ANNUAL REVIEWS

Annual Review of Biomedical Data Science

Data Science in the Food Industry

George-John Nychas,[1] Emma Sims,[2] Panagiotis Tsakanikas,[1] and Fady Mohareb[2]

[1]Laboratory of Microbiology and Biotechnology of Foods, Department of Food Science and Human Nutrition, School of Food and Nutritional Sciences, Agricultural University of Athens, 11855 Athens, Greece; email: gjn@aua.gr

[2]Bioinformatics Group, Department of Agrifood, School of Water, Energy and Environment, Cranfield University, Cranfield, Bedfordshire MK43 0AL, United Kingdom

Annu. Rev. Biomed. Data Sci. 2021. 4:341–67

First published as a Review in Advance on May 13, 2021

The *Annual Review of Biomedical Data Science* is online at biodatasci.annualreviews.org

https://doi.org/10.1146/annurev-biodatasci-020221-123602

Keywords

big data, food safety, omics, machine learning, food microbiology

Abstract

Food safety is one of the main challenges of the agri-food industry that is expected to be addressed in the current environment of tremendous technological progress, where consumers' lifestyles and preferences are in a constant state of flux. Food chain transparency and trust are drivers for food integrity control and for improvements in efficiency and economic growth. Similarly, the circular economy has great potential to reduce wastage and improve the efficiency of operations in multi-stakeholder ecosystems. Throughout the food chain cycle, all food commodities are exposed to multiple hazards, resulting in a high likelihood of contamination. Such biological or chemical hazards may be naturally present at any stage of food production, whether accidentally introduced or fraudulently imposed, risking consumers' health and their faith in the food industry. Nowadays, a massive amount of data is generated, not only from the next generation of food safety monitoring systems and along the entire food chain (primary production included) but also from the Internet of things, media, and other devices. These data should be used for the benefit of society, and the scientific field of data science should be a vital player in helping to make this possible.

1. INTRODUCTION

Consumer demands for high-quality foods that taste fresh, have crisp textures, and are nutritious are proportionately related to social behaviors. Consumer mindsets toward current food standards are also driving factors in food loss and waste, along with poor communication and coordination among different actors of the food supply chain. In parallel, consumers are increasingly aware of foodborne disease hazards and, thus, are also concerned about the safety and security of their food supply.

Indeed, at the consumer level, there is insufficient purchase planning, and in combination with the careless attitude of consumers who can afford to waste food, expiring best-before dates cause large amounts of waste (1). Moreover, the food industry's monitoring of the quality and safety of these highly perishable foods still relies heavily on regulatory inspection where analyses are performed via conventional means (e.g., International Organization for Standardization for total viable count) (2). This process is costly, time consuming, destructive to foods, and retrospective, and it precludes providing real-time information regarding remaining shelf life and a safety inspection throughout a product's life cycle. Food safety, along with security, represents the main challenge facing the agri-food industry in the twenty-first century. With the agri-food industry's estimated net worth of more than $8 trillion, along with the advancement of smart farming and changing attitudes toward popularized food trends, maintaining the safety standards of food has never been more important.

The worldwide population is predicted to reach 10 billion by 2050, and the demand for higher yields of nutritious food is expected to increase proportionally. It is envisaged that novel food chain technologies will increase productivity of the food chain and the quality of foods and could help address several societal challenges such as an ageing global population, the effects of climate change, and the reduced availability of resources. Thus, the agritech industries need to upscale applied control methods for safety by using accurate, rapid, and noninvasive technologies to detect unsafe and inauthentic food samples within the food chain itself.

Recently the European Union has required a mechanism called process analytical technology (PAT) to be deployed in the food manufacturing processing and production lines, which offers a solution to a broad need identified by agri-food industries, i.e., safety control of raw and in-process materials due to processing contaminants or existing or emerging hazards.

The quantification and correlation of changes associated with agri-food products can be used for the development of a distributed data repository at the heart of this approach that will integrate a massive amount of heterogeneous data (e.g., microbiological, metabolomics, and multi/hyperspectral fingerprints) taken from a wide range of handling, storage, and distribution conditions. This integrated multilevel and multidisciplinary knowledge will be used to build robust models of product quality and safety, as only through the correlation of multiple parameters of food quality can a robust and reliable monitoring service be developed and profitably applied. This knowledge can be tracked throughout the production, supply, and distribution chain and can be made available to stakeholders, food operators, retailers, and consumers through a live tracking system combining mobile and web technologies [e.g., by introducing specially designed QR (quick response) codes to food product packaging]. Examples of this approach are illustrated in **Figure 1**. The incorporation of knowledge from multiple disciplines, as depicted, would drive the agritech industry to rely on data science, a multidisciplinary field within computer science that combines data, mathematics, statistics, algorithms, and computing, to model various aspects of the supply chain in order to identify and improve current industry practices as inefficient.

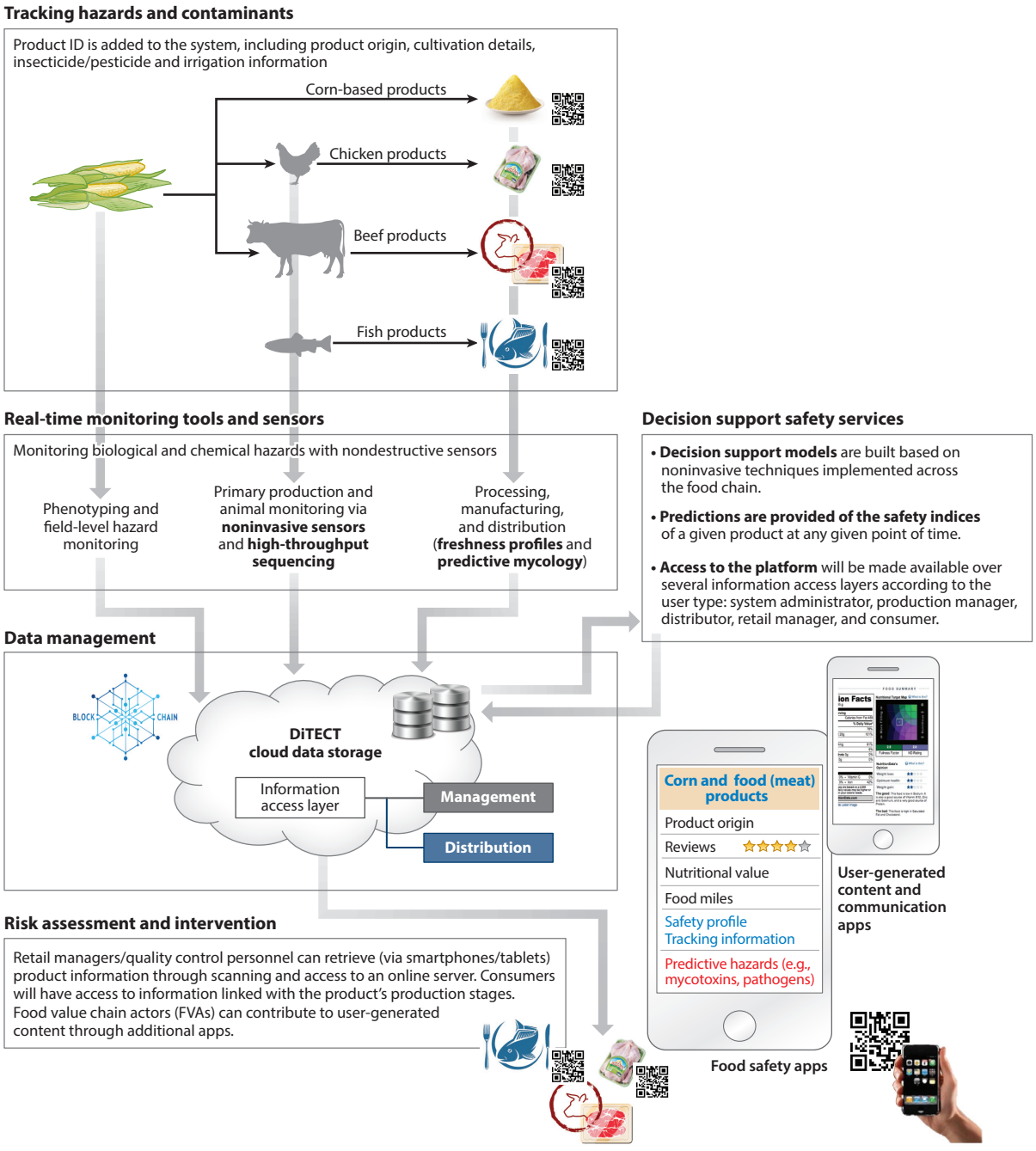

Figure 1

Overview of a holistic system, incorporating the knowledge acquired throughout the farm-to-fork food chain.

2. MEASURING (MONITORING) FOOD SAFETY AND QUALITY

2.1. Food Safety Versus Food Quality

The safety of food products, which is delineated by a series of challenges associated with either microbial pathogens or chemical contaminants, has been a major societal concern. Various explanations have been identified for the rising food safety concerns of recent years, including changes in food production, product processing, and distribution; increased international trade; increased worldwide food consumption; changing consumer needs and consumption patterns (e.g., preference for minimally processed foods); higher numbers of consumers at risk for infection; and increased interest, awareness, and scrutiny by consumers.

Despite the seriousness of the problem, there is no consistent, or accepted, definition of "food safety," as this term has different meanings among the public and food safety professionals. Food safety is defined roughly as the condition that ensures that food will not cause any harm to the consumer when it is prepared or eaten according to its intended use. Food safety deals with all those hazards, whether chronic or acute, that may make food injurious to the health of consumers [according to the FAO (Food and Agriculture Organization)/WHO (World Health Organization)], and as such, it is not negotiable. In contrast, food quality includes all other attributes that influence a product's value to the consumer (e.g., spoilage, flavor, texture, contamination, adulteration, authentication). This distinction between safety and quality has implications for public policy and influences the nature and content of the food control system most suited to meet predetermined national objectives.

Adulteration is another challenge that often either is associated with or directly impacts the safety of consumed goods. One motivation for food producers and processors to illegally alter food products and materials is to increase the quantity of final product with cheap substitutes, keeping net costs low and net profits high. The temptation to increase stocks to meet consumer demands has never been higher than throughout the initial COVID-19 (coronavirus disease 2019) lockdowns, which saw consumers stockpiling goods such as dried pastas and frozen meats, causing an unprecedented strain on procurement industries that operate on a just-in-time business strategy. These types of fraudulent scandals can rapidly escalate consumer distrust in both food production and marketing, as seen in the horsemeat incident in 2012, where both horse and pig DNA were found in a wide range of low-cost beef burgers (3). The magnitude of this scandal grew from the initial discovery in Ireland across the European Union, highlighting the vulnerability of the food processing and import chains at an international level. At the time, KPMG International Ltd. estimated that there were approximately 450 points at which the integrity of the supply chain could break down, leading to difficulties in tracing the initial incidence of fraudulent labeling that led to the adulteration of the final product.

In 2014 the Elliott review was published in the United Kingdom, which set out recommendations to strengthen the integrity of the food industry; this included the setup of the National Food Crime Unit, a branch of the British Food Standards Agency. The recommendations set out in this report are summarized best using eight key principles: consumers first, zero tolerance, intelligence gathering, laboratory services, auditing, government support, leadership, and crisis management (4). Since the Elliott review, there has been an international effort to minimize food crime incidence by implementing the eight ideals, but as repeatedly stated, the most effective methods of tackling food crime lie in government regulators co-operating with businesses to gather intelligence around methods of adulterating food covertly, as well as conducting rigorous laboratory testing regimes on an annual basis. As shown in the horsemeat scandal, some industries purchase adulterated materials purportedly unaware of their fraudulent contents; this still holds true today, since as early as last year there have been reports of refined olive oil being sold as extra virgin (5).

