KARYA ILHAM PUBLISHING
CONNECTING MINDS, BRIDGING IDEAS, INSPIRING READERS

International Journal of Advanced Research in Food Science and Agriculture Technology

Journal homepage:
https://karyailham.com.my/index.php/fsat/index
ISSN: 3030-556X

Integration of Spectroscopy and Machine Learning for Food Contamination Detection: A Review

Muhammad Adil[1], Matti Ur Rehman[2], Ehtesham Ali[3,*], Hassan Ibrahim[4], Maria Rukan[5], Muhammad Qasim Ali[6]

[1] Aquila's Global Tech Haripur, Pakistan
[2] National University of Computer & Emerging Sciences – FAST, Pakistan
[3] Faculty of Electrical and Electronics Engineering Technology, University Malaysia Pahang, Al-Sultan Abdullah, Pekan, Pahang, Malaysia
[4] Faculty of Computing Universiti Malaysia Pahang Al-Sultan Abdullah 26600 Pekan, Pahang, Malaysia
[5] Department of Microbiology, The University of Haripur, Haripur 22620, Pakistan
[6] Faculty of Chemical and Process Engineering Technology, Universiti Malaysia Pahang, 26300 Gambang, Kuantan, Pahang, Malaysia

ARTICLE INFO

Article history:
Received 17 July 2025
Received in revised form 24 August 2025
Accepted 28 August 2025
Available online 30 September 2025

Keywords:
Nir spectroscopy; food safety; machine learning

ABSTRACT

Ensuring food safety remains a pressing global challenge due to the growing threat of contamination from microbial pathogens, chemical adulterants, and physical impurities. Traditional detection methods, while accurate, are often labor intensive, costly, and time-consuming limiting their applicability for real-time monitoring. This review aims to explore the integration of spectroscopy and machine learning (ML) as a powerful, non-destructive approach for the rapid detection and quantification of food contaminants. The paper critically examines recent advancements in spectroscopic techniques including Near-Infrared (NIR), Hyperspectral Imaging (HSI), Fourier-Transform Infrared (FTIR), Raman, and Ultraviolet-Visible (UV-Vis) spectroscopy when combined with both conventional machine learning algorithms and modern deep learning models. A comparative analysis of their performance across various food matrices is presented, highlighting their sensitivity, specificity, and operational feasibility. The review also identifies key limitations in current systems, such as data standardization, model interpretability, and hardware portability. Future research directions are discussed with an emphasis on explainable AI, the development of portable sensing platforms, and the establishment of open-access spectral databases to support widespread adoption in food quality monitoring.

1. Introduction

Food safety is a worldwide issue in today's modern society, considering that millions are foodborne illness victims every year. According to the World Health Organization (WHO), over 600

* *Corresponding author.*
E-mail address: ehteshamali23@gmail.com

https://doi.org/10.37934/fsat.6.1.123

million cases of foodborne disease are reported on an annual basis, with an estimated 420,000 deaths worldwide [1]. Young children, especially under five years of age, are most vulnerable, and therefore the urgent for an immediate intervention worldwide. Global food chains have raised the issue of industrial agriculture, global warming, and globalization. All of the above issues bring greater challenges in tracing and controlling food safety, especially for microbiological, chemical, and physical food contaminants [2]. In addition to the health effect of food safety, food contamination also raises economic consequences, such as product recall, disruption of trade, legal liability, and customer loss of confidence in food systems.

New food safety risks are induced by contaminants like pesticide residues, veterinary drug residues, heavy metals, and economically motivated adulterants, which continue to be a significant toxicological challenge [3]. Recent literature has reported the common occurrence of toxic analytes, like the highly toxic herbicide paraquat, which is still on the market in some regions of the world, pesticide residues on fruits, and mycotoxins in dried fruits. The label-free electrochemical apt sensor that was capable of measuring paraquat levels in vegetables with high specificity and on-site applicability [4]. The hyperspectral imaging and deep machine learning algorithms to detect pesticide residues in grapes at more than 93% accuracy, enabling risk assessment of contamination and efficacy of the new detection technology [5]. The pinpointed mycotoxins, ochratoxin OTA and patulin PAT, as issues impacting fresh and dried fruits in their research on detection technologies for reducing their transmission. These new findings point to the fact that new detection methods are not only rapid and inexpensive but are also suitable for centralized laboratories and decentralized inspection stations in global food chains [6].

Previously, food contamination was identified with the help of analytical methods like high-performance liquid chromatography (HPLC), gas chromatography-mass spectrometry (GC-MS), microbial culture, and enzyme-linked immunosorbent assays (ELISA). Although such procedures are very sensitive and precise, they are subject to some limitations [7].

These procedures usually require vast amounts of time, which may take hours or even days to yield results, thus not being suitable for perishables. Such techniques also require large laboratory space and lengthy sample preparation with advanced laboratory equipment. Additionally, these methods are largely destructive and not suitable for field-level or real-time analysis. Therefore, traditional systems based on these techniques have extremely limited applications in fast changing food safety situations, especially in those places that are decentralized and resource constrained [8]. Although miniaturization and automation have improved, such traditional systems continue to remain largely unable to meet the increased demand for portable, rapid-response, and efficient systems indispensable for food safety assurance. Due to the limitations mentioned above, spectroscopic methods have become more popular for food safety testing. Spectroscopy is a set of non-invasive, real-time, and chemically inert analytical instruments that allow for instant analysis of food matrices. Analytical methods like near-infrared (NIR) spectroscopy, Fourier-transform infrared (FTIR) spectroscopy, Raman spectroscopy, hyperspectral imaging (HSI), and ultraviolet-visible (UV-Vis) fluorescence spectroscopy have proven useful in providing detailed molecular fingerprints. These methods are applied in food quality analysis, adulterant detection, and quantification of microbial contamination by specific spectral fingerprints [9]. These methods are specifically useful in discrimination of food quality, adulteration detection, and microbial contamination quantification by individual spectral fingerprints. For instance, NIR and FTIR spectroscopy are best suited for the analysis of chemical composition and detection of common adulterants. Raman spectroscopy, owing to its higher sensitivity to dry samples, is best suited for detection of microbial and chemical contaminants. In addition, hyperspectral imaging can capture spatial and spectral data concurrently, thus enabling surface inspection of fruits, vegetables, and meat for mold, bruising, or insect

infestation detection. These characteristics render spectroscopy a promising candidate to be incorporated in contemporary food monitoring systems [10].

However, the high dimensionality, noise, and complexity of spectral data pose formidable analytical challenges. The spectral data can exhibit gigantic variability in the environment, water content, and instrumentation and are therefore time-consuming and error-prone to handle manually. Machine learning (ML) and, more recently, deep learning (DL) have proven to be very potent tools for expanding the scope of spectroscopic analysis. ML enables automated decision-making and pattern recognition processes on high-dimensional big data. Support vector machines (SVM), random forests (RF), and partial least squares regression (PLSR) are the preferred algorithms for classifying and regressing. Principal component analysis (PCA) and other dimensionality reduction methods are useful for preprocessing spectral data and therefore enhance model performance [11]. Concurrently, deep learning models, namely convolutional neural networks (CNN) and their one-dimensional (1D CNN) counterpart, have also been found to be very efficient in the information extraction from raw spectral sequences derived from Fourier transform infrared (FTIR) or near-infrared (NIR) spectrometers. The models bypass the requirement of human-crafted features and allow end-to-end learning scenarios [12,13].

Recent research has demonstrated the efficacy of hybrid approaches. Fourier-transform infrared spectroscopy (FTIR) coupled with one-dimensional convolutional neural networks (1D CNN) to identify melamine and cyanuric acid in pet food, obtaining correlation coefficients greater than 0.99, which is superior to the results achieved using partial least squares regression (PLSR) and principal component regression (PCR) techniques [10]. Tang *et al*., [14] employed surface-enhanced Raman spectroscopy (SERS) coupled with convolutional neural networks (CNN) and radiofrequency (RF) models to identify bacterial contamination in meat and dairy products, achieving an accuracy rate greater than 98% [14].

The hyperspectral imaging coupled with deep learning techniques to identify pesticide residues in grapes, achieving a performance measure greater than 93% [5]. Collectively, these reports demonstrate the efficacy of coupling spectroscopy with machine learning for detection as well as quantitation and classification of complex food contaminants with high sensitivity and specificity.

The union of spectroscopy and ML is a revolution in food safety assessment. Spectroscopy offers rapid, portable, and scalable data acquisition, while machine learning offers interpretability, predictive improvement, and the ability to automate [15]. Combined, these technologies offer a platform to identify contamination in real time and enable proactive and reactive control in food safety. The platforms are being investigated for a range of applications ranging from automatic detection of microbial contaminants in raw meat to pesticide residue analysis on fruits and vegetables, and detection of concealed adulterants in beverages. Edge computing and embedded machine learning models enable real- time analysis on portable platforms, such as handheld spectrometers and smart sensors. These technologies can transform field inspections, customs testing, and even consumer testing.

Despite highly sophisticated work, numerous challenges remain. The generalization of models to large populations of food types, standardization of preprocessing procedures, instrument-to-instrument calibration transfer, and need for large, annotated datasets are significant challenges. Interpretability and transparency of deep learning models, or "black boxes," as regulatory agency acceptance issues. These will be met through cross-disciplinary effort involving food science, analytical chemistry, data science, and policy [16,17].

Concurrent with developments in food safety, the intersection of sensor-based and spectroscopy-based technologies has shown vast transformative capabilities in environmental monitoring. In our previous contribution, we conducted a comprehensive review of these techniques for real-time

determination of water quality in freshwater aquaculture systems with an emphasis on their application in the detection of critical physicochemical parameters such as pH, dissolved oxygen, and ammonia using traditional sensors as well as spectroscopic techniques [18]. The insight gained here from that field underscores the versatility of spectroscopy over a wide range of application fields and presents a conceptual framework for extrapolating similar methodologies to food contamination analysis. This review thus builds on that established platform to explore the nexus of spectroscopy and machine learning in food safety monitoring.

This review seeks to provide a thorough overview of the status and future direction concerning the conjunctive use of spectroscopy and machine learning for food contamination analysis. The basic principles and operation of the most used spectroscopic methods are described. The machine learning algorithms used to analyse spectral data and their applications in different types of contamination, including microbial, chemical, physical, and toxin-based threats, are documented. This study places specific emphasis on recent post-2020 publications, thereby enumerating recent developments in this area. In addition, the review accounts for typical challenges and elaborates on emerging trends, including explainable artificial intelligence, hybrid sensor systems, and open-access spectral databases [19]. In general, this review seeks to provide researchers, practitioners, and policymakers with an informed perspective regarding the potential and limitations of applying these technologies in real food safety environments.

