), 10426. https://doi.org/10.1038/s41598-024-61220-1

Zawawi, N., Chong, P.J., Mohd Tom, N.N., Saiful Anuar, N.S., Mohammad, S.M., Ismail, N. and Jusoh, A.Z. (2021). Establishing relationship between vitamins, total phenolic and total flavonoid content and antioxidant activities in various honey types. *Molecules*, 26(15), 4399. https://doi.org/10.3390/molecules26154399

Zhang, C., Wu, W., Zhou, L., Cheng, H., Ye, X. and He, Y. (2020). Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (*Lycium ruthenicum* Murr.) using near-infrared hyperspectral imaging. *Food Chemistry*, 319, 126536. https://doi.org/10.1016/j.foodchem.2020.126536

Zhang, D., Lin, Z., Xuan, L., Lu, M., Shi, B., Shi, J., He, F., Battino, M., Zhao, L. and Zou, X. (2024a). Rapid determination of geographical authenticity and pungency intensity of the red Sichuan pepper (*Zanthoxylum bungeanum*) using differential pulse voltammetry and machine learning algorithms. *Food Chemistry*, 439, 137978. https://doi.org/10.1016/j.foodchem.2023.137978.

Zhang, L., Dai, H., Zhang, J., Zheng, Z., Song, B., Chen, J., Lin, G., Chen, L., Sun, W and Huang, Y. (2023). A study on origin traceability of white tea (White Peony) based on near-infrared spectroscopy and machine learning algorithms. *Foods*, 12(3), 499. https://doi.org/10.3390/foods12030499

Zhang, L., Li, Y., Gu, Y., Fu, Y., Zhang, X. and Hu, J. (2024b). Atlantic salmon adulteration authentication by machine learning using bioimpedance non-destructive flexible sensing. *Microchemical Journal*, 196, 109638. https://doi.org/10.1016/j.microc.2023.109638

Zhang, L., Traore, S., Ge, J., Li, Y., Wang, S., Zhu, G., Cui, Y.

and Fipps, G. (2019). Using boosted tree regression and artificial neural networks to forecast upland rice yield under climate change in Sahel. *Computers and Electronics in Agriculture*, 166, 105031. https://doi.org/10.1016/j.compag.2019.105031

Zhao, H., Zhan, Y., Xu, Z., Nduwamungu, J.J., Zhou, Y., Powers, R. and Xu, C. (2022). The application of machine -learning and Raman spectroscopy for the rapid detection of edible oils type and adulteration. *Food Chemistry*, 373, 131471.https://doi.org/10.1016/j.foodchem.2021.131471

Zhao, Y., Tu, T., Tang, X., Zhao, S., Qie, M., Chen, A. and Yang, S. (2020). Authentication of organic pork and identification of geographical origins of pork in four regions of China by combined analysis of stable isotopes and multi-elements. *Meat Science*, 165, 108129. https://doi.org/10.1016/j.meatsci.2020.108129

Zheng, X., Li, Y., Wei, W. and Peng, Y. (2019). Detection of adulteration with duck meat in minced lamb meat by using visible near-infrared hyperspectral imaging. *Meat Science*, 149, 55-62. https://doi.org/10.1016/j.meatsci.2018.11.005

Zhou, F., Wu, Y., Zeng, X., Zhang, Z., Wang, S., Huang, P. and Kang, Z. (2024a). Authenticity Verification of Alcoholic Beverages in IoT Environment through Fluorescence Hyperspectral Technology and Machine Learning. *Journal of Circuits, Systems and Computers*, 38 (17), 2450314. https://doi.org/10.1142/S0218126624503146

Zhou, L., Zhang, C., Liu, F., Qiu, Z. and He, Y. (2019). Application of deep learning in food: a review. *Comprehensive Reviews in Food Science and Food Safety*, 18(6), 1793-1811. https://doi.org/10.1111/1541-4337.12492

Zhou, X., Li, L., Zheng, J., Wu, J., Wen, L., Huang, M., Ao, F., Luo, W., Li, M., Wang, H. and Zong, X. (2024b). Quantitative analysis of key components in Qingke beer brewing process by multispectral analysis combined with chemometrics. *Food Chemistry*, 436, 137739. https://doi.org/10.1016/j.foodchem.2023.137739

Zhu, H. and Xu, J.L. (2020). Authentication and Provenance of Walnut Combining Fourier Transform Mid-Infrared Spectroscopy with Machine Learning Algorithms. *Molecules*, 25, 4987. https://doi.org/10.3390/molecules25214987

Zhu, S.S., Zhou, L., Gao, P., Bao, Y.D., He, Y. and Feng, L. (2019a). Near infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties. *Molecules* 24(18), 3268. https://doi.org/0.3390/molecules24183268

Zhu, S.S., Zhou, L., Zhang, C., Bao, Y.D., Wu, B.H., Chu, H.J., Yu, Y., He, Y. and L. Feng, L. (2019b). Identification of soybean varieties using hyperspectral imaging coupled with convolutional neural network. *Sensors* 19(19), 4065. https://doi.org/10.3390/s19194065

REVIEW

https://doi.org/10.26656/lifr.1.e25050