These incidents show that although there has been progress, vulnerabilities in the food supply chain still exist, and with increasing pressures to keep food both cheap and in stock, the risk of adulteration is high.

2.2. Guaranteeing, Synchronizing, and Sharing Responsibilities for Food Safety Issues

At the World Food Summit held in Rome on November 13–17, 1996, the FAO reaffirmed "the right of everyone to have access to safe and nutritious food, consistent with the right to adequate food and the fundamental right of everyone to be free from hunger" (*6*). Guaranteeing this right is an important priority among the governments of United Nations member states. Among the member states of the European Union, more than 200 central EU laws have been incorporated into national legislation in the last 20 years. Similar efforts have been made by the Chinese Food and Drug Administration (CFDA) to harmonize/impose food safety laws in all provinces (*7*, *8*). In contrast, in the United States, the Food and Drug Administration (FDA) is implementing the recently passed Food Safety Modernization Act (FSMA) (21 U.S.C. § 301 et seq.), which recognizes that ensuring the safety of the food supply is a shared responsibility among many different points in the global supply chain for both human and animal food. In China, food producers and traders are responsible for the safety, while in the European Union, food safety along the food chain is a shared responsibility among food business operators, which have the primary responsibility; regulatory authorities, which monitor this responsibility; and consumers, who must also recognize that they are responsible for the proper storage, handling, and preparation of food (*9*).

2.3. Food Safety Control/Assessment: Current and Future Approaches

In all of the regions mentioned above, governments have attempted to fulfil the need for food safety by utilizing a common approach/practice that relies heavily on establishing that there are frequent standardized regulatory inspections and that sampling regimes are followed. Currently, a wide range of audits and inspections are used to evaluate the quality or safety of raw or processed materials and food products (*10*). This is largely based on good design of processes, products, and procedures, where testing of the end or finished product (analyzed for certain hazards) is considered to be the control measure of the production process.

Both microbiological and physicochemical analyses used to monitor food safety (*a*) are time consuming, providing retrospective results; (*b*) are costly; (*c*) sometimes require high-tech molecular tools and, thus, highly trained personnel; and (*d*) are usually destructive to test products, thus limiting their potential to be used on, in, or at line. It is clear that this approach cannot sufficiently guarantee consumer protection since inspection and sampling of 100% of food products are technically, financially and logistically impossible (*10*).

Bearing in mind the limitations mentioned above, novel approaches to food safety and control are being developed based on the latest advancements in data management technologies in order to integrate multilevel and trans-disciplinary data in a structured, semi-structured, or unstructured database (big data). Such knowledge is derived from (*a*) preharvest and primary production environmental conditions and (*b*) processing, handling, storage, and distribution conditions, using automated systems to record monitoring data in food production facilities (e.g., silos, animal facilities) and their final products in the entire food chain (**Figure 2**).

2.4. Monitoring Food Safety/Quality: Tools

It is only in recent years that omics technologies have been used as a very promising approach for the assessment of food safety, quality, and (food) crime. DNA sequences and DNA barcoding are

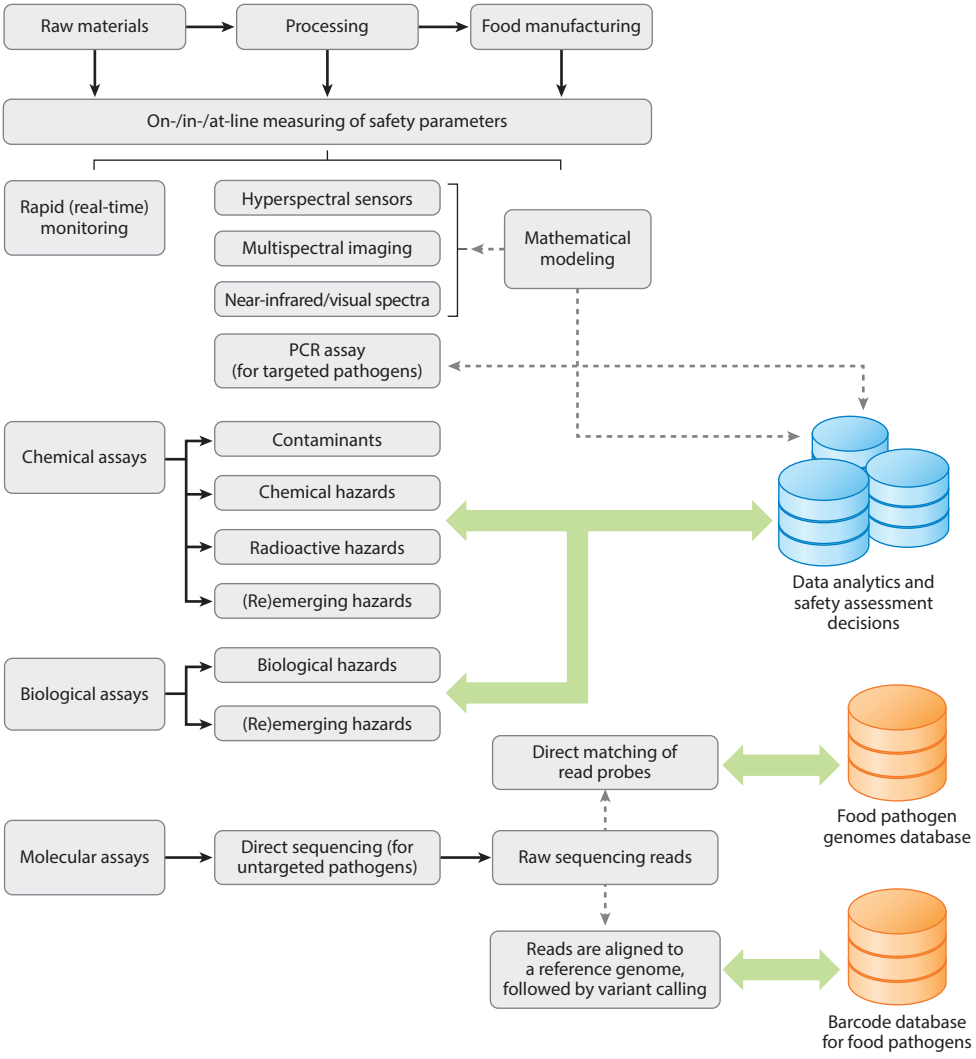

Figure 2

Flow chart of information and data along the food chain, which constitutes a traceable system for the assessment and prediction of safety risks. Solid black arrows indicate workflow processes. Dotted arrows indicate that a preprocessing step is required before the storage/query interaction with the database. Large green arrows indicate query interactions with a database rather than the actual physical storage of data. Abbreviation: PCR, polymerase chain reaction.

considered potential monitoring tools for food fraud (e.g., edible oils, halal meat, horsemeat) (11). Moreover, analytical platforms have been very useful since they provide vital information regarding the composition of foods. Proteomics and other recent omics disciplines (e.g., metabolomics, lipidomics, volatilomics, and fingerprinting), as well as artificial intelligence (AI), have proved to be excellent tools in the food sector (12, 13). Additionally, nondestructive and noninvasive sensors, as well as advanced high-throughput sequencing techniques, which are considered either nontargeted or targeted rapid detection methods, have recently been introduced to monitor food safety.

2.4.1. Sensors in food science. Food can be contaminated at any stage of the farm-to-fork food chain via either chemical, biological, or physical contaminants. Traditional in-lab safety assessments are neither practical nor applicable during the food processing and distribution chain. A wide range of well-established techniques using sensors have been employed for food safety hazard detection, such as high-performance liquid chromatography (14–16), gas chromatography (17, 18), high-performance liquid chromatography–mass spectrometry (19), gas chromatography–mass spectrometry (GC-MS) (20), nuclear magnetic resonance (21, 22), Fourier transform infrared spectroscopy (FT-IR) (23–25), and Raman spectrometry (26–29). Although these sensors have proved to be efficient and can provide trustworthy results, they must be used within strictly controlled laboratory conditions. Because of this, biological and chemical hazard sensors that are less targeted to specific markers of environmental contaminants must also be employed and applied in the field. In addition, sensors should be easily transported or installed at several stages of the food chain, which is a more rational choice for food safety and quality monitoring. To this end, alternative sensor spectroscopy approaches have also been used, including sensors with spectra in the ultraviolet (UV), near-infrared (NIR), mid-infrared, and visual (VIS) ranges. More recently multi- and hyperspectral imaging sensors have been employed as tools to simultaneously assess spectral and spatial information from food samples (30–35). What is encouraging about both vibrational spectroscopy and imaging sensors is that they have proved efficient enough, when coupled with appropriate data analysis methods, against the gold standard performance of sensors like FT-IR and GC-MS.

2.4.2. Application of next-generation sequencing to food safety. Advances in next-generation sequencing (NGS) technologies over the past decade have given researchers an unprecedented opportunity to enhance understanding of microbial behavior on a molecular level across all omics levels. Decreasing sequencing costs as well as the abundance of platforms have opened up the possibility of deep genome assemblies for orphan species outside of the model organisms, and sometimes at the strain level for key pathogens. This includes, for instance, reference genomes for several *Campylobacter* strains isolated from retail chickens (36), beef livers (37) and turkeys (38), as well as high-resolution functionally annotated reference genomes for key pathogens such as *Campylobacter jejuni* (39) and *Listeria monocytogenes* (40, 41). The completeness of reference genomes is nowadays an easily achievable target thanks to the availability of third-generation sequencing platforms from Oxford Nanopore Technologies and Pacific Biosciences, which can produce very long sequencing reads from a single high–molecular weight DNA molecule. The high error rate of such platforms can easily be compensated for by using a hybrid assembly strategy and short-read Illumina sequences for error correction, as was the case for the salmonella genome assembly, where four *Salmonella enterica* strains isolated from pistachio genomes were assembled to completion (42).

Additionally, metagenomics provides an excellent framework for studying the microbial ecosystem through monitoring the relationships and interactions among different species to identify the impact of some naturally occurring or spoiling species on the presence, growth suppression, or activation of pathogens (43, 44). Furthermore, the interactions between the food matrix and environmental conditions during food storage and their impact on the microbial ecosystem can be monitored across several products such as cheese (45, 46), vegetables (47, 48), meat (49, 50), and poultry (51).

2.4.3. Data mining and data analysis. The massive amount of data generated by various analytical and high-throughput platforms is a challenging issue for food safety. Monitoring and

real-time sensor data should be stored in a central repository with a semi-structured design to provide a flexible means of storing these highly heterogeneous and diverse data types. This data repository can be complemented with a flexible, smartphone- and cloud-enabled web interface.

As mentioned above, several signal processing and machine learning (ML) methods coupled with noninvasive sensors [e.g., imaging; VIS, NIR, and UV sensors; electronic nose (eNose)] have been developed and applied in the food sector (for a large variety of food products) for safety and assessment purposes, among others. **Table 1** and **Supplemental File 1** present a list of published manuscripts on this topic, along with the corresponding methods used. In general, the data analysis workflow follows the pipeline of data acquisition, preprocessing, normalization and data handling, feature extraction/selection or feature engineering, and classification or regression, which, according to the problem at hand, may be either supervised or unsupervised. More recently, following the success of AI in other scientific areas, deep neural networks and convolutional neural networks (CNNs) seem to provide superior results in machine vision over more traditional ML approaches like support vector machines, genetic algorithms, and partial least-squares discriminant analysis, among others (see **Table 1** and **Supplemental File 1**). CNNs have been used for identifying and classifying food types. While this technology is still in a primitive stage, it has thus far provided promising results with food type identification using image databases such as Food-101 and UECFood (52). One setback to the next step of utilizing deep learning image-based ML algorithms is the lack of reference images in the libraries for all food types in varying states of quality, adulteration, and contamination. Briefly, CNNs mimic the human visual system to gain information from visual media. The main features of an image that help identify its subject are color and shape, which can be extracted using mathematical methods such as edge detection and image segmentation, both of which can be automated using ML methods to gather data relating to regions of interest within an image. Edge detection uses convolutional kernels to distort an image by enhancing certain features within it such as pixel intensity gradient changes in particular directions. These features are used to define edges, and different kernels can extract different kinds of image data, such as corners, straight lines, textures, and blank spaces. The expression of these characteristics builds up a profile of an image subject, which a neural network can use to predict the class/type of the subject (53).