2. Spectroscopy Techniques in Food Contamination Detection

Spectroscopic methods have become leading analytical tools in modern food safety analysis because they can offer rapid, non-destructive, and chemically non-intrusive analysis. Spectroscopic methods are based on the investigation of interaction of the food constituents with electromagnetic radiation and hence offer spectral information defining the molecular structure, chemical bonding, and structural characteristics. Quality and nature of the spectral information offered are a function of the applied technique with each offering different advantages based on the matrix, type of contamination, and sensitivity needed. The following are the major spectroscopic methods used to analyse food contamination, how they work, the type of spectral information, and food safety applications [19-21].

Near-infrared (NIR) spectroscopy exploits the absorption of electromagnetic radiation between 780 and 2500 nm by molecular combination bands of vibrations and overtones, for example, O–H, C–H, and N–H bonds. Its advantages, for example, penetration deep into the material, minimal sample preparation requirements, and short analysis times, have resulted in increased use of NIR in the food sector, increasingly used to determine quality parameters like moisture, protein, and fat content, as well as to detect adulterants and contaminants. The overlapping and broad application of NIR spectra make manual interpretation time-consuming. To circumvent this, preprocessing operations such as multiplicative scatter correction (MSC), standard normal variate (SNV), and derivative transformation are used to eliminate noise and baseline drift. Coupling with Machine Learning (ML) models improves the discriminability of the NIR so that subtle chemical differences in multispectral food matrices can be detected. NIR coupled with support vector machine (SVM) and artificial neural network (ANN) for aflatoxin and antibiotic residues analysis in milk and cereal commodities and recorded high classification performance [22].

Fourier transform infrared (FTIR) spectroscopy utilizes mid-infrared radiation (4000–400 cm^{-1}) to measure the fundamental vibrational transitions of the molecular bonds, resulting in sharp and characteristic spectral peaks. This molecular fingerprinting capability allows FTIR to provide in-depth information on the presence of target chemical compounds and thus is very valuable for the

identification of food adulterants and chemical residues. FTIR spectra are more chemically resolved and selective than NIR spectra but water sensitive. Baseline correction, smoothing (e.g., Savitzky–Golay), and normalization preprocessing steps are required for reproducibility [23]. Recent application included the integration of FTIR and deep learning for end-to-end predictive models. Joshi *et al*., [10] used FTIR and 1D convolutional neural network (1D-CNN) for the identification of melamine and cyanuric acid in pet food with R^2 values greater than 0.99 [10,14]. RF to identify bacterial contamination in meat and dairy products and attained classification accuracies of >98% [6]. These examples illustrate the capability of FTIR to conduct high resolution contaminant detection with the application of advanced ML approaches

Raman spectroscopy detects the inelastic scattering of monochromatic light upon collision with molecular vibration. In contrast to IR absorption methods, Raman spectroscopy detects information relevant to molecular polarizability and is optimally used for the analysis of low-moisture food and aqueous systems. The method yields well-defined and narrow peaks that enable highly specific detection of compounds, e.g., microbial and chemical contaminants. Surface enhanced Raman spectroscopy (SERS) amplifies the weak Raman signal so that trace levels become detectable even in complicated matrices [24] . Deep convolutional neural network models to raw Raman spectra to identify bacterial contamination with a detection rate of about 100% [25]. Raman's high specificity and ML ability to identify patterns make it beneficial for on-site and laboratory-based food safety analysis.

Hyperspectral imaging (HSI) is a fusion of traditional spectroscopy with digital imaging, which generates a three-dimensional data cube of spatial and spectral information. Each pixel contains a full spectrum, enabling chemical identification and contaminant localization on food surfaces. HSI is very well suited for surface-level defect inspection, mold, bruising, and contamination of fruits, vegetables, and meat. It generates humongous high-dimensional datasets that require dimensionality reduction methods, e.g., PCA, t-SNE, or CARS, prior to ML model classification. HSI and convolutional neural networks (CNNs) to identify pesticide residues in grapes with a classification accuracy of over 93% [5]. Likewise, Ekramirad *et al*., [26] employed HSI and gradient tree boosting to classify insect infestation in apples with 97.4% accuracy. How HSI, in combination with ML, enables efficient detection of microbial and chemical contamination of foods [27].

Ultraviolet-visible spectroscopy operates within 200 to 800 nm light wavelengths. Ultraviolet-visible spectroscopy is since ultraviolet and visible radiation is absorbed by colored compounds in foods. Ultraviolet-visible spectroscopy is primarily applied to investigate electronic transitions and is more efficient in the detection of coloured compounds such as pigments and polyphenols with distinctive absorption profiles in the UV-Vis. UV-Vis spectroscopy is typically applied to identify adulterants as well as to verify color alteration due to oxidation or spoilage. The analysis is also applied to verify if food products such as juices, oils, wine, and spices are authentic. Vibrational spectroscopy methods, such as UV-Vis, for detecting milk product adulteration. The authors pointed out the effectiveness of UV-VIS spectroscopy in the detection of changes in composition and quality differences by examining spectral patterns in the visible and ultraviolet region [28]. Although less molecular-specific than FTIR or Raman spectroscopic techniques, UV-VIS spectroscopy is still a potential candidate due to the ease of use, the rate of data acquisition, and field-portable instrument compatibility.

Fluorescence spectroscopy is based on the phenomenon of luminescence of a compound on the absorption of visible or ultraviolet light. The analytical method is highly sensitive and is able to quantify naturally fluorescent and labelled analytes in very trace amounts. Nan *et al*., [6] discussed the use of fluorescence-based methods for determining ochratoxin A and patulin mycotoxins in fruit samples. A label-free, fluorescence amplified electrochemical apt sensor for the determination of

paraquat residues in vegetable samples, thus developing an ultra-sensitive and selective detection system most suitable for field testing [4]. These studies collectively illustrate the growing relevance of fluorescence and UV-Vis spectroscopy in food contaminant tracking, particularly when coupled with machine learning methods for quantitative determination and classification.

These spectroscopic methods provide an array of equipment for detecting contaminants in different food categories. For example, HSI allows for non-destructive screening of fruits, whereas FTIR and Raman spectrometry analyse adulterants at the molecular level. Their use depends on food matrix complexity, nature of the contaminant, requirement of sensitivity, and working range. If used along with sophisticated machine learning algorithms, their analytical capability is further increased to screen out the contaminants in real-time, high-throughput, and high accuracy throughout the entire food supply chain.

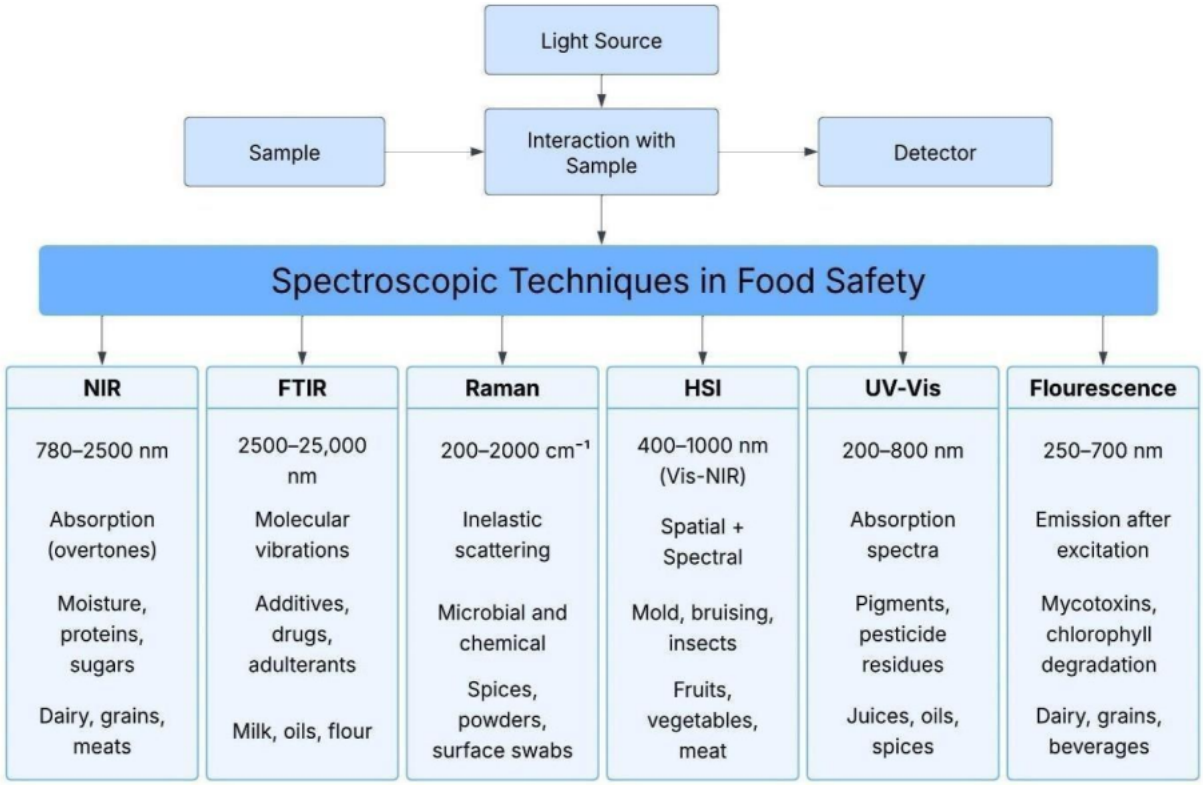

Fig. 1. Schematic representation of major spectroscopic techniques used in food contamination detection

3. Machine learning

The union between spectroscopy and machine learning (ML) has revolutionised food safety analysis, allowing sensitive, fast, and non-destructive identification of contaminants at near real-time. Although several spectroscopic methods near-infrared (NIR), Fourier-transform infrared (FTIR), Raman, ultraviolet-visible (UV-Vis), and hyperspectral imaging (HSI) can provide chemical and physical information from food matrices, the application of powerful computing capacity is needed to realize the maximum benefits of these methods [34]. Raw spectral data provided by these methods are by nature high-dimensional consisting of thousands of variables that are highly correlated to describe reflectance or absorbance at many wavelengths. In addition to this, data are influenced by noise, baseline drift, light scattering, environmental variability, and instrument variability and therefore are hard to interpret.