One major advantage of deep learning approaches is that feature engineering and feature selection are not as crucial as in other ML methods since the model learns, during the training phase, the most informative information for the problem at hand. Nevertheless, there are currently two difficulties for the application of AI in the food safety sector. The first is the stakeholders' disbelief in the system's output since a lot of the underlying algorithms work as black boxes and their decisions are not interpretable. The second issue is the limited amount of data, which is one of the biggest issues in data science in general, including the food sector. Food, just like any living organism, exhibits vast variability among samples even within the same batch of products, and this issue also arises if one considers the environmental impact, contamination impact, storage and processing conditions, etc. Thus, it is apparent that massive amounts of data are required for the development of robust and efficient models. Through advancements in sensor technology, there are increasingly many publications employing data fusion at different levels: data, features, and prediction/decision fusion (99). Data fusion itself is the practice of merging sensory inputs to generate new and more informative datasets. This can be done over seven levels according to the JDL/DFIG (Joint Directors of Laboratories/Data Fusion Information Group) model: source preprocessing (of raw data), object assessment, situation assessment, impact assessment, process refinement, user refinement, and mission refinement (100).

Currently, there are apparent limitations for using preexisting multiplatform information-fusion methods, including the lack of storage space, as well as the suboptimal processing speeds.

Table 1 Representative rapid food sensor applications and corresponding data analysis methodologies

Sensor types	Food types	Purpose	Data analysis methods	Reference
Imaging	Beef fillets	Spoilage detection	HCA, PLSR, PLS-DA	54
	Salmon	Spoilage (LAB) detection	LS-SVM	55
	Mushrooms	Bruise detection	PCA (application to hypercube data)	56
	Meat	Monitoring meat color	PCA versus linear, nonlinear, and kernel-based regression methods (ANN and SVM)	57
	Milk powders	Adulteration detection	Spectral similarity measures (SAM, SCM, EDM)	58
	Pork, beef	Adulteration detection	HCA, PCA, LDA, PLS-DA	59
	Prawn	Adulteration detection	UVE-SPA-LS-SVM	60
	Chicken fillets	*Pseudomonas* detection	PLS regression	61
	Packaged beef	Spoilage detection	SVR, GMM, MDL	62
	Beef and horsemeat (minced)	Detection of adulteration of minced beef with horsemeat	PCA, PLS-DA, RF, SVM	63
	Narrow-leaved oleaster	Geographical origin identification of dry narrow-leaved oleaster fruits	PCA, PLS-DA, SVM, CNN	64
	Honey	Botanical origin classification	SVC, kNN	65
Spectroscopy	Minced beef	Spoilage detection/sensory classification	PLSR, GA-GP, GA-ANN, SVR (various kernel functions)	66
	Milk	Fatty acid composition determination	PLS, modified PLS	67
	Edible oils	Trans fatty acid determination	PLSR	68
	Chicken, pork, turkey, lamb, beef	Authentication of species and the distinct muscle groups within these species	PC-DFA, GA-MLR	69
	Beef, turkey	Adulteration detection	PCA, LDA, PLSR	70
	Beef burger	Adulteration detection	PLS-DA, SIMCA, low- and mid-level fusion strategies based on PLS	71
	Beef, horsemeat	Discrimination between beef and horsemeat	PCA	72
	Tommy Atkins mangoes	Quality control analysis	PLSR	73
	Barley, chickpeas, sorghum	Cultivar identification	SVM, PLS-DA	74
	Green salads	Assessment of microbial contamination	PLSR	75
	Chinese tea	Testing of total polyphenols, caffeine, free amino acids	PLS	76
	Brazilian coffee	Quality determination of arabica coffee (identification and quantification of adulterations such as robusta coffee)	PLS, PCA	77
	Rice	Quality and authenticity analysis of rice	PCA, kNN, SVM	78
	Salted minced meat (green ham)	Minced meat composition diagnostics	PLS, RF regression	79
	Avocadoes	Quality assessment of avocado fruit (fresh dry matter content)	PLSR	80
eNose	Beef fillets	Spoilage detection/sensory classification	PCA → DFA-SVM and SVR (RBF kernel)	81
	Table olives	Sensory classification	PCA, HCA, DFA, MLP-NN	82
	Tomatoes	Detection of microbial contamination	PCA	83
	Strawberries	Detection of fungal disease	PCA, ANN (MLP)	84
	Catfish fillets	Sensory classification/off-flavor detection	PCA, ANN, QF	85
	Peaches	Assessment of firmness, sugar content, acidity	PCA, LDA, PCR, PLSR	86

(*Continued*)

Table 1 (*Continued*)

Sensor types	Food types	Purpose	Data analysis methods	Reference
eNose, acoustic	Mangoes	Assessment of ripeness/maturity	PCA, LDA, LDA-ANN	87
eNose, eTongue	Strawberry juice	Discrimination among processing approaches	LDA, PLSR, SVM, RF	88
	Robusta coffee	Discrimination among varietal origins	PCA, kNN, PLS-DA, BP-ANN	89
	Minced mutton	Adulteration detection (proportion of pork in minced mutton)	MLR-PLS, BP-ANN	90
eTongue	Orange beverage, Chinese vinegar	Authentication/discrimination among brands	PCA, BP-ANN, SVM, RF	91
Imaging, spectroscopy, eNose	Pork	Freshness detection (TVB-N content)	PCA, BP-ANN	92
	Minced beef	Meat spoilage prediction	PCA, OLS, SLR, PC-R, PLSR, RF, kNN-R, CVM-R	93
Imaging, spectroscopy	Minced beef	Identification of frozen-then-thawed minced beef labeled as fresh	PLS-DA, SVM	94
	Lychee fruit	Micro-damage detection	PLS-DA, LS-SVM	95
	Minced pork	Estimation of microbiological spoilage	PLSR	96
	Pineapple	Quality assessment	PLSR, SVM, PLS-DA	97

Table adapted with permission from Reference 98; copyright 2016 Elsevier.

Abbreviations: ANN, artificial NN; BP, backpropagation; CNN, convolutional NN; CVM-R, Cramer-Von Mises in R; DFA, discriminant function analysis; EDM, Euclidean distance measure; GA, genetic algorithm; GMM, Gaussian mixture modeling; GP, genetic programming; HCA, hierarchical cluster analysis; kNN, *k*-nearest neighbor; LAB, lactic acid bacteria; LDA, linear discriminant analysis; LS, least-squares; MDL, minimum description length; MLP, multilayer perceptron; MLR, multiple linear regression; NN, neural network; OLS, ordinary least squares; PC, principal component; PC-R, PC regression; PCA, PC analysis; PCR, polymerase chain reaction; PLS-DA, partial LS discriminant analysis; PLSR, partial LS regression; QF, quality factor; RBF, radial basis function; RF, random forest; SAM, spectral angle measure; SCM, spectral correlation measure; SIMCA, soft independent modeling by class analogy; SLR, stepwise linear regression; SPA, successive projections algorithm; SVC, support vector clustering; SVM, support vector machine; SVR, support vector regression; TVB-N, total volatile basic nitrogen; UVE, uninformative variable elimination.

We can surpass the former limitation by employing the cloud or distributed storage space, and processing speeds are getting higher as we speak. To this end, noninvasive, rapid, and more efficient detection methods would provide a holistic approach toward food safety with increased classification accuracy, traceability, and authenticity, as presented in the cases of oil, beer, and almonds (99).

3. NEXT-GENERATION STRATEGIES

Alongside the abovementioned approaches and tools, traceability can also be achieved through the real-time monitoring of production, supply, and distribution chains, which can be made available to primary production and final (products) food business operators and retailers through a live tracking system based on, e.g., blockchain technology that combines mobile and web technologies. The quantification of parameters will be used to (*a*) monitor the safety of foods derived from a diversity of supply chains, (*b*) develop and implement PAT (10) within food processing and manufacturing to ensure process efficacy and validation for hazard control, and (*c*) assess food safety by providing simple and practical decision-support tools for agri-food business operators and stakeholders. In parallel, live tracking methods for labeled food products will be implemented [e.g., unique identifiers such as QR codes and NFC (near-field communication) tags] to allow for continuous food quality control in an online food chain management platform. Since biological, chemical, and environmental hazards can occur, evolve, and differ in their type and level at the various stages of food processing, safety controls in the form of sensitive sensors, microbiological tests, and metagenomics have to be implemented at critical production stages.

The targeted scientific breakthrough is the implementation of the abovementioned food safety tracking system in tandem with data science, big data, data analytics, Internet of things (IoT)

concepts (defined as the network of devices that gather and convey data via the Internet), and real-time data feeds acquired from different sensors under a common infrastructure where, e.g., blockchain can be used to efficiently record transactions, execute dynamic actions via smart contracts, and provide trustworthiness and transparency for food industry actors.

The ultimate goal of this holistic approach is to establish the foundation for developing next-generation monitoring platforms for food safety through a simple, smart tool driven by data science and big data that enables real-time predictions for the safety profile of a given food product. To achieve this, the European Food for Life (101) platform has proposed (*a*) strategic research programs in the use of omics technologies and utilization of big data analytics in the food sector, (*b*) strategic ecosystems (looking outside the box) through collaborations with companies/institutions who have the new skills and resources necessary to help generate and interpret the data (e.g., scientific computing experts), and (*c*) appropriate consortia that will drive progress and standardization (e.g., the COMPARE consortium on genome sequencing of infectious pathogens).

The establishment of a virtual center for food microbiome and food-omics would require both multiagency collaborations and extensive interdisciplinary efforts. Such a virtual center could boost progress in all fields of food-related research with the attendant reduction of costs that such collaborative initiatives offer.

3.1. Next-Generation Sequencing

The application of NGS in food safety is still considered in its infancy (102), especially when compared to the study of infectious diseases and even plant and soil genomics. Although systems-level omics analysis and sequencing in general are crucial for fundamentally understanding microbial behavior across the manufacturing, processing, and storage of foods, their application as truly integrated risk management or diagnostic tools remains limited to regulatory authority sampling plans and the back-tracing of pathogens following an outbreak. The main reason for this is the very nature of the sequencing techniques, which, despite their massively parallel high-throughput yield, involve complex and often time-consuming steps prior to sequencing, such as molecular extraction, library preparation, and quality control. Additionally, the bioinformatics and downstream analysis of raw sequencing reads remains one of the main bottlenecks to applying NGS as a real-time or near-real-time diagnostic tool for food safety (50).

While the sheer size of the data generated through NGS and third-generation sequencing platforms is likely to remain a challenge, the latest advances in computational methods and storage solutions from data science can provide an integrated framework for real-time risk management and control of foodborne pathogens. Additionally, the methods in current practice from other research fields including clinical setup could be adapted to develop a molecular-based diagnostic framework for near-real-time food safety monitoring and management. For instance, raw sequencing reads have been previously used to predict drug resistance for *Mycobacterium tuberculosis*, which has noticeably improved the resolution and timeliness of tuberculosis diagnosis (103). This was achieved through the development of a web-based tool, TB-Profiler, which accepts raw sequencing reads in the FASTQ format of sequenced samples extracted from hospital patients. The reads are preprocessed and aligned to the reference genome on the server side, before single-nucleotide polymorphisms (SNPs) and indels are called. The list of variants is then compared to a large database of drug resistance polymorphisms and the drug resistance is then predicted based on the variant scores. The speed of the current process can be massively improved through the direct matching of read probes, instead of relying on read alignment and variant calling. In order to develop a similar system for pathogen monitoring in foods, several objectives need to be met. This includes (*a*) the creation of an open database for functionally annotated genomes of food pathogens

coupled with gene expression and proteomic profiling data; (*b*) the implementation of a dynamic barcoding system for accurate species identification [a barcode is made up of a small combination of SNPs that together express a pattern of variation specific to a particular species, strain, or even isolate (104)]; (*c*) the creation of a publicly available web-based platform allowing food producers, regulatory authorities, and researchers to submit sequencing data extracted from a given sample for rapid identification and quantification; and, finally, (*d*) the creation of a genomic and barcoding database with a modular design, allowing knowledge to expand as new data become available.

3.2. Data Science in the Food Sector

A lot of emphasis on food safety occurs during the paddock-to-purchase phase of the supply chain; however, a significant number of food poisoning incidents happen postpurchase due to improper handling by consumers. Proposed methods to tackle this include applications (apps) that contain safety information about safe food transport and storage practices, but this requires active engagement from consumers (105). With higher-resolution equipment becoming increasingly available, more studies are being conducted with the purpose of assessing food quality, safety, and authentication. The data generated from these experiments often need novel preprocessing and analytical methods to reduce the dimensionality of and model the resulting dataset for prediction or classification. Although the increasing variety of approaches to food safety detection allows for more comprehensive experimental designs, it also raises the issue of nonapplicable algorithms due to experiment-specific results; this is leading to a movement toward meta-analyses with open-source data as well as efficient information-fusion methodologies (99).

Data science aims to apply advanced statistical approaches such as ML, AI, and pattern recognition to unravel hidden patterns within large volumes of data, combined with modern solutions for data integration, storage, and visualization. With rapid advances in sensors, mobile technologies, and computer processors, data science has been gaining popularity across various sectors, including in almost every field of research and everyday applications. With the power to run informative apps and with built-in sensory devices like cameras, mobile phones are becoming an increasingly powerful, although still underutilized, scientific research tool.

Nowadays, there are several new but not yet adopted data science technologies that could transform and shape the food safety field of the future, such as big data approaches, AI, blockchain, IoT, and the digital twin (DT) (a digital representation of a physical object such as a city or factory). The combination of these technologies would have a great impact toward the so-called Industry 4.0 in the food sector. Each technology, while if not complementary to the others, has critical limitations and properties that when combined with the others results in an enhanced system, e.g., AI coupled with DT for predictive modeling risk assessment. When integrated with IoT technology, AI will be enhanced through the abundance of and inherent variabilities in IoT data; furthermore, the integration of blockchain with AI would reinforce the transparency and traceability of such data. Importantly, Open-source libraries, social media, and mobile devices are expected to offer great solutions in the food sector. These issues are analyzed below.

3.2.1. Internet of things/foods. IoT is crucial for significantly optimizing traceability (106–108) and food safety (109–112) and minimizing waste (113). IoT is based on the interconnection of all things (e.g., sensors, devices, machines, computing devices) via communication media [e.g., WiFi, Bluetooth, RFID (radio-frequency identification)]. Recently efforts have been focused on embedded sensors, low-power wireless communications, and signal processing algorithms, allowing for devices like mobile phones to collect and transfer data to repositories via transferring channels like WiFi, which facilitate real-time monitoring and control. In other words, the connected

sensors could be deployed in, on, or at line throughout the production and distribution chain, where multidimensional and multivariate data are sent over on-the-fly to a cloud-based central or distributed data repository. On the cloud server side, validated mathematical models can receive the POST (power-on self-test) request from the sensors as input variables and can send back to the sensor the safety profile and risk parameters in real time.

Recent advances in technology have led to the implementation of IoT frameworks in the agricultural sector, advancing the prospects of real-time production analytics, which in the case of food supply chains can be tailored to global and local regulations. Automated hazard analysis and critical control points (HACCP) (114) checklists are being used throughout manufacturing, production, and transporting procedures so that companies can get access to meaningful and consistent data that enable them to put into practice some food safety solutions. Additionally, IoT, coupled with big data approaches, will lead to safer and more sustainable transportation around the globe (115, 116). For example, meat and fish need to be stored and shipped at a certain monitored temperature, humidity, etc. in order to ensure safety, which can be constantly and in real-time monitored by wireless sensors (e.g., RFID). In the event of contamination, sensors can be used as an early alert system and for traceability of the products. To date, most applications of IoT to food safety are still limited and only in early stages of development.

3.2.2. Big data. Big data is a vastly underutilized tool in the sectors of food safety and quality. It is broadly defined as "high-volume, high-velocity, high-veracity, and/or high-variety information assets that require new forms of processing to enable enhanced decision-making, insight discovery, and process optimization" (117), which matches the description of sensory data generated within the agricultural industry. After the initial cost of purchasing and installation, sensory devices are cost efficient to operate and generate an overabundance of data that can be processed for free. There has been a surge in the development of sensory devices with higher sensitivity to subtle changes, thus improving the ability to identify small but significantly influential feature changes. The downside to higher-resolution sensory devices, along with the need for increased storage space, is the need for increased computational power and optimized algorithms to tackle the volume of data generated in real time, problems which may be eased by sensory fusion technology, allowing for a holistic multivariate modeling approach. Cloud storage is the most recent addition to the armory of big data science, allowing for wider availability and long-term remotely accessible historical datasets. This holds true for the food safety sector, as there are many online databases such as the WHO's FOSCOLLAB (food safety collaborative platform), which provides food and chemical risk reports collected by GEMS/Food (Global Environment Monitoring System's Food Contamination and Assessment Program) (105). With roughly 600–800 entries per month, FOSCOLLAB is the largest freely accessible dataset for food safety monitoring purposes. In conclusion, big data science can lead to real transparency across the farm-to-fork chain, facilitating unbiased data-driven decisions on safety and quality, process line sanitation, and supply chain issues (faster shipments, lower temperatures, etc.), as well as predictive risk assessments of shelf life and spoilage (118–120).

3.2.3. Blockchain. Since the emergence of blockchain technology in 2008 for use in financial systems (121), its potential uses have grown into other sectors such as healthcare, legal contracts, and food supply chain regulations. Blockchain provides transparency, efficiency, security, and safety, benefitting every stakeholder in the food sector due to their need to demonstrate the quality of their adopted methods and products (122) and satisfy national and international legislative requirements concerning the traceability of food (FSMA; see also 123). Olsen & Borit (124) provided an extensive overview of a general food traceability system and its components.

Blockchain technology can be utilized along all main stages of an agri-food supply chain (125): production, processing, distribution, retailing, and consumption. Considering the number of supply chain phases and the ever-increasing complexity of globalized trade, tracing food products is increasingly difficult for stakeholders in the food sector (126).

Roughly, food safety can be thought of as the hygienic way of processing and managing the food so as to prevent illnesses. Several flows of goods have compromised food safety and quality, as described by Creydt & Fischer (127), and the WHO has reported that more than 23 million people suffer from contaminated food in Europe every year (128, 129). Blockchain provides solutions for the improvement of traceability and transparency by recording information at every stage of the supply chain to drive better hygienic conditions and help identify contaminated products, fraudulent activity, and risks as early as possible.

Blockchain does, however, face challenges, such as how to handle the overabundance of information resulting from the physical flow of goods and how to handle a high number of transactions in a single system cost-effectively. Another significant shortcoming is that while advancing and addressing transparency within the industry, blockchain risks a loss of confidentiality, a major issue that has to be surpassed. Although the technology is considered new, it should be stressed that applications and proofs-of-concept are fast emerging (see the literature cited and tables in References 130–132).

3.2.4. Artificial intelligence. The food industry is rapidly employing AI (and ML as a subset of techniques), which has been proven to help and advance food waste minimization (133), hygiene at processing sites (134), food sorting (133, 135, 136) food safety (137), etc. The recent success of AI in the food industry is based on automated systems and noninvasive sensors that provide abundant data for near-real-time assessment of various properties of the food (quality, quantity, contamination, shelf life, etc.). Through AI, food stakeholders will be enabled to utilize predictive modeling in food processing/preparation facilities (138) to exclude high-risk manipulations and environments and provide time-crucial alerts to managers so as to prevent contamination and take proactive or corrective actions. Thus, food stakeholders using AI would be able to collaborate with public health agencies and research (academic or industry) partners toward the identification of contamination sources (139) so as to prevent foodborne outbreaks (140).

3.2.5. Social media and mobile devices serving food safety and quality. In addition to information and communication technology (ICT), IoT, and sensors, food safety and quality have a number of allies: (*a*) social media, (*b*) mobile devices, and (*c*) apps. Several existing apps have been available to consumers. This is the case with dairy products and especially milk; for example, milk can be classified in classes according to protein content and quantification (141). Recently, several publications have claimed that hazards and quality indices can be detected in various foods, although certain limitations do not allow the implementation in real life (142–144).

Since social media platforms like Facebook, Twitter, LinkedIn, and YouTube are receiving more attention as a potential source of food safety data (145), knowledge can be gained through food safety–related discussions, opinions, or online questionnaires. Web mining is one of the most commonly used approaches to collect and mine social media data, and the majority of these platforms offer APIs (application programming interfaces) to facilitate this. The derived information is based on a fully expandable database, which allows for the collection of publications from all available media that can be scanned and digitized. Digital sources (news sites, blogs, forums, and social media), TV, and radio stations can be also monitored.

3.2.6. Digital twins. Through advances in cyber–physical systems interconnections (cloud technology, IoT, AI, and big data analytics), the DT solution (146) has gained increasing attention, especially due to the possibilities it offers toward Industry 4.0. As with industries in other fields, the food industry has just begun to take advantage of (*a*) real-time simulations, (*b*) the possibility of intelligent decision-making, and (*c*) cost-effective solutions addressing stakeholders' demands. A DT is a digital model of a physical element or process with data connections, enabling a convergence between the digital and physical versions at an appropriate rate of synchronization (147). In essence, the concept is that of a mathematical model that can describe a process/product so as to analyze and help make decisions about it according to the needs of stakeholders. Concerning food safety, by employing a DT of physical assets related to product safety and simulations of various parameters (temperature, risk of environmental contamination, etc.) while also using real-time data from sensors and other IoT-connected services, one can use a DT not only as a monitoring tool but also as an early-warning system. Thus, any stakeholder would have a predictive simulation of poor-quality products or safety issues that could emerge. We must note that every simulation would be tested in the real world, not only for validation purposes but also for bringing the DT closer to reality, increasing its efficiency. A very recent example of this in the food sector is the use of DTs of products and production by Siemens, where data are acquired from the factory (148, 149) and across the supply chain, increasing farm-to-fork visibility for the management of food safety events, leading to faster reaction times and the elimination of potentially hazardous products (149, 150). Other interesting applications of DT in the food sector include aquaponics for production optimization purposes (151).

3.2.7. Satellite analysis. The motivation behind food adulteration is to increase the quantity of a final product cheaply, which often happens due to low levels of food security for high-value or less-nutritious food products. It is estimated that globally there are 475 million smallholder farms, which are less than two hectares in size. These exist predominantly in impoverished areas, such as rural China (152). In 2008, the World Bank produced the *World Development Report 2008* as a means of identifying pathways to improve the financial standing of citizens in such areas, which can be achieved by providing funding to increase productivity and food security. Smallholder farms often rely on family labor due to the low profitability of selling net crop yields in impoverished communities. As such, they cannot afford to buy high-quality technology or systematically run ground surveys to accurately record the crop yields data that policy makers use to target areas of improvement in such areas, with errors frequently above 50% (153). One tested approach to counter inaccurate yields reports is to use high-resolution satellite imagery, in tandem with ML methods and a computer model that uses local weather forecasts, to predict crop growth. Newer models of CubeSat satellites can provide a resolution less than 5 m, eliminating issues of small and irregularly bounded fields that hindered previous attempts at satellite analysis (154). Results in these approaches have seen ML outperform traditional methods of measuring NDVI (normalized difference vegetation index) and EVI (enhanced vegetation index), likely due to the sensitivity of sensory devices in picking up changes in chlorophyll-rich plants associated with nutritional deficiency and therefore low yield (155). Progress in studies measuring crop yields via remote sensors such as satellite imagery will help policy makers to maximize food security in vulnerable communities and unregulated areas.