To solve such problems, linear chemometric methods, including principal component analysis (PCA) and partial least squares regression (PLS-R), have been conventionally employed [29]. These are nonetheless linear assumption-based methods and might fail to capture the nonlinear, complex dynamics of real spectroscopic data. With increasingly complex food matrices and increasingly complex contaminants from microbial, chemical, and physical adulterants, there is equally an urgent demand for equally accurate but also flexible, scalable, and capable of learning hierarchical representations directly from raw inputs. Machine learning has proven to be a suitable framework for the above demands, facilitating automated pattern recognition, anomaly detection, and measurement with minimal or no human intervention [30]. Developments in the last decade in deep learning (DL), including convolutional neural networks (CNNs), support vector machines (SVMs), random forests (RF), and ensemble methods, have proven to provide better performance in spectroscopic applications compared to traditional methods. The models particularly excel in recovering information from large, complex data, detecting weak spectral signatures of contamination that would be undetectable to human experts, as well as traditional algorithms [31] [32].

Applications of ML to food spectroscopy span the entire data pipeline, from signal preprocessing and feature selection to model training and performance estimation. Preprocessing operations attempt to denoise and normalize the spectral input to minimize the impact of external and instrumental variability. Feature selection techniques specify the most discriminatory wavelengths or bands of spectral data that are useful for discrimination or prediction. Classification models are trained to separate contaminated and clean samples, and regression models are employed to predict contaminant levels or degradation levels. Finally, rigorous model validation ensures robustness, generalizability, and acceptability for regulatory use in food inspection pipelines [33].

This section provides each part of the ML pipeline within the context of spectroscopic data for food contamination detection. Particular focus is on recent developments, best practice in methodology, and application-specific challenges, following recent empirical studies and overall reviews post-2021. By the overview of each step in the ML pipeline, this review aims to provide researchers and practitioners with a clearer idea of how such methods facilitate the development of next-generation food safety diagnostics that are not only accurate and trustworthy but also field-deployable, scalable, and interpretable.

3.1 Preprocessing

Preprocessing is a critical step in the application of machine learning (ML) algorithms to spectroscopic data analysis in food safety, for example. Food sample spectra obtained through methods like near-infrared (NIR), Fourier-transform infrared (FTIR), Raman, or hyperspectral imaging (HSI) will typically contain redundant variability in the form of distortions. This variability will commonly take the form of random noise, baseline drifts, effects of light scattering, and inter-sample variability as a function of parameters like moisture content, surface roughness, or compositional heterogeneity. Without preprocessing, such differences will dampen weak but meaningful chemical signatures, thus undermining the effectiveness and integrity of ML models [34].

To have coherence and to increase detection of informative signal, researchers apply preprocessing techniques like Standard Normal Variance (SNV) and Multiplicative Scatter Correction (MSC) to eliminate the influence of path length and scatter, particularly with powdered or irregular samples. Savitzky Golay filtering is also commonly applied to enhance spectral information by reducing high-frequency noise without diminishing peak shape integrity. First- and second-order

derivatives are utilized for improving peak features, facilitating deconvolution of overlapped spectral bands, and eliminating baseline drift [35].

One such recent systematic study highlighted the importance of preprocessing in the majority of machine learning efforts in the identification of mycotoxins and other food adulterants. The review showed that more than 85% of published research employed preprocessing methods, and models trained on raw spectra were always inferior to those trained on pre-processed data [36]. Specifically, the use of SNV, derivatives, and PCA dependent dimensionality reduction was linked to the optimal classification accuracy in studies conducted with FTIR and NIR spectra In in-line or handheld food analysis applications, where operator-dependent sensor calibration variation, illumination, and control are larger, standardized preprocessing is mandatory. As ML-based spectroscopic technology is more widely applied in non- laboratory settings, high-quality preprocessing pipelines will be necessary to ensure model accuracy, interpretability, and acceptability to regulators. Some newer architectures have also built in automated preprocessing algorithms, where optimal transform selection is dynamically determined based on the properties of the dataset a fast-growing area of interest for making ML workflows more scalable and less human-optimization dependent [37].

3.2 Feature Selection

The most critical steps in the analysis of spectral data are feature extraction and selection followed by preprocessing. Because spectroscopic data typically contain hundreds or thousands of variables, each a unique wavelength or frequency, the most critical feature is dimensionality reduction without information loss related to food contamination. Not only does this enable computationally efficient and better-performing models, but also generalizability and minimization of overfitting. Of the older techniques, PCA is mostly used to project data into a lower-dimensional space without losing most of its variance. PCA is useful for exploratory data analysis and visualization but is an unsupervised method and will not necessarily select the best features for a specific classification or regression problem. More advanced techniques, such as competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA), and recursive feature elimination, have been proposed to address this by picking variables on the basis of correlation with class labels or response variables. These are especially useful when diagnostic information is carried only by thin wavelength bands, an issue commonly faced in adulterant detection or microbial contamination research [38]. A case in point is within their evaluation of hyperspectral imaging in food microbiology, where the necessity of dimensionality reduction to prevent the curse of dimensionality inherent in spectral imaging systems. From their findings, feature selection methods not only reduce the computational burden but also lead to more robust model performance when coupled with machine learning classifiers such as SVMs or random forests (RFs). The past decade has seen improvements in the deep learning framework give rise to end-to-end architectures such as convolutional neural networks (CNNs), which learn hierarchical spectral representations from lightly processed data. This is revolutionary since it eliminates hand-engineered feature selection and can lead to better accuracy with big data. In food safety applications that still remain dominated by small data, traditional feature selection routines remain the dominant choice since they are interpretable and can result in identification of specific chemical markers. The selection of using traditional vs. deep learning-based feature extraction is thus a function of the specific analytical setup, data availability, and interpretability vs. the performance trade-off [39,40].

3.3 Classification Algorithms

Following feature extraction, classification models are the core component of the majority of machine learning (ML) applications to spectroscopic food contaminant analysis. The models are trained to distinguish between classes such as "contaminated" vs. "noncontaminated" samples or between specific microbial strains or adulterants. Of the classical models, support vector machines (SVMs) are especially popular as they are capable of handling high-dimensional and nonlinear data with the aid of kernel functions. They are also resilient, even with relatively small datasets, and are less prone to overfitting [41]. Likewise, k-nearest neighbors (k-NN) offer an easy-to-implement solution for local proximity-based classification in feature space but are not robust and prone to noise and irrelevant features [42]. Ensemble models such as decision trees and random forests (RFs) are widely employed for their robustness and interpretability; RFs, in turn, calculate the mean of the output of numerous trees to reduce variance and enhance generalization. More advanced models such as artificial neural networks (ANNs) and convolutional neural networks (CNNs), are capable of extracting the nonlinear relationships regardless of manual feature selection. CNNs, especially in their 1D and 2D implementations, have been effectively employed to directly process spectral and hyperspectral data. For instance, in a recent study, Liu *et al*., [39] employed CNNs for tea quality classification via near-infrared spectra, with enhanced performance compared to conventional classifiers, demonstrating the growing applicability of deep learning models to food analysis [43].

3.4 Regression Algorithms

Apart from classification, most spectroscopic machine learning applications in food safety are quantitative prediction, such as quantitation of contaminant concentration or extent of spoilage. Regression models are the target for these. Partial least squares regression (PLSR) has been the most prevalent chemometric tool employed for spectroscopic quantitation for the last two decades. PLSR is ideally suited to linear modeling of target output against spectral input even when there is multicollinearity [44,45]. For handling nonlinearity and higher-order interactions, however, more sophisticated regression techniques have come to the fore. Support vector regression (SVR), a direct extension of support vector machines to continuous output variables, has also gained popularity considering robustness and insensitivity to outliers. Neural network-based regression is also highly promising, especially in deep learning models. For example, a recent review of infrared spectroscopy in food safety applications highlighted the following case studies: one used second derivative preprocessing of near-infrared spectra for pesticide residue quantitation in strawberries and cabbage using PLSR and LS SVM. Predicted prediction correlation coefficients (RP) were above 0.93 with a root mean square error of prediction (RMSEP) below 3.22 mg kg^{-1}, while LS SVM was superior to PLSR, highlighting the advantages of nonlinear regression in complex matrices. These studies indicate that while PLSR remains of interest, the application of nonlinear regression techniques can deliver high accuracy, especially in applications involving chemically heterogeneous composition or trace-level analytes [22]. With more stringent food safety laws and sample heterogeneity, the application of strong regression models based on high-quality spectral data has become an imperative for quantitative, accurate food contaminant analysis.

3.5 Model Evaluation Metrics

To ascertain the performance and reliability of such models, strict performance metrics have to be employed. For classification, some of the most widely employed metrics include accuracy,

sensitivity (true positive rate), specificity (true negative rate), and precision, and F1 score. Accuracy provides an overall success rate, while sensitivity and specificity provide specific interest in food safety contexts where false negatives (i.e., failure to identify a contaminant) are of particular concern. For regression models, evaluation can include the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). These assess the quality of the predicted values against true measurements [46]. NIR spectroscopy coupled with random forest classification, which was 97.7% accurate on the test set. They also employed Support Vector Regression (SVR) for the prediction of adulterant content, $R^2 > 0.98$, RMSE < 1.7%, which showed the model's accuracy for both classification and quantitative uses. Their study is an outstanding example of good practice in model validation, e.g., the application of k-fold cross-validation and the application of more than one metric of performance to render models robust and deployable in food safety contexts [18]. The figure 2 below shows the key steps in the process of applying machine learning methods in spectroscopic food analysis.

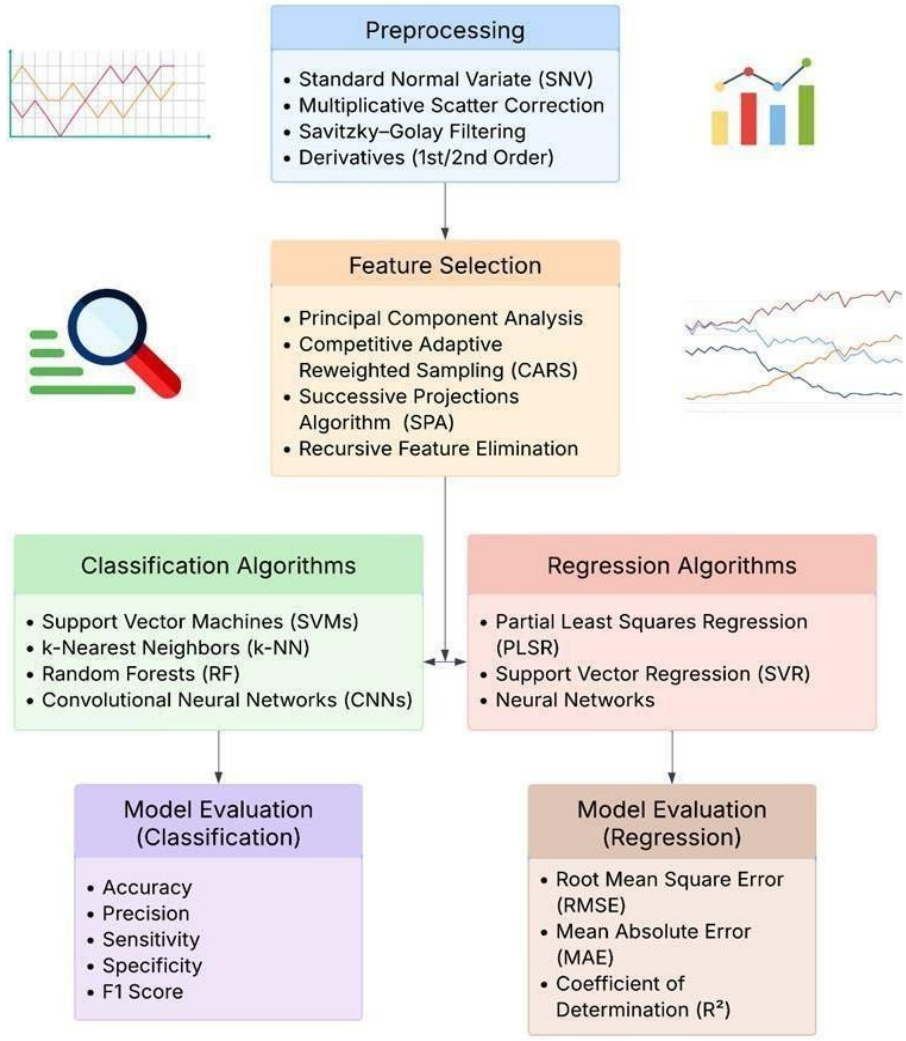

Fig. 2. Machine learning pipeline for spectroscopic food contamination analysis

4. Integration of Spectroscopy and Machine Learning: Applications In Food Contamination Detection

Spectroscopy combined with machine learning has transformed food contaminant analysis to produce fast, non-destructive, and scalable methods that supplant the traditional assays. The analytical power of spectroscopic platforms typically limited by noise, high dimensionality, and non-linearity is significantly improved when coupled with ML methods capable of extracting latent features and recognizing discriminative patterns. In our earlier work, we explained how Wireless Sensor Networks (WSNs) and Internet of Things (IoT)-based systems can facilitate smart, real-time monitoring of the environment in agriculture, aquaculture, and smart cities. The results on energy-efficient data acquisition, wireless communication, and sensor deployment strategies provide a relevant background for the implementation of similar smart monitoring systems in the application of food safety applications, particularly in the integration of spectroscopy and machine learning models. This section offers a critical overview of the use of such hybrid platforms in solving microbial, chemical, and physical contamination problems in a broad range of food matrices. Recent studies between 2023 and 2025 have demonstrated advancements in model structure, spectral preprocessing, and data fusion techniques. A concise table is presented at the end of this section.

4.1 Microbial Contamination

Food contamination by microbes is the biggest food safety issue in the world. The most prevalent foodborne illness causing microbes are Escherichia coli, Salmonella, and Listeria monocytogenes. Culture- or quantitative polymerase chain reaction-based detection platforms are now being drastically expedited and field-deployed through the help of spectroscopic sensors and ML pipelines [8].

In 2024, Wisuthiphaet *et al*., [47] used a bacteriophage-induced excitation–emission matrix (EEM) fluorescence spectroscopy method to find E. coli in spinach homogenate. Their work demonstrated that phage-induced spectral changes—indirect indicators of bacterial presence—were accurately described by a Gaussian Process classifier with more than 92% classification accuracy at bacterial concentrations as low as 10^2 CFU/mL [47]. This was achieved with only six important wavelengths, selected based on feature importance ranking, and thus eliminating spectral redundancy and allowing for deployment on low-resolution, miniaturized optical devices. These advancements demonstrate that the incorporation of spectral fingerprinting into traditional machine learning models enhances the sensitivity of real-time food safety measurements.

In deep learning, a multi-scale convolutional neural network (CNN) for Salmonella serovar classification from Raman spectral data. Preprocessed spectra through Savitzky–Golay smoothing and normalized using standard normal variate (SNV) were employed as input to the CNN, which consisted of hierarchical convolutional layers optimized for the extraction of local vibrational peaks as well as more general spectral pattern. The model was confirmed to be approximately 98.4% accurate and was also significantly robust against baseline drift, demonstrating the application of deep spectral structures to classification problems of high complexity [48].confocal micro-Raman spectroscopy to perform one of the largest pathogen identification trials. The research enhanced computational efficiency by integrating classic classifiers such as support vector machines (SVM) and random forests (RF) with that of CARS-SPA feature selection, optimally thus enhancing the input feature space. Importantly, it employed Shapley Additive Explanations (SHAP) to determine which Raman shifts had the greatest impact in classification, thus evading the long-standing issue of machine learning "black box" behavior in regulatory contexts. This degree of transparency is an

important step towards the application of ML-spectroscopy tools towards developing actionable and auditable diagnostic protocols [49].

Collectively, these studies recognize both the utility and growing maturity of machine learning-based spectroscopic methods in microbial contamination analysis. In addition, they demonstrate a clear evolution from simple SVM models to increasingly complex explainable and deep learning models capable of distinguishing different pathogens with little preprocessing necessary.

4.2 Chemical Contamination

Trace amounts of chemical food pollutants such as pesticide residues, veterinary medications, and heavy metals are of significant toxicological interest even at trace concentrations. Food chemometrics' use of non-destructive identification of these contaminants has been a long-standing dream, and the developments in spectroscopic instrumentation complemented by its ML models have well advanced the limits of what is possible [50].

The short-wave infrared hyperspectral imaging (SWIR-HSI) for the detection of pesticide residues on the surface of Hami melon. The model used a 1D convolutional neural network (1D-CNN) that used spatial and spectral information, hence capable of sensing chemical absorption and microstructural parameters that signified contamination. Background correction, median filtering, and normalization were the preprocessing methods adopted that rendered the performance of the model to be insensitive to changes in ambient light. The resulting model achieved classification accuracies of approximately 94%, which is comparable to traditional chromatography techniques, but without the need for any reagents or sample destruction [51].

Though classification might be necessary, chemical analysis will usually demand quantitation, determining the levels of concentration of the impurities. These models, including partial least squares regression (PLSR), Support Vector Regression (SVR), and Multilayer Perceptrons (MLPs), have functioned according to this trend [52]. For instance, Ye *et al*., [5] showed the application of hyperspectral imaging (HSI) for pesticide grape residues discrimination using support vector machines (SVM) and convolutional neural networks (CNN). Their models' accuracy range was 93.3% to 97.2% depending on the fruit variety type and the spectral preprocessing methods applied. While regression analysis as such was not employed in their article, most recent papers on pesticide quantitation by near-infrared (NIR) and Raman spectroscopy typically report R^2 values > 0.90 with root mean square error of prediction (RMSEP) values < 2 mg/kg when feature selection methods like variable importance in projection (VIP) and genetic algorithms (GA) are applied. Whether the models can be transferred across fruit types is being investigated. Some authors have begun to tinker with transfer learning—train on one fruit, test on another with some success, progressing to more general contamination sensors [53].

In conclusion, chemical contamination analysis has come a long way from the conventional static lab testing to more dynamic, mobile, and intelligent systems that can make binary classification as well as quantitative prediction. Hybrid deep learning algorithms, particularly with spatial contexts.

4.3 Physical Contamination and Adulteration

Identification of these physical impurities, such as plastics, bone, stones, and insect material, by conventional optical inspection systems is challenging owing to their varied morphology and close spectral resemblance to the food matrix [54].

The integration of spectroscopy with object-detection-based machine learning algorithms has enabled novel means of real-time and non-destructive physical contamination screening.

A 2023 paper on chili peppers employed Vis-NIR imaging with pixel-wise classification (with SVM and RF) and subsequent object-level detection with YOLOv5. The proposed system achieved 86%-pixel classification accuracy and 96% object detection accuracy, which are appropriate for conveyor belt automation in factory sorting lines. Of particular interest, the paper emphasizes the importance of hybrid pipelines where spectral segmentation comes before object recognition, enhancing interpretability and minimizing false positives [55].

The challenge of detecting small, heterogeneous foreign contaminants in chicken meat using a semi-supervised Generative Adversarial Network (GAN) model trained with a 1D U-Net from near-infrared (NIR) spectral information. In contrast to supervised methods that need large, annotated datasets, their approach allowed the model to learn spectral representations from pure data directly while identifying labelled anomalies as potential contaminants. The model achieved a high F1-score of 96.8% and demonstrated good generalization capability for new object types. This new method reduces the annotation cost and is particularly well-suited to scenarios with high contamination uncertainty [56].

These principles in rice food processing operations. They proposed a new YOLOv-MA (multi-scale dilated attention) framework, which resulted in a 5%–11% mAP increase over the traditional YOLOv5 in the detection of contaminants such as husk fragments and soil dust. Their research underscores the imperative to develop adaptive architectures that are capable of learning across multiple spatial resolutions a concern of growing significance to many food matrices [57].

Vis-NIR HSI to detect minced meat adulteration, or beef-chicken mixing. Spectral-spatial transformation and SVM classification resulted in 94.9% accuracy. The application of this technique highlights the increasing feasibility of spectral fingerprinting in combating economically motivated adulterationa large compliance issue in developed and emerging economies [58].

4.4 Application across Food Matrices

Apart from the type of contamination, spectroscopy–ML integration performance also greatly depends on physical and chemical properties of the target food matrix. Water content, for instance, surface roughness, biochemical composition, influences the quality of the spectrum and model performance [59].