3.2.8. Online food safety databases and toolboxes for data analytics. Over the past decades, food regulatory authorities [e.g., European Food Safety Authority in the European Union, FDA and USDA in the United States, CFDA in China, Food Standards Australia New Zealand] have

identified assurance of food safety and quality as one of their ultimate goals. In the European Union the Rapid Alert System for Food and Feed (156) has been developed and is considered to be the main food safety online database used by regulatory authorities, industry, and scientists. Similar online food safety databases can be found in the US Import Refusal Report (157), which reports on those products for which the FDA refused admission to part or all of the product offered for importation. Alerts and notifications are also available in China's food safety system (158).

Numerous farm-to-fork models for quantitative microbiological risk analysis (QMRA) are available in the literature. QMRA requires the development of complex mathematical models; dedicated software packages include @RISK (Palisade Company, Ithaca, NY) and the mc2d package in R (159). More details of the available software are provided in a review by Tenenhaus-Aziza & Elluze (160). Further information regarding the available toolboxes for data analytics are given in **Table 2**.

4. CHALLENGES OF DATA SCIENCES IN FOOD SAFETY AND BEYOND

Fragmented information and the lack of communication can have a major impact on the food supply chain, causing inefficiency, waste, and mistrust among producers, suppliers, and their customers. A key strategy that should be followed is to encourage all food chain actors to improve food integrity and integrate their content, and especially their data, on a secure cloud platform (including heterogeneous information derived from food microbial ecosystems throughout a product's life) in order to enhance the predictive power of product-specific food safety models. The food sector generates datasets in each step of the food chain that then need to be stored in a data repository until they are needed (**Figure 1**). Currently, there are no defined/established standards or data exchange pipelines to allow these datasets to integrate and merge in order to be more useful and reusable. However, there are several data standards for specific stages/sites of the food chain that have emerged in the last few years to enable research laboratories and food value-chain actors to ensure data availability and provide querying functionalities in standardized formats. Examples of data standards at different stages of the farm-to-fork chain include AGROVOC (161), FoodEx2 (162), and FoodOn (**https://foodon.org/**) for data in the agri-food sector; GS1 + EPCIS/GTIN (Electronic Product Code Information Services/Global Trade Item Number) (**https://www.gs1.org/standards/**) for food supply chain data; HACCP Ontology (**http://www.haccpalliance.org/**) for hazard data; and FoodEx2 (162) and E-lab (163) for laboratory data.

Novel data management solutions are required for the massive amount of high-dimensional data generated by the large diversity of available analytical and high-throughput platforms. One possible approach toward Industry 4.0 (164) should be the storage of data in a central repository complemented with a flexible, user-friendly, and possibly cloud-enabled web interface. Specifically, creating efficient online monitoring systems in real time requires the ability to integrate all the available sensor data and to run analyses in real time. As research in the food sector becomes increasingly data driven, so does the interactive process among multiple stakeholders across the world. High-performance computing, visualizations, simulations, workflows, risk assessments, and other advanced tools and approaches such as distributed networks or grid computing need to be developed and employed in order to enable the necessary worldwide collaboration. Data play a key role in the process of accelerating and democratizing this modernization. It is also important that the data abide by the four principles of FAIR (findable, accessible, interoperable, and reusable) (165). One of the most important issues for enabling interoperability via data standardization is the sufficient knowledge representation via ontologies, semantic nets, and rules, as stressed by Jonquet et al. (166) and the references they cite. When this obstacle is surpassed,

Table 2 MATLAB, R, and Python libraries available for chemometric and machine learning applications

Language	Library	Methods	URL
MATLAB	Statistics and Machine Learning Toolbox	HCA, *k*-means, ANOVA, MLR, LDA, kNN, SVM, RF, other methods	**http://www.mathworks.com/**
	Neural Network Toolbox	ANNs	**http://www.mathworks.com/**
	PLS Toolbox	MLR, PLS, PC-R, preprocessing methods	**http://www.eigenvector.com/**
	LibPLS	PLSR; PLS-DA; LDA; various methods for preprocessing, variable selection, and outlier detection	**http://libpls.net/**
	iToolbox	PLS variants with intervals (iPLS, biPLS, fiPLS, siPLS, mwPLS)	**http://www.models.life.ku.dk/iToolbox/**
	PLS-Genetic Algorithm Toolbox	GA-PLS	**http://www.models.life.ku.dk/GAPLS/**
	LIBSVM	SVM	**http://www.csie.ntu.edu.tw/~cjlin/libsvm/**
	LS-SVMlab	LS-SVM	**http://www.esat.kuleuven.be/sista/lssvmlab/**
R	The R Stats Package	HCA, PCA, *k*-means, other statistical functions	**http://www.r-project.org/**
	chemometrics	PCA, PLSR, other regression methods (lasso, ridge), tools for CV, clustering, etc.	**http://cran.r-project.org/web/packages/chemometrics/**
	ChemometricsWithR	PCA, GA for variable selection	**http://cran.r-project.org/web/packages/ChemometricsWithR/**
	pls	PLSR, PC-R	**http://cran.r-project.org/web/packages/pls/**
	plsgenomics	PLSR, DA, ridge PLS	**http://cran.r-project.org/web/packages/plsgenomics/**
	gpls	Generalized PLS	**http://bioconductor.org/packages/release/bioc/html/gpls.html**
	cluster	Cluster analysis methods	**http://cran.r-project.org/web/packages/cluster/**
	neuralnet	ANNs	**http://cran.r-project.org/web/packages/neuralnet/**
	e1071	SVM (LIBSVM) and other clustering methods	**http://cran.r-project.org/web/packages/e1071/**
	randomForest	RF	**http://cran.r-project.org/web/packages/randomForest/**
	gbm	Generalised boosted regression	**http://cran.r-project.org/web/packages/gbm/**
	robustbase	Robust linear regression	**http://cran.r-project.org/web/packages/robustbase/**
	FNN	Fast nearest neighbor search algorithms and applications	**http://cran.r-project.org/web/packages/FNN/**
	tidyverse	Data visualization	**https://cran.r-project.org/web/packages/tidyverse/**
	BLR	Bayesian linear regression	**https://cran.r-project.org/web/packages/BLR/**
	ranger	Rapid RF	**https://cran.r-project.org/web/packages/ranger/**
Python	NumPy	Array computing	**https://pypi.org/project/numpy/**
	pandas	Flexible data structures	**https://pypi.org/project/pandas/**
	Matplotlib	Data visualization	**https://pypi.org/project/matplotlib/**
	scikit-learn	Comprehensive regression and classification model suite	**https://pypi.org/project/scikit-learn/**
	scikit-image	Image edge detection and segmentation	**https://pypi.org/project/scikit-image/**
	Keras	Deep neural networks	**https://pypi.org/project/Keras/**

Table adapted with permission from Reference 98; copyright 2016 Elsevier.

Abbreviations: ANN, artificial neural network; ANOVA, analysis of variance; biPLS, backward interval PLS; CV, computer vision; DA, discriminant analysis; fiPLS, forward interval PLS; GA, genetic algorithm; HCA, hierarchical cluster analysis; iPLS, interval PLS; kNN, *k*-nearest neighbor; LDA, linear discriminant analysis; LS, least-squares; MLR, multiple linear regression; mwPLS, moving window PLS; PC-R, principal component regression; PCA, principal component analysis; PLS, partial LS; PLSR, partial LS regression; RF, random forest; siPLS, synergy interval PLS; SVM, support vector machine.

computer systems will be able to reach out to the data autonomously (findable principle) and use them (reusable principle) for solving complex tasks such as predictive modeling development, near-real-time monitoring, safety assessments, risk assessments, etc. Thus, FAIRification of food safety data could facilitate data standardization and interoperability for scientific data management, accelerating research and food chain actors toward an Industry 4.0 for the food sector.

Finally, another critical issue for data from the agri-food sector, related to the accessibility principle of FAIR, is privacy and anonymization. In some research domains, such as health, data are anonymized by efficient removal or encryption of personal information, which is relatively easy to do due to the tight and controlled nature of clinical studies. However, anonymization has not been addressed so far in the case of agri-food data (within a farm-to-fork approach for safety). The main difficulty is the spatiotemporal nature of the data (167), either from farms/fields or from food processing units and distribution, a problem that does not yet have an adequate solution since it leads to granularity distortions. A recent study on anonymizing public participation GIS (geographic information systems) data (168) proposed a method of anonymizing GIS data inside the context of GDPR (General Data Protection Regulation), but more studies are still needed for adapting such approaches to the food sector.

All of the recent advances in data science technologies can lead to updated data acquisition, enhanced analysis, and traceability across the whole farm-to-fork chain (**Figure 1**). In addition to the emerging opportunities, we should acknowledge issues that need to be prioritized. A first critical area that needs to be transformed in order to enable Industry 4.0 for the food sector is a data and information infrastructure that will support open-access data operating under the FAIR principles. Data science and food safety should converge, and each affects and shapes the other via intercommunication, influencing their synchronized development. In this way, ML will adapt to the needs of food research while food safety will harvest the benefits of data science and information technology. Another critical issue that needs to be addressed, mainly concerning stakeholders and not the final consumers, is the development of data anonymization and privacy methods that do not distort the data or the context of the original study. One hope is that AI and ML, coupled with other information technologies (discussed above) and data not produced in a laboratory, could be used for advance surveillance and alert systems for foodborne outbreaks and diseases (169, 170). Data that can be used in tandem with lab-based sensors can originate from social media as well.

5. SUMMARY POINTS

Innovative, integrated knowledge repositories that employ big data from all stages of production to entire food chains, that are based on product history, and that can be accessed globally via a cloud database that also accommodates decision-making tools establish the foundation for a new line of technology that can predict food safety. The changes that ICT, IoT, and big data will bring to the food sciences and their stakeholders are much greater than most people can anticipate, especially with the extremely dynamic advancements in smart devices. Users' measurements of food quality/safety conditions will be stored, analyzed, and shared in the cloud. The in-depth data obtained from thorough line production measurements will offer multifaceted approaches to upgrading current regulatory methods. The prospect of observing singular samples within the food chain at multiple time points will help identify, pinpoint, and analyze current weaknesses in maintaining food safety and quality standards, changing and refining the operations and policy-making decisions of food stakeholders, such as food operators, inspectors, and researchers. These new activities will be related to data and computer science, while the needs of conventional food scientists

may be impacted dramatically. In the era before big data and smart devices, food science studies were conducted almost exclusively by academics, regulatory authorities, and the media. Consumer education is a big issue, and it has been recognized that it could take years or decades to educate the general population and build a food safety culture and awareness of food sector. It is clear that in the age of big data, scientific suggestions will be oriented toward consumers and (specific) food products, and this knowledge will be readily available on consumers' smart devices. In other words, the impact of technological and educational advancements in food sciences will be amplified with the use of big data and data science. Users (consumers, authorities, food operators, etc.) will not only get the recommendations they need but also have online the tools to fulfil those needs.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Our approaches and ideas were introduced, developed, and implemented over the last five years with support from the SYMBIOSIS-EU project [EU FP7 (7th Framework Program)]; the PhasmaFOOD, IMPAQT, and DiTECT projects (EU Horizon 2020 program); the iMeatSense project (cofounded by the European Social Fund and the Greek General Secretariat of Research and Technology); and the QAPP project (cofounded by the European Union and the Greek Competitiveness, Entrepreneurship & Innovation program). We would like to apologize to all colleagues whose work could not be mentioned due to space limitations.