In highly aqueous foods such as meat and milk, overlapping bands of water absorption and scattering effects make analysis challenging. however, demonstrated the effectiveness of short-wave infrared hyperspectral imaging (SWIR-HSI) with a 1D convolutional neural network (CNN) in the identification of Escherichia coli, Listeria monocytogenes, and Staphylococcus aureus on mutton surfaces. Its success was thanks to sophisticated spectral preprocessing, i.e., the use of first-derivative transformations and continuum removaland the capacity of the CNN model to learn features irrespective of changing moisture conditions. The system was capable of generating a classification accuracy of approximately 98.6%, vouching for its credibility and its potential in meat-processing facilities [60].

Fermentation products are challenging since their composition varies with time. Inception-based 2D-CNN model with NIR spectra of bacterial broth as the training dataset. Their model yielded $R^2 \approx 0.90$ and RMSE ≈ 0.52 in the estimation of the total viable counts (TVC). The utilization of inception modules, which have the capability to process multiple receptive fields in parallel, allowed the model to learn across batch conditions and thus be usable for in-line quality control [61].

Together, these pieces of work demonstrate that model building and preprocessing will need to be modified not only to the type of contamination but also to the matrix-dependent optical response.

Next-generation systems will probably incorporate adaptive algorithms that are capable of self-calibrating to changing food environments.
Recent uses of spectroscopy and machine learning for food contamination detection are outlined in Table 1.

Table 1
Summary of recent studies integrating spectroscopy and machine learning in food contamination detection

Contamination Type	Food Matrix	Spectroscopy Technique	ML Model	Performance	Reference (Year)
Microbial (E. coli)	Spinach homogenate	EEM fluorescence	SVC, RF, Gaussian Process	Accuracy > 92 %	[62,63]
Salmonella serovars	Bacterial isolates	Raman spectroscopy	Multi-scale CNN	Accuracy ~ 98.4 %	[64,65]
Mixed pathogens (9 types)	Bacterial isolates	Confocal microRaman	SVM, RF + CARS-SPA + SHAP	Accuracy ~ 98.9 %	[19]
Chemical (pesticides)	Hami melon	SWIR-HSI	1D-CNN + fusion	Accuracy ≈ 94 %	[20]
Physical (foreign objects)	Chili peppers	Vis-NIR HSI + YOLOv5	RF, SVM, YOLOv5	Pixel 86 %, Detection 96 %	[21]
Physical (foreign matter)	Poultry breast	NIR-HSI + Semi-supervised GAN	1D U-Net + GAN	F1-score 96.8 %, Accuracy 96.5 %	[22]
Physical (foreign objects)	Rice production	Vis-NIR imaging + YOLOv-MA	YOLOv-MA	+5–11 % mAP gain	[23]
Adulteration (meat)	Beef mince	Vis-NIR HSI	SVM + spectral transformation	Accuracy 94.9 %	[24]
Microbial (pathogens)	Mutton	SWIR-HSI	PLS-DA, SVM, RF	Accuracy 94.3 %	[25]
Quantitative microbial load	Fermentation broth	NIR spectroscopy	Inception 2D-CNN	$R^2 \approx 0.90$; RMSE ≈ 0.52	[26]

Despite significant progress in the integration of spectroscopy and machine learning (ML) for food contamination detection, a number of sophisticated and interrelated challenges remain to stand in the way of their advancement from laboratory research to practical and scalable deployment [62]. These challenges are not merely technical hurdles—such as equipment variability, data dimensionality, and algorithmic computational expense—but also more general infrastructural and regulatory hurdles. The integration of spectroscopy with ML often entails the analysis of high dimensional noisy data from various food matrices, each having unique chemical and physical characteristics dictating spectral behaviour. Additionally, the sensitivity of spectrometric response to the environment, operator handling, and equipment configuration presents a further layer of complexity that must be successfully navigated by ML models. This challenge is extremely problematic when models are deployed over many instruments or geographic locations, where no standardization is the norm. Additionally, the dependence of machine learning models on the quantity and quality of annotated training data is extremely high. For food, the collection of large and diverse annotated datasets is not only time consuming but is also constrained by the food samples' perishable nature and diversity. Algorithms that show robust performance under laboratory-controlled conditions fail to generalize in industrial or field settings. Additionally, the lack of standard acquisition protocols, the computational

expense of real-time execution on edge hardware, model interpretability, and regulatory clearance are all present formidable challenges [63]. Addressing these multifaceted challenges is necessary to fully realize the potential of spectroscopy-ML systems in facilitating scalable, accurate, and non-invasive solutions for the next generation of food safety monitoring.

5. Challenges in Integration

One of the primary and most common problems in the machine learning-spectroscopy integration for the detection of food contamination is high spectral variability due to intrinsic and extrinsic variables. Food matrix spectra are parameters that depend on moisture content, texture, color, and particle size distribution. e.g., in NIR and MIR spectroscopy, samples with high water content (e.g., fresh meat, vegetables, or fruit) yield strong water absorption bands, which mask underlying detail about contaminants. Surface roughness heterogeneity of contaminant scattering and distribution artifacts, especially with diffuse reflectance methods, also cause baseline shifts. Extrinsic sources like instrument calibration, non- uniform illumination source, and non-uniform ambient experimental environmental conditions (e.g., temperature or humidity) also add to the problem. Minor variations in the incidence angle, path length, or sample position can add to nonreproducible spectral profiles. These changes significantly reduce the robustness and reproducibility of the machine learning models trained on such spectra, especially when such models are implemented using different instruments or sample acquisition geometries [64]. When the same preprocessing protocols, e.g., standard normal variate (SNV) and multiplicative scatter correction (MSC), were employed, model transfer of NIR models from one miniaturized device to another resulted in a catastrophic loss of classification accuracy. The observation underscores the critical importance of advanced calibration transfer methods, spectral standardization protocols, and standardized acquisition protocols to guarantee device-agnostic MLspectroscopy pipeline performance under a wide range of operating conditions [65].

Another persistent issue is the high dimensionality of spectral data and the consequent risk of overfitting, especially when working with small or sparsely distributed datasets. Hyperspectral imagery and spectral data are normally liable to contain hundreds or thousands of bands, most of which are redundant or collinear. Redundancy is computationally costly and increases the risk of fitting noise instead of an informative signal. Dimension reduction techniques such as PCA, variable selection techniques such as competitive adaptive reweighted sampling (CARS) and genetic algorithms (GA- SPA) are commonly employed, but their performance is normally outshone by natural variability and noise in the data. This issue is much more pronounced in deep learning models such as CNNs, where they need vast amounts of training data to prevent memorization of patterns corresponding to individual batches or environments [66,67].In their work on monitoring fermentation operations, The robust initial performances ($R^2 \approx 0.90$, RMSE ≈ 0.52) with NIR spectra-trained CNN models. The models, however, had a sharp decline in performance about 10% when they were tested on unseen batches, even after the use of the same instrumentation and preprocessing protocols. This situation indicates the burning need for mechanisms such as cross-batch normalization, data augmentation, dropout regularization, and domain-invariant learning to provide the robustness and reproducibility of models in different conditions [61].

Overfitting is also directly related to the lack of well-defined standard operating procedures for data collection, preprocessing, and metadata documentation, which severely hinders reproducibility across different laboratories and institutions. Additionally, spectroscopic workflows also have inconsistencies in details that might appear minor; however, these parameters are more important—such as the choice of spectral range, resolution parameters, illumination geometry, sample thickness,

and background correction techniques—each introducing variability that influences model transferability. Additionally, preprocessing methods used on raw spectral data differ in studies seldom [68]. The variations in ambient light and sample handling could alter classification accuracy in portable FTIR instruments regardless of concealed sample types. In most instances, critical metadata concerning acquisition conditions, instrument calibration, and preprocessing protocols are not recorded and render replication of published results impossible. Such a condition calls for universally accepted spectral data standards and reporting protocols, including the FAIR principles (Findability, Accessibility, Interoperability, and Reusability), to enhance collaborative spectral data sharing and cross-validation among various laboratories [69].

The use of spectroscopy–machine learning systems in field-portable or edge devices presents new challenges of sensor performance, algorithmic speed, and real-time responsiveness [70]. In contrast to high-spectral resolution laboratory-based spectroscopic instruments with abundant computational resources operating in controlled environments, field-portable instruments need to cope with severe compromises. Miniature spectrometers are beset by low dynamic range, low signalto-noise ratios, and limited spectral bands, which compromise their sensitivity to trace-abundance analytes. As an additional challenge, the use of machine learning models in these devices necessitates extensive algorithmic fine-tuning since onboard processors are incapable of executing standard convolutional neural network (CNN) models or ensemble methods without suffering the penalty of compromised latency or overheating. In a recent article, A portable spectroscopic platform with on-device convolutional neural networks (CNNs) for real-time prediction of moisture and protein content in raw cereals [71]. The method showed the applicability of TinyML models for fast quality inspection in field settings; however, the system also showed subtle decreases in predictive accuracy compared to benchscale analysis, especially in variable optical density samples. The results illustrate the unavoidable performance– portability tradeoff, where mobility and real-time access optimizations come at the price of compromised analytical accuracy. To limit this gap, future research should explore hybrid architectures that bridge local feature extraction on edge devices with secure, low-latency cloud-based inference without sacrificing speed while preserving spectral fidelity.

The greatest and perhaps most challenging task is the machine learning model generalizability to generalize across diverse food matrices, contaminant types, and operating conditions. Models are good when trained and tested on one matrix (e.g., apples or grapes); however, accuracy is noted to be decreased upon expansion to other food materials due to variations in spectral characteristics. The variations occur as a result of variations in texture, water content, pigment, or chemical characteristics. Calibration transfer and domain adaptation techniques seek to close this gap by transferring models from small new-domain sets [72,73]. The investigated transferability of NIR-based SSC models among strawberries, grapes, and apples using calibration transfer techniques such as PLS correction and semi-supervised parameter-free calibration (SS-PFCE). They proved that by intentional updating of samples (e.g., 20% new-domain calibration), SSC prediction errors could be minimized to be as close as those of fruit-specific models—thus proving the applicability of cross-matrix spectral modeling to food quality determination [74].

Overall, solving these difficult tasks is at the core of the evolution of spectroscopy-ML–ML systems into deployable, generalizable, and strong solutions. Investment in data standardization, model interpretability, and hardware-software co- design will be required to optimize technology usage for next-generation food safety monitoring.