LITERATURE CITED

1. FAO (Food Agric. Organ.). 2011. *Global food losses and food waste: extent, causes and prevention*. Tech. Rep., Food Agric. Organ. U.N., Rome
2. Eur. Comm. 2005. Commission regulation (EC) no 2073/2005 on microbiological criteria for foodstuffs. *Off. J.* L388:1–26. **https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32005R2073**
3. Lawrence F. 2013. Horsemeat scandal: the essential guide. *Guardian*, Feb. 15. **https://www.theguardian.com/uk/2013/feb/15/horsemeat-scandal-the-essential-guide**
4. Elliott C. 2014. *Elliott review into the integrity and assurance of food supply networks: final report*. Tech. Rep., U.K. Gov., London
5. Natl. Food Crime Unit. 2020. *Food crime strategic assessment 2020*. Tech. Rep., Food Stand. Agency, London
6. FAO (Food Agric. Organ.). 1996. *Rome declaration on food safety*. World Declar., Food Agric. Organ. U.N., Rome. **http://www.fao.org/3/w3613e/w3613e00.htm**
7. Yasuda JK. 2015. Why food safety fails in China: the politics of scale. *China Q.* 223:745–69
8. Guo Z, Bai L, Gong S. 2019. Government regulations and voluntary certifications in food safety in China: a review. *Trends Food Sci. Technol.* 90:160–65
9. Eur. Comm.. 1999. *White paper on food safety*. White Pap. COM/99/719 final, Eur. Comm.
10. Nychas G-JE, Panagou EZ, Mohareb F. 2016. Novel approaches for food safety management and communication. *Curr. Opin. Food Sci.* 12:13–20
11. Ellis DI, Muhamadali H, Allen DP, Elliott CT, Goodacre R. 2016. A flavour of omics approaches for the detection of food fraud. *Curr. Opin. Food Sci.* 10:7–15
12. Kosmides AK, Kamisoglu K, Calvano SE, Corbett SA, Androulakis IP. 2013. Metabolomic fingerprinting: challenges and opportunities. *Crit. Rev. Biomed. Eng.* 41:205–21

13. Lytou AE, Panagou EZ, Nychas G. 2019. Volatilomics for food quality and authentication. *Curr. Opin. Food Sci.* 28:88–95
14. Macrae R. 1981. Recent applications of high pressure liquid chromatography* to food analysis. *Int. J. Food Sci. Technol.* 16:93–110
15. Nollet LML, Toldra F. 2012. *Food Analysis by HPLC*. Boca Raton, FL: CRC Press. 3rd ed.
16. Yashin YI, Yashin AY. 2004. Analysis of food products and beverages using high-performance liquid chromatography and ion chromatography with electrochemical detectors. *J. Anal. Chem.* 59:1121–27
17. Lehotay SJ, Hajšlová J. 2002. Application of gas chromatography in food analysis. *Trends Anal. Chem.* 21:686–97
18. Rohn S. 2014. Gas chromatography in food analysis. In *Practical Gas Chromatography: A Comprehensive Reference*, ed. K Dettmer-Wilde, W Engewald, pp. 745–66. Berlin: Springer-Verlag
19. Hird SJ, Lau BPY, Schuhmacher R, Krska R. 2014. Liquid chromatography-mass spectrometry for the determination of chemical contaminants in food. *Trends Anal. Chem.* 59:59–72
20. Vazquez-Roig P, Pico Y. 2012. Gas chromatography and mass spectroscopy techniques for the detection of chemical contaminants and residues in foods. In *Chemical Contaminants and Residues in Food*, ed. D Schrenk, pp. 17–61. Cambridge, UK: Woodhead
21. Hatzakis E. 2019. Nuclear magnetic resonance (NMR) spectroscopy in food science: a comprehensive review. *Compr. Rev. Food Sci. Food Saf.* 18:189–220
22. van Duynhoven JPM, Belton PS, Webb GA, van As H, eds. 2013. *Magnetic Resonance in Food Science: Food for Thought*. London: RSC Books
23. Rodriguez-Saona LE, Allendorf ME. 2011. Use of FTIR for rapid authentication and detection of adulteration of food. *Annu. Rev. Food Sci. Technol.* 2:467–83
24. van de Voort FR. 1992. Fourier transform infrared spectroscopy applied to food analysis. *Food Res. Int.* 25:397–403
25. Valand R, Tanna S, Lawson G, Bengtström L. 2020. A review of Fourier transform infrared (FTIR) spectroscopy used in food adulteration and authenticity investigations. *Food Addit. Contam. A* 37:19–38
26. Li Y-S, Church JS. 2014. Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. *J. Food Drug Anal.* 22:29–48
27. Li-Chan ECY. 1996. The applications of Raman spectroscopy in food science. *Trends Food Sci. Technol.* 7:361–70
28. Jin H, Lu Q, Chen X, Ding H, Gao H, Jin S. 2016. The use of Raman spectroscopy in food processes: a review. *Appl. Spectrosc. Rev.* 51:12–22
29. Zhang Z. 2017. Raman spectroscopic sensing in food safety and quality analysis. In *Sensing Techniques for Food Safety and Quality Control*, ed. X Lu, pp. 1–16. London: R. Soc. Chem.
30. Feng Y-Z, Sun D-W. 2012. Application of hyperspectral imaging in food safety inspection and control: a review. *Crit. Rev. Food Sci. Nutr.* 52:1039–58
31. Liu Y, Pu H, Sun D-W. 2017. Hyperspectral imaging technique for evaluating food quality and safety during various processes: a review of recent applications. *Trends Food Sci. Technol.* 69:25–35
32. Siche R, Vejarano R, Aredo V, Velasquez L, Saldaña E, Quevedo R. 2016. Evaluation of food quality and safety with hyperspectral imaging (HSI). *Food Eng. Rev.* 8:306–22
33. Khan MJ, Khan HS, Yousaf A, Khurshid K, Abbas A. 2018. Modern trends in hyperspectral image analysis: a review. *IEEE Access* 6:14118–29
34. Bonah E, Huang X, Aheto JH, Osae R. 2019. Application of hyperspectral imaging as a nondestructive technique for foodborne pathogen detection and characterization. *Foodborne Pathog. Dis.* 16:712–22
35. Tsakanikas P, Pavlidis D, Panagou E, Nychas G-J. 2016. Exploiting multispectral imaging for non-invasive contamination assessment and mapping of meat samples. *Talanta* 161:606–14
36. He Y, Reed S, Strobaugh TP Jr. 2020. Complete genome sequence and annotation of *Campylobacter jejuni* YH003, isolated from retail chicken. *Microbiol. Resour. Announc.* 9:e01307-19
37. He Y, Yan X, Reed S, Xie Y, Chen C-Y, Irwin P. 2015. Complete genome sequence of *Campylobacter jejuni* YH001 from beef liver, which contains a novel plasmid. *Genome Announc.* 3:e01492-14
38. Nielsen DW, Maki JJ, Looft T, Ricker N, Sylte MJ. 2020. Complete genome sequence of *Campylobacter jejuni* strain NADC 20827, isolated from commercial turkeys. *Microbiol. Resour. Announc.* 9:e01403-19

39. Carraro L, Marotta F, Janowicz A, Patavino C, Piccirillo A. 2019. Draft whole-genome sequences of 16 *Campylobacter jejuni* isolates obtained from wild birds. *Microbiol. Resour. Announc.* 8(26):e00359-19
40. Laksanalamai P, Steyert SR, Burall LS, Datta AR. 2013. Genome sequences of *Listeria monocytogenes* serotype 4b variant strains isolated from clinical and environmental sources. *Genome Announc.* 1:e00771-13
41. Hurley D, Luque-Sastre L, Parker CT, Huynh S, Eshwar AK, et al. 2019. Whole-genome sequencing-based characterization of 100 *Listeria monocytogenes* isolates collected from food processing environments over a four-year period. *mSphere* 4:e00252-19
42. Haendiges J, Gonzalez-Escalona N, Miller JD, Hoffmann M. 2019. Complete genome sequences of four *Salmonella enterica* strains associated with pistachios assembled using a combination of short- and long-read sequencing. *Microbiol. Resour. Announc.* 8:e00975-19
43. den Besten HMW, Wells-Bennik MHJ, Zwietering MH. 2018. Natural diversity in heat resistance of bacteria and bacterial spores: impact on food safety and quality. *Annu. Rev. Food Sci. Technol.* 9:383–410
44. Imran M, Bré J-M, Guéguen M, Vernoux J-P, Desmasures N. 2013. Reduced growth of *Listeria monocytogenes* in two model cheese microcosms is not associated with individual microbial strains. *Food Microbiol.* 33:30–39
45. Alessandria V, Ferrocino I, De Filippis F, Fontana M, Rantsiou K, et al. 2016. Microbiota of an Italian Grana-like cheese during manufacture and ripening, unraveled by 16S rRNA-based approaches. *Appl. Environ. Microbiol.* 82:3988–95
46. Spyrelli E, Stamatiou A, Tassou C, Nychas G-J, Doulgeraki A. 2020. Article microbiological and metagenomic analysis to assess the effect of container material on the microbiota of feta cheese during ripening. *Fermentation* 6:12
47. Doulgeraki AI, Papaioannou M, Nychas G-JE. 2016. Targeted gene expression study of *Salmonella enterica* during biofilm formation on rocket leaves. *LWT* 65:254–60
48. Argyri K, Doulgeraki A, Manthou E, Grounta A, Argyri A, et al. 2020. Microbial diversity of fermented Greek table olives of Halkidiki and Konservolia varieties from different regions as revealed by metagenomic analysis. *Microorganisms* 8:1241
49. Chaillou S, Chaulot-Talmon A, Caekebeke H, Cardinal M, Christieans S, et al. 2015. Origin and ecological selection of core and food-specific bacterial communities associated with meat and seafood spoilage. *ISME J.* 9:1105–18
50. Mohareb F, Iriondo M, Doulgeraki AI, Van Hoek A, Aarts H, et al. 2015. Identification of meat spoilage gene biomarkers in *Pseudomonas putida* using gene profiling. *Food Control* 57:152–60
51. Nieminen TT, Koskinen K, Laine P, Hultman J, Säde E, et al. 2012. Comparison of microbial communities in marinated and unmarinated broiler meat by metagenomics. *Int. J. Food Microbiol.* 157:142–49
52. Zhou L, Zhang C, Liu F, Qiu Z, He Y. 2019. Application of deep learning in food: a review. *Compr. Rev. Food Sci. Food Saf.* 18:1793–811
53. Shelhamer E, Long J, Darrell T. 2017. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39:640–51
54. Panagou EZ, Papadopoulou O, Carstensen JM, Nychas G-JE. 2014. Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage. *Int. J. Food Microbiol.* 174:1–11
55. He H-J, Sun D-W, Wu D. 2014. Rapid and real-time prediction of lactic acid bacteria (LAB) in farmed salmon flesh using near-infrared (NIR) hyperspectral imaging combined with chemometric analysis. *Food Res. Int.* 62:476–83
56. Gowen AA, O'Donnell CP, Taghizadeh M, Cullen PJ, Frias JM, Downey G. 2008. Hyperspectral imaging combined with principal component analysis for bruise damage detection on white mushrooms (*Agaricus bisporus*). *J. Chemom.* 22:259–67
57. Sharifzadeh S, Clemmensen LH, Borggaard C, Støier S, Ersbøll BK. 2014. Supervised feature selection for linear and non-linear regression of L*a*b* color from multispectral images of meat. *Eng. Appl. Artif. Intell.* 27:211–27
58. Fu X, Kim MS, Chao K, Qin J, Lim J, et al. 2014. Detection of melamine in milk powders based on NIR hyperspectral imaging and spectral similarity analyses. *J. Food Eng.* 124:97–104