6. Recent Trends and Future Prospects

Along with such technologies that use machine learning (ML) and spectroscopy moving from laboratory experiments to practical applications in industry and commerce, the field is going through its very core technological change [75]. Although recent emphasis has been on moving away from achieving high detection accuracy to making these systems interpretable, scalable, and hardware-efficient, as well as data privacy law-compliant, academics are answering the growing call for real-time, embedded applications in food safety by investigating lightweight deployment techniques, such as TinyML In-device inference is supported by TinyML through techniques such as model quantization, pruning, and architecture optimization [76]. These advances enable the deployment of convolutional neural networks and other predictive models on microcontrollers in handheld spectroscopic instruments, enabling localized, low- power, and rapid detection of contaminants without the use of cloud resources. Likewise, highlighted that the future of embedded intelligence is in the balance of energy-efficient hardware and small model architectures, particularly in applications such as food quality monitoring, where latency, portability, and cost effectiveness are paramount [77].

In addition to efficiency, the new trends include multi-institutional collaboration and data privacy. Commercial and institutional data silos have hindered the general demand for annotated spectral data in diverse food matrices. Federated learning (FL) is a promising substitute by facilitates model training in a decentralized manner across multiple sites or devices without the need for centralized data pooling. Fendor *et al*., [78] were highlighted in a systematic review of more than 40 FL applications in the food and agriculture industries, showcasing their relevance in pesticide monitoring, water safety, and food spoilage forecasting. However, the challenges of heterogeneity in sensor platforms, nonuniform data distribution, and issues in synchronization remain paramount. The constraints of vertical FL, particularly when consumers inhabit varying feature spaces or sensor modalities, present challenges in food spectroscopy when acquisition protocols and hardware are diverse. The combination of FL with TinyML and swarm learning platforms for IoT networks, once more affirming the suitability of such structures for real- time, privacy-respecting contamination tracking in dispersed food networks [78].

One of the most important issues of concern is the explainability of machine learning (ML) decisions, especially in highly regulated industries like food safety. Although extremely accurate, conventional deep learning architectures will continue to be black box systems and are thus not generally accepted by industry auditors and government regulators. SHapley Additive exPlanations (SHAP) to hyperspectral classifiers and demonstrated how important spectral bands impacting model decisions can be visualized and understood [79]. In another work, Zhang and Abdullah introduced an XAI-guided wavelength selection approach that can be applied to honey authentication [80]. By integrating the SHAP and LIME techniques, they achieved better classification performance while significantly lowering the dimensionality of the hyperspectral input data. Such approaches not only build trust in AI-based decisions but also improve model performance by eliminating noise or redundant information. Arrighi *et al*., [81] recently systematically reviewed and categorized XAI approaches for food datasets and provided a framework for choosing the right interpretability methods based on model type, food matrix complexity, and application intent. The use of such approaches is likely to become a routine requirement for ML models employed in food quality control and contamination risk evaluation.

At the same time, there is a growing need for reproducible and automated machine learning pipelines. Spectral models must be retrained periodically due to changes in food matrices, sensor settings, or environmental conditions. Software frameworks like TensorFlow Lite, MLflow, and ONNX

Runtime are designed to meet this need by providing model versioning features, performance monitoring metrics, and automated retraining cadences. These frameworks also provide continuous concept drift detection, which is critical in cases where patterns of contamination are seasonal-based or where new supply chains are added to the system. By integrating such deployment frameworks into food inspection workflows, models can be updated with minimal user effort or intervention and hence attain scalability and long-term sustainability.

Lastly, the value of open-access spectral archives is increasingly being appreciated. The reproducibility of contemporary research is often marred by proprietary data and inconsistent acquisition protocols. Initiatives to develop shared spectral repositories with contaminant levels, sample data, and sensor data annotations is becoming critical in facilitating benchmarking and transfer learning. The datasets can standardize preprocessing for any data across the board and enhance model generalization across food groups and geographies. Furthermore, the availability of datasets will facilitate the development of machine learning infrastructures that can be tested and certified for commercial applications across institutions.

Briefly, the convergence of machine learning and spectroscopy for food contamination detection is moving toward deployable, explainable, and privacy-aware systems. Developments in TinyML, federated learning, explainable AI, modular deployment software, and shared data infrastructures are not only enhancing technical capability but also overcoming the practical constraints in the context of food safety monitoring. Interdisciplinary collaboration will be central to be globally scalable and compliant with regulatory standards, hiring food scientists, machine learning engineers, software developers, and policymakers on a common mission of secure, transparent, and intelligent food quality assurance.

7. Conclusions

The integration of spectroscopy and machine learning (ML) has provided a new era of non-destructive, rapid, and intelligent food contamination analysis. This review critically examined the spectroscopic method repertoire ranging from traditional NIR, FTIR, and UV-Vis to more advanced Raman and hyperspectral imaging and their complementary application with ML algorithms for classification, quantitation, and anomaly detection in foods. The experiments on microbial, chemical, and physical contaminants have demonstrated that the synergy between spectral resolve and smart data analysis has the potential to outperform conventional laboratory-based assays as real-time, scalable, and portable diagnostic tools. Despite this progress, several interrelated challenges remain. They include spectral variation, small, annotated datasets, model transferability, real-time edge deployment challenges, and the absence of standardized acquisition and preprocessing protocols. New technologies like TinyML, federated learning, XAI, and open-access collaborative databases hold the key to revolutionary solutions for such challenges. Yet, interdisciplinary collaboration would still be required to close the gap with deployable and regulation-compliant systems. The strategy for ML-based food safety spectroscopy requires scalable system design, interpretable modelling pipelines, and harmonized evaluation frameworks. Investments in edge cloud hybrid architectures, cross-domain spectral modelling, and open benchmarking datasets are crucial. The intersection of analytical chemistry, computer vision, embedded systems, and food regulatory science will enable the future of food safety monitoring to transition from reactive risk management to proactive, global, smart surveillance systems with worldwide reach.

Acknowledgement
This research was not funded by any grant.

References

[1] "Food safety." Accessed: Jun. 23, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/food-safety

[2] Onyeaka, Helen, Soumya Ghosh, KeChrist Obileke, Taghi Miri, Olumide A. Odeyemi, Ogueri Nwaiwu, and Phemelo Tamasiga. "Preventing chemical contaminants in food: Challenges and prospects for safe and sustainable food production." *Food Control* 155 (2024): 110040. https://doi.org/10.1016/j.foodcont.2023.110040

[3] Wahab, Shadma, Khursheed Muzammil, Nazim Nasir, Mohammad Suhail Khan, Md Faruque Ahmad, Mohammad Khalid, Wasim Ahmad, Adam Dawria, Lingala Kalyan Viswanath Reddy, and Abdulrahman Mohammed Busayli. "Advancement and new trends in analysis of pesticide residues in food: A comprehensive review." *Plants* 11, no. 9 (2022): 1106. https://doi.org/10.3390/plants11091106

[4] Wu, Qiaoling, Han Tao, Yuangen Wu, Xiao Wang, Qili Shi, and Donglin Xiang. "A label-free electrochemical aptasensor based on Zn/Fe bimetallic MOF derived nanoporous carbon for ultra-sensitive and selective determination of paraquat in vegetables." *Foods* 11, no. 16 (2022): 2405. https://doi.org/10.3390/foods11162405

[5] Ye, Weixin, Tianying Yan, Chu Zhang, Long Duan, Wei Chen, Hao Song, Yifan Zhang, Wei Xu, and Pan Gao. "Detection of pesticide residue level in grape using hyperspectral imaging with machine learning." *Foods* 11, no. 11 (2022): 1609. https://doi.org/10.3390/foods11111609

[6] Nan, Mina, Huali Xue, and Yang Bi. "Contamination, detection and control of mycotoxins in fruits and vegetables." *Toxins* 14, no. 5 (2022): 309. https://doi.org/10.3390/toxins14050309

[7] Haji, Abdulmajid, Kasahun Desalegn, and Hayat Hassen. "Selected food items adulteration, their impacts on public health, and detection methods: A review." *Food Science & Nutrition* 11, no. 12 (2023): 7534-7545. https://doi.org/10.1002/fsn3.3732

[8] Quintela, Irwin A., Tyler Vasse, Chih-Sheng Lin, and Vivian CH Wu. "Advances, applications, and limitations of portable and rapid detection technologies for routinely encountered foodborne pathogens." *Frontiers in Microbiology* 13 (2022): 1054782. https://doi.org/10.3389/fmicb.2022.1054782

[9] Anagaw, Yeniewa Kerie, Wondim Ayenew, Liknaw Workie Limenh, Derso Teju Geremew, Minichil Chanie Worku, Tewodros Ayalew Tessema, Wudneh Simegn, and Melese Legesse Mitku. "Food adulteration: Causes, risks, and detection techniques." *SAGE Open Medicine* 12 (2024): 20503121241250184. https://doi.org/10.1177/20503121241250184

[10] Joshi, Rahul, Lakshmi Priya Gg, Mohammad Akbar Faqeerzada, Tanima Bhattacharya, Moon Sung Kim, Insuck Baek, and Byoung-Kwan Cho. "Deep learning-based quantitative assessment of melamine and cyanuric acid in pet food using fourier transform infrared spectroscopy." *Sensors* 23, no. 11 (2023): 5020. https://doi.org/10.3390/s23115020

[11] Sarkar, Chittabarni, Deepak Gupta, Umesh Gupta, and Barenya Bikash Hazarika. "Leaf disease detection using machine learning and deep learning: Review and challenges." *Applied Soft Computing* 145 (2023): 110534. https://doi.org/10.1016/j.asoc.2023.110534

[12] Jernelv, Ine L., Dag Roar Hjelme, Yuji Matsuura, and Astrid Aksnes. "Convolutional neural networks for classification and regression analysis of one-dimensional spectral data." *arXiv preprint arXiv:2005.07530* (2020). doi: 10.48550/arXiv.2005.07530.

[13] Zeng, Hao, Xiaoqing Han, and Qingjie Liu. "Mineral detection from hyperspectral images using a spatial-spectral residual convolution neural network." In *Journal of Physics: Conference Series*, vol. 1894, no. 1, p. 012104. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1894/1/012104

[14] Tang, Jia-Wei, Quan Yuan, Xin-Ru Wen, Muhammad Usman, Alfred Chin Yen Tay, and Liang Wang. "Label-free surface-enhanced Raman spectroscopy coupled with machine learning algorithms in pathogenic microbial identification: Current trends, challenges, and perspectives." *Interdisciplinary Medicine* 2, no. 3 (2024): e20230060. https://doi.org/10.1002/INMD.20230060

[15] Cozzolino, Daniel. "Advances in spectrometric techniques in food analysis and authentication." *Foods* 12, no. 3 (2023): 438. https://doi.org/10.3390/foods12030438

[16] Class, Lisa-Carina, Gesine Kuhnen, Sascha Rohn, and Jürgen Kuballa. "Diving deep into the data: a review of deep learning approaches and potential applications in foodomics." *Foods* 10, no. 8 (2021): 1803. https://doi.org/10.3390/foods10081803

[17] Tseng, Yufeng Jane, Pei-Jiun Chuang, and Michael Appell. "When machine learning and deep learning come to the big data in food chemistry." *ACS omega* 8, no. 18 (2023): 15854-15864. https://doi.org/10.1021/acsomega.2c07722

[18] E. Ali et al., "A Comprehensive Review of Sensor-Based and Spectroscopy-Based Systems for Monitoring Water Quality in Freshwater Aquaculture System," J. Adv. Res. Appl. Sci. Eng. Technol., vol. 56, no. 3, pp. 37–54, 2026. https://doi.org/10.37934/araset.56.3.3754

[19] Artavia, Graciela, Carolina Cortés-Herrera, and Fabio Granados-Chinchilla. "Selected instrumental techniques applied in food and feed: Quality, safety and adulteration analysis." *Foods* 10, no. 5 (2021): 1081. https://doi.org/10.3390/foods10051081

[20] He, Wanchong, and Qinghua Zeng. "Non-Invasive Food Authentication Using Vibrational Spectroscopy Techniques for Low-Resolution Food Fingerprinting." *Applied Sciences* 15, no. 11 (2025): 5948. https://doi.org/10.3390/app15115948

[21] Xie, Chuanqi, and Weidong Zhou. "A review of recent advances for the detection of biological, chemical, and physical hazards in foodstuffs using spectral imaging techniques." *Foods* 12, no. 11 (2023): 2266. https://doi.org/10.3390/foods12112266

[22] Qi, Wenliang, Yanlong Tian, Daoli Lu, and Bin Chen. "Research progress of applying infrared spectroscopy technology for detection of toxic and harmful substances in food." *Foods* 11, no. 7 (2022): 930. https://doi.org/10.3390/foods11070930

[23] Cherniienko, Alina, Roman Lesyk, Lucjusz Zaprutko, and Anna Pawełczyk. "IR-EcoSpectra: Exploring sustainable ex situ and in situ FTIR applications for green chemical and pharmaceutical analysis." *Journal of Pharmaceutical Analysis* 14, no. 9 (2024): 100951. https://doi.org/10.1016/j.jpha.2024.02.005

[24] Hussain, Mubashir, Jun Zou, He Zhang, Ru Zhang, Zhu Chen, and Yongjun Tang. "Recent Progress in Spectroscopic Methods for the Detection of Foodborne Pathogenic Bacteria." *Biosensors* 12, no. 10 (2022): 869, https://doi.org/10.3390/bios12100869

[25] Maruthamuthu, Murali K., Amir Hossein Raffiee, Denilson Mendes De Oliveira, Arezoo M. Ardekani, and Mohit S. Verma. "Raman spectra-based deep learning: A tool to identify microbial contamination." *MicrobiologyOpen* 9, no. 11 (2020): e1122. https://doi.org/10.1002/mbo3.1122

[26] Ekramirad, Nader, Alfadhl Y. Khaled, Lauren E. Doyle, Julia R. Loeb, Kevin D. Donohue, Raul T. Villanueva, and Akinbode A. Adedeji. "Nondestructive detection of codling moth infestation in apples using pixel-based nir hyperspectral imaging with machine learning and feature selection." *Foods* 11, no. 1 (2021): 8. https://doi.org/10.3390/foods11010008

[27] Soni, Aswathi, Yash Dixit, Marlon M. Reis, and Gale Brightwell. "Hyperspectral imaging and machine learning in food microbiology: Developments and challenges in detection of bacterial, fungal, and viral contaminants." *Comprehensive Reviews in Food Science and Food Safety* 21, no. 4 (2022): 3717-3745. https://doi.org/10.1111/1541-4337.12983

[28] Windarsih, Anjar, Abdul Rohman, Irnawati, and Sugeng Riyanto. "The combination of vibrational spectroscopy and chemometrics for analysis of milk products adulteration." *International Journal of Food Science* 2021, no. 1 (2021): 8853358. https://doi.org/10.1155/2021/8853358

[29] Kharbach, Mourad, Mohammed Alaoui Mansouri, Mohammed Taabouz, and Huiwen Yu. "Current application of advancing spectroscopy techniques in food analysis: data handling with chemometric approaches." *Foods* 12, no. 14 (2023): 2753. https://doi.org/10.3390/foods12142753

[30] Zhang, Huanle, Nicharee Wisuthiphaet, Hemiao Cui, Nitin Nitin, Xin Liu, and Qing Zhao. "Spectroscopy approaches for food safety applications: Improving data efficiency using active learning and semi-supervised learning." *Frontiers in Artificial Intelligence* 5 (2022): 863261. https://doi.org/10.3389/frai.2022.863261

[31] Houston, James, Frank G. Glavin, and Michael G. Madden. "Robust classification of high-dimensional spectroscopy data using deep learning and data synthesis." *Journal of Chemical Information and Modeling* 60, no. 4 (2020): 1936-1954. 2020. https://doi.org/10.1021/acs.jcim.9b01037

[32] Castorena, Juan, Diane Oyen, Ann Ollila, Carey Legett, and Nina Lanza. "Deep spectral CNN for laser induced breakdown spectroscopy." *Spectrochimica Acta Part B: Atomic Spectroscopy* 178 (2021): 106125. https://doi.org/10.1016/j.sab.2021.106125

[33] Tsakanikas, Panagiotis, Apostolos Karnavas, Efstathios Z. Panagou, and George-John Nychas. "A machine learning workflow for raw food spectroscopic classification in a future industry." *Scientific Reports* 10, no. 1 (2020): 11212. https://doi.org/10.1038/s41598-020-68156-2

[34] Martens, Harald. "Understanding the root cause (s) of nonlinearities in near infrared spectroscopy." *NIR news* 32, no. 1-2 (2021): 20-26. https://doi.org/10.1177/09603360211003758

[35] Chen, Tingting, Chun Wang, Jingqiu Liang, and Jingsong Li. "Application of Derivative Transform Spectroscopy in Gas Detection." In *E3S Web of Conferences*, vol. 299, p. 02005. EDP Sciences, 2021. https://doi.org/10.1051/e3sconf/202129902005

[36] Inglis, Alan, Andrew C. Parnell, Natarajan Subramani, and Fiona M. Doohan. "Machine learning applied to the detection of mycotoxin in food: A systematic review." *Toxins* 16, no. 6 (2024): 268. https://doi.org/10.3390/toxins16060268

[37] Mumuni, Alhassan, and Fuseini Mumuni. "Automated data processing and feature engineering for deep learning and big data applications: a survey." *Journal of Information and Intelligence* 3, no. 2 (2025): 113-153. https://doi.org/10.1016/j.jiixd.2024.01.002

[38] Nturambirwe, Jean Frederic Isingizwe, Eslam A. Hussein, Mattia Vaccari, Christopher Thron, Willem Jacobus Perold, and Umezuruike Linus Opara. "Feature reduction for the classification of bruise damage to apple fruit using a contactless FT-NIR spectroscopy with machine learning." *Foods* 12, no. 1 (2023): 210. https://doi.org/10.3390/foods12010210

[39] Liu, Xuyang, Hongle An, Wensheng Cai, and Xueguang Shao. "Deep learning in spectral analysis: Modeling and imaging." *TrAC Trends in Analytical Chemistry* 172 (2024): 117612. https://doi.org/10.1016/j.trac.2024.117612

[40] Jin, Zhuo-Lin, Lu Chen, Yu Wang, Chao-Ting Shi, Yan Zhou, and Bing Xia. "Application of machine learning in LC-MS-based non-targeted analysis." *TrAC Trends in Analytical Chemistry* (2025): 118243. https://doi.org/10.1016/j.trac.2025.118243

[41] Roman, Ibai, Roberto Santana, Alexander Mendiburu, and Jose A. Lozano. "In-depth analysis of SVM kernel learning and its components." *Neural Computing and Applications* 33, no. 12 (2021): 6575-6594. doi: 10.1007/s00521-020-05419-z. https://doi.org/10.1007/s00521-020-05419-z

[42] Brandt, John, and Fred Stolle. "A global method to identify trees outside of closed-canopy forests with medium-resolution satellite imagery." *International Journal of Remote Sensing* 42, no. 5 (2021): 1713-1737. https://doi.org/10.1080/01431161.2020.1841324

[43] Ding, Yuhan, Yuli Yan, Jun Li, Xu Chen, and Hui Jiang. "Classification of tea quality levels using near-infrared spectroscopy based on CLPSO-SVM." *Foods* 11, no. 11 (2022): 1658. https://doi.org/10.3390/foods11111658

[44] Silalahi, Divo Dharma, Habshah Midi, Jayanthi Arasan, Mohd Shafie Mustafa, and Jean-Pierre Caliman. "Kernel partial least square regression with high resistance to multiple outliers and bad leverage points on near-infrared spectral data analysis." *Symmetry* 13, no. 4 (2021): 547. https://doi.org/10.3390/sym13040547

[45] Sadergaski, Luke R., Sawyer B. Irvine, and Hunter B. Andrews. "Partial least squares, experimental design, and near-infrared spectrophotometry for the remote quantification of nitric acid concentration and temperature." *Molecules* 28, no. 7 (2023): 3224. https://doi.org/10.3390/molecules28073224

[46] Plevris, Vagelis, German Solorzano, Nikolaos P. Bakas, and Mohamed El Amine Ben Seghier. "Investigation of performance metrics in regression analysis and machine learning-based prediction models." (2022). https://doi.org/10.23967/eccomas.2022.155

[47] Wisuthiphaet, Nicharee, Huanle Zhang, Xin Liu, and Nitin Nitin. "Detection of Escherichia coli Using Bacteriophage T7 and Analysis of Excitation-Emission Matrix Fluorescence Spectroscopy." *Journal of Food Protection* 87, no. 12 (2024): 100396. https://doi.org/10.1016/j.jfp.2024.100396

[48] Sun, Jiazheng, Di He, and Yuxin You. "Raman spectroscopy powered by machine learning methods for rapid identification of foodborne pathogens." *Food Bioscience* 66 (2025): 106281. https://doi.org/10.1016/j.fbio.2025.106281