59. Ropodi AI, Pavlidis DE, Mohareb F, Panagou EZ, Nychas GJE. 2015. Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. *Food Res. Int.* 67:12–18
60. Wu D, Shi H, He Y, Yu X, Bao Y. 2013. Potential of hyperspectral imaging and multivariate analysis for rapid and non-invasive detection of gelatin adulteration in prawn. *J. Food Eng.* 119:680–86
61. Feng Y-Z, Sun D-W. 2013. Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of *Pseudomonas* loads in chicken fillets. *Talanta* 109:74–83
62. Tsakanikas P, Pavlidis D, Panagou E, Nychas GJ. 2016. Exploiting multispectral imaging for non-invasive contamination assessment and mapping of meat samples. *Talanta* 161:606–14
63. Ropodi AI, Panagou EZ, Nychas G-JE. 2017. Multispectral imaging (MSI): a promising method for the detection of minced beef adulteration with horsemeat. *Food Control* 73:57–63
64. Gao P, Xu W, Yan T, Zhang C, Lv X, He Y. 2019. Application of near-infrared hyperspectral imaging with machine learning methods to identify geographical origins of dry narrow-leaved oleaster (*Elaeagnus angustifolia*) fruits. *Foods* 8:620
65. Noviyanto A, Abdulla WH. 2020. Honey botanical origin classification using hyperspectral imaging and machine learning. *J. Food Eng.* 265:109684
66. Argyri AA, Jarvis RM, Wedge D, Xu Y, Panagou EZ, et al. 2013. A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage. *Food Control* 29:461–70
67. Coppa M, Revello-Chion A, Giaccone D, Ferlay A, Tabacco E, Borreani G. 2014. Comparison of near and medium infrared spectroscopy to predict fatty acid composition on fresh and thawed milk. *Food Chem.* 150:49–57
68. da Costa Filho PA. 2014. Developing a rapid and sensitive method for determination of *trans*-fatty acids in edible oils using middle-infrared spectroscopy. *Food Chem.* 158:1–7
69. Ellis DI, Broadhurst D, Clarke SJ, Goodacre R. 2005. Rapid identification of closely related muscle foods by vibrational spectroscopy and machine learning. *Analyst* 130:1648–54
70. Alamprese C, Casale M, Sinelli N, Lanteri S, Casiraghi E. 2013. Detection of minced beef adulteration with turkey meat by UV–vis, NIR and MIR spectroscopy. *LWT* 53:225–32
71. Zhao M, Downey G, O'Donnell CP. 2014. Detection of adulteration in fresh and frozen beefburger products by beef offal using mid-infrared ATR spectroscopy and multivariate data analysis. *Meat Sci.* 96:1003–11
72. Boyacı İH, Temiz HT, Uysal RS, Velioğlu HM, Yadegari RJ, Rishkan MM. 2014. A novel method for discrimination of beef and horsemeat using Raman spectroscopy. *Food Chem.* 148:37–41
73. Marques EJN, de Freitas ST, Pimentel MF, Pasquini C. 2016. Rapid and non-destructive determination of quality parameters in the 'Tommy Atkins' mango using a novel handheld near infrared spectrometer. *Food Chem.* 197:1207–14
74. Kosmowski F, Worku T. 2018. Evaluation of a miniaturized NIR spectrometer for cultivar identification: the case of barley, chickpea and sorghum in Ethiopia. *PLOS ONE* 13:e0193620
75. Tsakanikas P, Fengou L-C, Manthou E, Lianou A, Panagou EZ, Nychas G-JE. 2018. A unified spectra analysis workflow for the assessment of microbial contamination of ready-to-eat green salads: comparative study and application of non-invasive sensors. *Comput. Electron. Agric.* 155:212–19
76. Wang J, Wang Y, Cheng J, Wang J, Sun X, et al. 2018. Enhanced cross-category models for predicting the total polyphenols, caffeine and free amino acids contents in Chinese tea using NIR spectroscopy. *LWT* 96:90–97
77. Correia RM, Tosato F, Domingos E, Rodrigues RRT, Aquino LFM, et al. 2018. Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta* 176:59–68
78. Teye E, Amuah CLY, McGrath T, Elliott C. 2019. Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics. *Spectrochim. Acta A* 217:147–54
79. Kartakoullis A, Comaposada J, Cruz-Carrión A, Serra X, Gou P. 2019. Feasibility study of smartphone-based near infrared spectroscopy (NIRS) for salted minced meat composition diagnostics at different temperatures. *Food Chem.* 278:314–21
80. Subedi PP, Walsh KB. 2020. Assessment of avocado fruit dry matter content using portable near infrared spectroscopy: method and instrumentation optimisation. *Postharvest Biol. Technol.* 161:111078

81. Papadopoulou OS, Panagou EZ, Mohareb FR, Nychas G-JE. 2013. Sensory and microbiological quality assessment of beef fillets using a portable electronic nose in tandem with support vector machine analysis. *Food Res. Int.* 50:241–49
82. Panagou EZ, Sahgal N, Magan N, Nychas GJE. 2008. Table olives volatile fingerprints: potential of an electronic nose for quality discrimination. *Sens. Actuators B* 134:902–7
83. Concina I, Falasconi M, Gobbi E, Bianchi F, Musci M, et al. 2009. Early detection of microbial contamination in processed tomatoes by electronic nose. *Food Control* 20:873–80
84. Pan L, Zhang W, Zhu N, Mao S, Tu K. 2014. Early detection and classification of pathogenic fungal disease in post-harvest strawberry fruit by electronic nose and gas chromatography–mass spectrometry. *Food Res. Int.* 62:162–68
85. Wilson A, Oberle C, Oberle D. 2013. Detection of off-flavor in catfish using a conducting polymer electronic-nose technology. *Sensors* 13:15968–84
86. Zhang H, Wang J, Ye S, Chang M. 2012. Application of electronic nose and statistical analysis to predict quality indices of peach. *Food Bioprocess Technol.* 5:65–72
87. Zakaria A, Shakaff AY, Masnan MJ, Saad FS, Adom AH, et al. 2012. Improved maturity and ripeness classifications of *Magnifera Indica* cv. Harumanis mangoes through sensor fusion of an electronic nose and acoustic sensor. *Sensors* 12:6023–48
88. Qiu S, Wang J, Gao L. 2014. Discrimination and characterization of strawberry juice based on electronic nose and tongue: comparison of different juice processing approaches by LDA, PLSR, RF, and SVM. *J. Agric. Food Chem.* 62:6426–34
89. Dong W, Zhao J, Hu R, Dong Y, Tan L. 2017. Differentiation of Chinese robusta coffees according to species, using a combined electronic nose and tongue, with the aid of chemometrics. *Food Chem.* 229:743–51
90. Tian X, Wang J, Ma Z, Li M, Wei Z. 2019. Combination of an E-nose and an E-tongue for adulteration detection of minced mutton mixed with pork. *J. Food Q.* 2019:10
91. Liu M, Wang M, Wang J, Li D. 2013. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar. *Sens. Actuators B* 177:970–80
92. Huang L, Zhao J, Chen Q, Zhang Y. 2014. Nondestructive measurement of total volatile basic nitrogen (TVB-N) in pork meat by integrating near infrared spectroscopy, computer vision and electronic nose techniques. *Food Chem.* 145:228–36
93. Estelles-Lopez L, Ropodi A, Pavlidis D, Fotopoulou J, Gkousari C, et al. 2017. An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling. *Food Res. Int.* 99:206–15
94. Ropodi AI, Panagou EZ, Nychas GE. 2018. Rapid detection of frozen-then-thawed minced beef using multispectral imaging and Fourier transform infrared spectroscopy. *Meat Sci.* 135:142–47
95. Xiong J, Lin R, Bu R, Liu Z, Yang Z, Yu L. 2018. A micro-damage detection method of litchi fruit using hyperspectral imaging technology. *Sensors* 18(3):700
96. Fengou LC, Spyrelli E, Lianou A, Tsakanikas P, Panagou EZ, Nychas GE. 2019. Estimation of minced pork microbiological spoilage through Fourier transform infrared and visible spectroscopy and multispectral vision technology. *Foods* 8(7):238
97. Manthou E, Lago S-L, Dagres E, Lianou A, Tsakanikas P, et al. 2020. Application of spectroscopic and multispectral imaging technologies on the assessment of ready-to-eat pineapple quality: a performance evaluation study of machine learning models generated from two commercial data analytics tools. *Comput. Electron. Agric.* 175:105529
98. Ropodi AI, Panagou EZ, Nychas GJE. 2016. Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science disciplines. *Trends Food Sci. Technol.* 50:11–25
99. Marini F. 2016. Data fusion strategies for food authentication. In *Proceedings of the 2nd IMEKO FOODS Conference*, pp. 238–42. Benevento, Italy: Univ. Sannio
100. Sabeur Z, Zlatev Z, Melas P, Veres G, Arbab-Zavar B, et al. 2017. Large scale surveillance, detection and alerts information management system for critical infrastructure. In *Environmental Software Systems:*

Computer Science for Environmental Protection, ed. J Hřebíček, R Denzer, G Schimak, T Pitner, pp. 237–46. Cham, Switz.: Springer
101. Food Life. 2016. *Food for tomorrow's consumer: strategic research and innovation agenda of the European Technology Platform Food for Life*. Tech. Rep., Food Life, Brussels. **http://etp.fooddrinkeurope.eu/news-and-publications/news/5-public-consultation-on-draft-etp-food-for-life-strategic-research-and-innovation-agenda.html**
102. den Besten HMW, Amezquita A, Bover-Cid S, Dagnas S, Ellouze M, et al. 2018. Next generation of microbiological risk assessment: potential of omics data for exposure assessment. *Int. J. Food Microbiol.* 287:18–27
103. Phelan JE, O'Sullivan DM, Machado D, Ramos J, Oppong YEA, et al. 2019. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med*. 11:41
104. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, et al. 2008. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar. J.* 7:223
105. Marvin HJP, Janssen EM, Bouzembrak Y, Hendriksen PJM, Staats M. 2017. Big data in food safety: an overview. *Crit. Rev. Food Sci. Nutr.* 57:2286–95
106. Yan B, Hu D, Shi P. 2012. A traceable platform of aquatic foods supply chain based on RFID and EPC Internet of Things. *Int. J. RF Technol.* 4:55–70
107. Mededjel M, Belalem G, Neki A. 2017. Towards a traceability system based on cloud and fog computing. *Multiagent Grid Syst*. 13:47–68
108. Chen R-Y. 2015. Autonomous tracing system for backward design in food supply chain. *Food Control* 51:70–84
109. Yu X, Liu P, Ren W, Zhang C, Wang J, Zheng Y. 2018. Safety traceability system of livestock and poultry industrial chain. In *Proceedings of the 4th International Conference on Cloud Computing and Security (ICCCS 2018)*, ed. X Sun, Z Pan, E Bertino, pp. 3–12. Cham, Switz.: Springer Int.
110. Gupta K, Rakesh N. 2018. IoT-based solution for food adulteration. In *Proceedings of First International Conference on Smart System, Innovations and Computing*, ed. AK Somani, S Srivastava, A Mundra, S Rawat, pp. 9–18. Singapore: Springer
111. Nirenjena S, Subramanian D, Monisha M. 2018. Advancement in monitoring the food supply chain management using IOT. *Int. J. Pure Appl. Math*. 119:1193–96
112. Shih C-W, Wang C-H. 2016. Integrating wireless sensor networks with statistical quality control to develop a cold chain system in food industries. *Comput. Standards Interfaces* 45:62–78
113. Bouzembrak Y, Klüche M, Gavai A, Marvin HJP. 2019. Internet of Things in food safety: literature review and a bibliometric analysis. *Trends Food Sci. Technol.* 94:54–64
114. Feng T. 2017. A supply chain traceability system for food safety based on HACCP, blockchain & Internet of things. In *Proceedings of the 2017 International Conference on Service Systems and Service Management*. New York: IEEE
115. Musa Z, Vidyasankar K. 2017. A fog computing framework for blackberry supply chain management. *Procedia Comput. Sci.* 113:178–85
116. Verdouw CN, Wolfert J, Beulens AJM, Rialland A. 2016. Virtualization of food supply chains with the internet of things. *J. Food Eng.* 176:128–36
117. Beyer MA, Laney D. 2012. *The importance of 'big data': a definition*. Tech. Rep. G00235055, Gartner Research, Stamford, CT
118. Deloitte China. 2015. *Big data and analytics in the automotive industry*. Tech. Rep., Deloitte China, Shanghai
119. Soc. Actuar. 2019. *The use of big data and data analytics to enhance insurer operations in Asia-Pacific*. Tech. Rep., Soc. Actuar., Schaumberg, Ill.
120. Pollard S, Namazi H, Khaksar R. 2019. Big data applications in food safety and quality. In *Encyclopedia of Food Chemistry*, ed. L Melton, F Shahidi, P Varelis, pp. 356–63. Oxford: Academic
121. Bhardwaj S, Kaushik M. 2018. Blockchain—technology to drive the future. In *Smart Computing and Informatics*, ed. SC Satapathy, V Bhateja, S Das, pp. 263–71. Singapore: Springer
122. Smith BG. 2008. Developing sustainable food supply chains. *Philos. Trans. R. Soc. B* 363:849–61