[49] Tu, Qiancheng, Miaoyun Li, Zhiyuan Sun, Huimin Niu, Lijun Zhao, Yanxiao Wang, Lingxia Sun, Yanxia Liu, Yaodi Zhu, and Gaiming Zhao. "Rapid and accurate identification of foodborne bacteria: a combined approach using confocal Raman micro-spectroscopy and explainable machine learning." *Analytical and Bioanalytical Chemistry* (2025): 1-12https://doi.org/10.1007/s00216-025-05816-0

[50] Chen, Zhuzi, Zamar Daka, Liying Yao, Jiamin Dong, Yuqi Zhang, Peiqi Li, Kaidi Zhang, and Shunli Ji. "Recent progress in the application of chromatography-coupled mass-spectrometry in the analysis of contaminants in food products." *Food Chemistry: X* (2025): 102397. https://doi.org/10.1016/j.fochx.2025.102397

[51] Hu, Yating, Benxue Ma, Huting Wang, Yuanjia Zhang, Yujie Li, and Guowei Yu. "Detecting different pesticide residues on Hami melon surface using hyperspectral imaging combined with 1D-CNN and information fusion." *Frontiers in Plant Science* 14 (2023): 1105601. https://doi.org/10.3389/fpls.2023.1105601

[52] Mohammadi, Mahsa, Mohammadreza Khanmohammadi Khorrami, Hamid Vatanparast, Amirmohammad Karimi, and Mina Sadrara. "Classification and determination of sulfur content in crude oil samples by infrared spectrometry." *Infrared Physics & Technology* 127 (2022): 104382. https://doi.org/10.1016/j.infrared.2022.104382

[53] Nana Hermana, Asep, Dewi Rosmala, and Milda Gustiana Husada. "Transfer learning for classification of fruit ripeness using vgg16." In *Proceedings of the 2021 4th International Conference on Computers in Management and Business*, pp. 139-146. 2021. https://doi.org/10.1145/3450588.3450943

[54] Cho, B. K. "Application of spectral imaging for safety inspection of fresh cut vegetables." In *IOP Conference Series: Earth and Environmental Science*, vol. 686, no. 1, p. 012001. IOP Publishing, 2021. https://doi.org/10.1088/1755-1315/686/1/012001

[55] Shu, Zhan, Xiong Li, and Yande Liu. "Detection of Chili Foreign Objects Using Hyperspectral Imaging Combined with Chemometric and Target Detection Algorithms." *Foods* 12, no. 13 (2023): 2618. https://doi.org/10.3390/foods12132618

[56] Campos, Rodrigo Louzada, Seung-Chul Yoon, Soo Chung, and Suchendra M. Bhandarkar. "Semisupervised deep learning for the detection of foreign materials on poultry meat with near-infrared hyperspectral imaging." *Sensors* 23, no. 16 (2023): 7014. https://doi.org/10.3390/s23167014

[57] Wang, Jiahui, Mengdie Jiang, Tauseef Abbas, Hao Chen, and Yuying Jiang. "YOLOv-MA: A High-Precision Foreign Object Detection Algorithm for Rice." *Agriculture* 15, no. 13 (2025): 1354. https://doi.org/10.3390/agriculture15131354

[58] Yu, Yuewen, Wei Chen, Dongjie Zhao, Hanwen Zhang, Wenliang Chen, Rong Liu, and Chenxi Li. "Meat species authentication using portable hyperspectral imaging." *Frontiers in Nutrition* 12 (2025): 1577642. https://doi.org/10.3389/fnut.2025.1577642

[59] Sun, Qiannan, Yide Dong, Xin Wen, Xu Zhang, Shijiao Hou, Wuduo Zhao, and Dan Yin. "A review on recent advances in mass spectrometry analysis of harmful contaminants in food." *Frontiers in Nutrition* 10 (2023): 1244459. https://doi.org/10.3389/fnut.2023.1244459

[60] Bai, Zongxiu, Dongdong Du, Rongguang Zhu, Fukang Xing, Chenyi Yang, Jiufu Yan, Yixin Zhang, and Lichao Kang. "Establishment and comparison of in situ detection models for foodborne pathogen contamination on mutton based on SWIR-HSI." *Frontiers in Nutrition* 11 (2024): 1325934. https://doi.org/10.3389/fnut.2024.1325934

[61] Banerjee, Shantanu, Shyamapada Mandal, Naveen G. Jesubalan, Rijul Jain, and Anurag S. Rathore. "NIR spectroscopy—CNN-enabled chemometrics for multianalyte monitoring in microbial fermentation." *Biotechnology and Bioengineering* 121, no. 6 (2024): 1803-1819. https://doi.org/10.1002/bit.28681

[62] Ling, Zhiming, Luyu Yang, Wanqing Zhang, Ting Yao, and Hengyi Xu. "Detection of food contaminants: A review of established rapid analytical techniques and their applications and limitations." *Food Safety and Health* 2, no. 1 (2024): 72-95. https://doi.org/10.1002/fsh3.12032

[63] Jia, Wenyang, Konstantia Georgouli, Jesus Martinez-Del Rincon, and Anastasios Koidis. "Challenges in the use of AI-driven non-destructive spectroscopic tools for rapid food analysis." *Foods* 13, no. 6 (2024): 846. https://doi.org/10.3390/foods13060846

[64] Zhong, Xiaoting, Brian Gallagher, Keenan Eves, Emily Robertson, T. Nathan Mundhenk, and T. Yong-Jin Han. "A study of real-world micrograph data quality and machine learning model robustness." *npj Computational Materials* 7, no. 1 (2021): 161. https://doi.org/10.1038/s41524-021-00616-3

[65] Amirvaresi, Arian, and Hadi Parastar. "Miniaturized NIR spectroscopy and chemometrics: A smart combination to solve food authentication challenges." *Frontiers in Analytical Science* 3 (2023): 1118590. https://doi.org/10.3389/frans.2023.1118590

[66] Xie, Zeke, Fengxiang He, Shaopeng Fu, Issei Sato, Dacheng Tao, and Masashi Sugiyama. "Artificial neural variability for deep learning: On overfitting, noise memorization, and catastrophic forgetting." *Neural computation* 33, no. 8 (2021): 2163-2192. https://doi.org/10.1162/neco_a_01403

[67] Zhang, Chiyuan, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. "Understanding deep learning (still) requires rethinking generalization." *Communications of the ACM* 64, no. 3 (2021): 107-115. https://doi.org/10.1145/3446776

[68] Sitorus, Agustami, Muhamad Muslih, Irwin Syahri Cebro, and Ramayanty Bulan. "Dataset of adulteration with water in coconut milk using FTIR spectroscopy." *Data in Brief* 36 (2021): 107058. https://doi.org/10.1016/j.dib.2021.107058

[69] Cebi, Nur, Hatice Bekiroglu, Azime Erarslan, and Luis Rodriguez-Saona. "Rapid sensing: Hand-held and portable FTIR applications for on-site food quality control from farm to fork." *Molecules* 28, no. 9 (2023): 3727. https://doi.org/10.3390/molecules28093727

[70] Tabani, Hamid, Ajay Balasubramaniam, Elahe Arani, and Bahram Zonooz. "Challenges and obstacles towards deploying deep learning models on mobile devices." *arXiv preprint arXiv:2105.02613* (2021). doi: 10.48550/arXiv.2105.02613.

[71] Ricci, Cosimo, Agata Gadaleta, Annamaria Gerardino, Angelo Didonna, Giuseppe Ferrara, and Francesca Romana Bertani. "Portable optical spectroscopy and machine learning techniques for quantification of the biochemical content of raw food materials." *CABI Agriculture and Bioscience* 5, no. 1 (2024): 39. https://doi.org/10.1186/s43170-024-00244-z

[72] Bergmann, Paul, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. "The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection." *International Journal of Computer Vision* 129, no. 4 (2021): 1038-1059. https://doi.org/10.1007/s11263-020-01400-4

[73] Surkova, Anastasiia, Andrey Bogomolov, Andrey Legin, and Dmitry Kirsanov. "Calibration transfer for LED-based optical multisensor systems." *ACS sensors* 5, no. 8 (2020): 2587-2595. https://doi.org/10.1021/acssensors.0c01018

[74] Guo, Cheng, Jin Zhang, Wensheng Cai, and Xueguang Shao. "Enhancing transferability of near-infrared spectral models for soluble solids content prediction across different fruits." *Applied Sciences* 13, no. 9 (2023): 5417. https://doi.org/10.3390/app13095417

[75] Konstantinova, Tatiana, Phillip M. Maffettone, Bruce Ravel, Stuart I. Campbell, Andi M. Barbour, and Daniel Olds. "Machine learning enabling high-throughput and remote operations at large-scale user facilities." *Digital Discovery* 1, no. 4 (2022): 413-426.Aug. 2022. https://doi.org/10.1039/D2DD00014H

[76] Capogrosso, Luigi, Federico Cunico, Dong Seon Cheng, Franco Fummi, and Marco Cristani. "A machine learning-oriented survey on tiny machine learning." *IEEE Access* 12 (2024): 23406-23426. https://doi.org/10.1109/ACCESS.2024.3365349

[77] Lin, Ji, Ligeng Zhu, Wei-Ming Chen, Wei-Chen Wang, and Song Han. "Tiny machine learning: Progress and futures [feature]." *IEEE Circuits and Systems Magazine* 23, no. 3 (2023): 8-34. https://doi.org/10.1109/MCAS.2023.3302182

[78] Dritsas, Elias, and Maria Trigka. "Federated Learning for IoT: A Survey of Techniques, Challenges, and Applications." *Journal of Sensor and Actuator Networks* 14, no. 1 (2025): 9. https://doi.org/10.3390/jsan14010009

[79] Oveis, Amir Hosein, Elisa Giusti, Giulio Meucci, Selenia Ghio, and Marco Martorella. "Explainability in hyperspectral image classification: a study of XAI through the Shap algorithm." In *2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1-5. IEEE, 2023.https://ieeexplore.ieee.org/abstract/document/10430776/https://doi.org/10.1109/WHISPERS61460.2023.10430776

[80] Zhang, Guyang, and Waleed Abdulla. "Explainable AI-driven wavelength selection for hyperspectral imaging of honey products." *Food Chemistry Advances* 3 (2023): 100491. doi: 10.1016/j.focha.2023.100491.

[