123. Eur. Parliam. Counc. 2002. Regulation (EC) no 178/2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. *Off. J.* L31:1–24. **http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32002R0178**
124. Olsen P, Borit M. 2018. The components of a food traceability system. *Trends Food Sci. Technol.* 77:143–49
125. Pincheira Caro MR, Ali MS, Vecchio M, Giaffreda R. 2018. Blockchain-based traceability in agri-food supply chain management: a practical implementation. In *Proceedings of 2018 IoT Vertical and Topical Summit on Agriculture (IoT Tuscany)*. New York: IEEE
126. Demestichas K, Peppes N, Alexakis T, Adamopoulou E. 2020. Blockchain in agriculture traceability systems: a review. *Appl. Sci.* 10:4113
127. Creydt M, Fischer M. 2019. Blockchain and more—algorithm driven food traceability. *Food Control* 105:45–51
128. WHO (World Health Organ.) Reg. Off. Eur. 2015. *More than 23 million people in the WHO European Region fall ill from unsafe food every year*. Web Resour., WHO Reg. Off. Eur., Copenhagen
129. WHO (World. Health Organ.). 2015. *WHO estimates of the global burden of foodborne diseases: foodborne disease burden epidemiology reference group 2007–2015*. Tech. Rep., WHO, Geneva
130. Antonucci F, Figorilli S, Costa C, Pallottino F, Raso L, Menesatti P. 2019. A review on blockchain applications in the agri-food sector. *J. Sci. Food Agric.* 99:6129–38
131. Kamilaris A, Fonts A, Prenafeta-Boldú FX. 2019. The rise of blockchain technology in agriculture and food supply chains. *Trends Food Sci. Technol.* 91:640–52
132. Galvez JF, Mejuto JC, Simal-Gandara J. 2018. Future challenges on the use of blockchain for food traceability analysis. *Trends Anal. Chem.* 107:222–32
133. TOMRA. *TOMRA food processing equipment: smart investments*. Web Resour., TOMRA, Leuven, Belg. **https://www.tomra.com/en/sorting/food**
134. Garbie IH. 2010. Enhancing the performance of industrial firms through implementation of lean techniques. In *Proceedings of the Institute of Industrial Engineers Annual Conference and Expo 2010*, Vol. 2, ed. A Johnson, J Miller, pp 1303–8. Red Hook, NY: Curran Assoc.
135. Kosior E, Mitchell J, Davies K, Kay M, Ahmad R, et al. 2017. Plastic packaging recycling using intelligent separation technologies for materials. In *Proceedings of the 2017 USENIX Annual Technical Conference*, pp. 500–6. Berkeley, CA: USENIX
136. Tsakanikas P, Karnavas A, Panagou EZ, Nychas G-J. 2020. A machine learning workflow for raw food spectroscopic classification in a future industry. *Sci. Rep.* 10:11212
137. Chacón Ramírez E, Albarrán JC, Cruz Salazar LA. 2020. The control of water distribution systems as a holonic system. In *Service Oriented, Holonic and Multi-agent Manufacturing Systems for Industry of the Future*, ed. T Borangiu, D Trentesaux, P Leitão, AG Boggino, V Botti, pp. 352–65. Cham, Switz.: Springer Int.
138. Zoellner C, Jennings R, Wiedmann M, Ivanek R. 2019. EnABLe: an agent-based model to understand *Listeria* dynamics in food processing facilities. *Sci. Rep.* 9:495
139. Horn AL, Friedrich H. 2019. Locating the source of large-scale outbreaks of foodborne disease. *J. R. Soc. Interface* 16:20180624
140. Sadilek A, Caty S, DiPrete L, Mansour R, Schenk T, et al. 2018. Machine-learned epidemiology: real-time detection of foodborne illness at scale. *NPJ Digital Med.* 1:36
141. Silva AFS, Rocha FRP. 2020. A novel approach to detect milk adulteration based on the determination of protein content by smartphone-based digital image colorimetry. *Food Control* 115:107299
142. Liu Z, Zhang Y, Xu S, Zhang H, Tan Y, et al. 2017. A 3D printed smartphone optosensing platform for point-of-need food safety inspection. *Anal. Chim. Acta* 966:81–89
143. Jung Y, Heo Y, Lee JJ, Deering A, Bae E. 2020. Smartphone-based lateral flow imaging system for detection of food-borne bacteria *E. coli* O157:H7. *J. Microbiol. Methods* 168:105800
144. Alfian G, Syafrudin M, Rhee J. 2017. Real-time monitoring system using smartphone-based sensors and NoSQL database for perishable supply chain. *Sustainability* 9:2073
145. Wang H, Xu Z, Fujita H, Liu S. 2016. Towards felicitous decision making: an overview on challenges and trends of big data. *Inf. Sci.* 367–68:747–65

146. Grieves M. 2015. *Digital twin: manufacturing excellence through virtual factory replication*. White Pap., Florida Inst. Technol., Melbourne, FL
147. ISO (Int. Org. Standard.). 2019. *ISO/DIS 23247-1: Automation systems and integration—digital twin manufacturing framework—part 1: overview and general principles*. Indust. Process Standard, ISO, Geneva
148. Siemens. 2020. *Unprecedented transparency with the digital twin of the supply chain*. Webinar, Siemens, Munich. **https://www.plm.automation.siemens.com/global/en/webinar/retail-digital-twin-supply-chain/57730**
149. EIT Food. 2018. *Digital twin management*. Web Resour., EIT Food, Leuven, Belg. **https://www.eitfood.eu/projects/digital-twin-management-2019**
150. Srai J, Settanni E, Tsolakis N, Aulakh P. 2019. *Supply Chain Digital Twins: Opportunities and Challenges Beyond the Hype*. Cambridge, UK: Univ. Cambridge. **https://doi.org/10.17863/CAM.45897**
151. Ahmed A, Zulfiqar S, Ghandar A, Chen Y, Hanai M, Theodoropoulos G. 2019. Digital twin technology for aquaponics: towards optimizing food production with dynamic data driven application systems. In *Methods and Applications for Modeling and Simulation of Complex Systems*, ed. G Tan, A Lehmann, YM Teo, W Cai, pp. 3–14. Singapore: Springer
152. Rapsomanikis G. 2015. *The economic lives of smallholder farmers*. Tech. Rep., Food Agric. Organ. U.N., Rome
153. Veltmeyer H. 2009. The World Bank on "agriculture for development": a failure of imagination or the power of ideology? *J. Peasant Stud.* 36(2):393–410
154. Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353:790–94
155. Burke M, Lobell DB. 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *PNAS* 114:2189–94
156. Eur. Comm. 2014. *RASFF—food and feed safety alerts*. Web Resour., Eur. Comm. Publ. Off., Luxembourg. **https://ec.europa.eu/food/safety/rasff_en**
157. FDA (US Food Drug Admin.). 2015. *Import refusal report*. Web Resour., Food Drug Admin., Silver Spring, MD. **https://www.accessdata.fda.gov/scripts/importrefusals/**
158. GACC (Gen. Admin. Cust. China). 2020. [Food information not allowed to enter]. Web Resour., Gen. Admin. Cust. China, Beijing (In Chinese) **http://jckspj.customs.gov.cn/spj/zwgk75/2706876/fe2490a9-1.html**
159. Pouillot R, Delignette-Muller ML. 2010. Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. *Int. J. Food Microbiol.* 142:330–40
160. Tenenhaus-Aziza F, Ellouze M. 2015. Software for predictive microbiology and risk assessment: a description and comparison of tools presented at the ICPMF8 Software Fair. *Food Microbiol.* 45:290–99
161. FAO (Food Agric. Organ.). 2021. *AGROVOC multilingual thesaurus*. Web. Resour., Agric. Inf. Manag. Standard. Portal, Food Agric. Organ. U.N., Rome. **http://www.fao.org/agrovoc/search**
162. EFSA (Eur. Food Safety Auth.). n.d. *Data standardisation*. Web Resour., EFSA, Parma, Italy. **https://www.efsa.europa.eu/en/data/data-standardisation**
163. Lehr H. n.d. *Electronic management and exchange of laboratory information*. PowerPoint Present. **https://www.unescap.org/sites/default/files/05%20-%20Electronic%20Management%20and%20Exchange%20of%20Laboratory%20Analysis%20Information%20V151210a.pdf**
164. Hasnan NZN, Yusoff YM. 2018. Short review: application areas of Industry 4.0 technologies in food processing sector. In *Proceedings of the 2018 IEEE Student Conference on Research and Development (SCOReD)*. New York: IEEE. **https://doi.org/10.1109/SCORED.2018.8711184**
165. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018
166. Jonquet C, Toulet A, Arnaud E, Aubin S, Dzalé Yeumo E, et al. 2018. AgroPortal: a vocabulary and ontology repository for agronomy. *Comput. Electron. Agric.* 144:126–43
167. Shekhar S, Colletti J, Muñoz-Arriola F, Ramaswamy L, Krintz C, et al. 2017. Intelligent infrastructure for smart agriculture: an integrated food, energy and water system. arXiv:1705.01993 [cs.CY]
168. Hasanzadeh K, Kajosaari A, Häggman D, Kyttä M. 2020. A context sensitive approach to anonymizing public participation GIS data: from development to the assessment of anonymization effects on data quality. *Comput. Environ. Urban Syst.* 83:101513

169. Oldroyd RA, Morris MA, Birkin M. 2018. Identifying methods for monitoring foodborne illness: review of existing public health surveillance techniques. *JMIR Public Health Surveill.* 4:e57
170. Ribeiro FDS, Caliva F, Swainson M, Gudmundsson K, Leontidis G, Kollias S. 2018. An adaptable deep learning system for optical character verification in retail food packaging. In *Proceedings of the 2018 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*. New York: IEEE. **https://doi.org/10.1109/EAIS.2018.8397178**

Annual Review of Biomedical Data Science

Volume 4, 2021

Contents

Errata

An online log of corrections to *Annual Review of Biomedical Data Science* articles may be found at http://www.annualreviews.org/errata/biodatasci