Current Research in Food Science 8 (2024) 100679

Contents lists available at ScienceDirect

Current Research in Food Science

journal homepage: www.sciencedirect.com/journal/current-research-in-food-science

Review Article

Machine learning assisted biosensing technology: An emerging powerful tool for improving the intelligence of food safety detection

Zixuan Zhou [a,1], Daoming Tian [a,b,1], Yingao Yang [a], Han Cui [a,c], Yanchun Li [a], Shuyue Ren [a,*], Tie Han [a,**], Zhixian Gao [a,***]

[a] Tianjin Key Laboratory of Risk Assessment and Control Technology for Environment and Food Safety, Tianjin Institute of Environmental and Operational Medicine, Tianjin, 300050, China
[b] Beidaihe Rest and Recuperation Center of PLA, Qinhuangdao, 066000, China
[c] State Key Laboratory of Food Nutrition and Safety, Tianjin University of Science & Technology, Tianjin, 300457, China

ARTICLE INFO

Keywords:
Deep learning
Biosensor
Food safety
Classic algorithms
Monitoring

ABSTRACT

Recently, the application of biosensors in food safety assessment has gained considerable research attention. Nevertheless, the evaluation of biosensors' sensitivity, accuracy, and efficiency is still ongoing. The advent of machine learning has enhanced the application of biosensors in food security assessment, yielding improved results. Machine learning has been preliminarily applied in combination with different biosensors in food safety assessment, with positive results. This review offers a comprehensive summary of the diverse machine learning methods employed in biosensors for food safety. Initially, the primary machine learning methods were outlined, and the integrated application of biosensors and machine learning in food safety was thoroughly examined. Lastly, the challenges and limitations of machine learning and biosensors in the realm of food safety were underscored, and potential solutions were explored. The review's findings demonstrated that algorithms grounded in machine learning can aid in the early detection of food safety issues. Furthermore, preliminary research suggests that biosensors could be optimized through machine learning for real-time, multifaceted analyses of food safety variables and their interactions. The potential of machine learning and biosensors in real-time monitoring of food quality has been discussed.

1. Introduction

Recently, significant advances have occurred in the field of artificial intelligence and machine learning, impacting various facets of human activities such as healthcare, agriculture, weather forecasting, and food safety. An enormous amount of data directly or indirectly linked to the topic of food safety has been generated worldwide (Marvin et al., 2017). The US Food and Drug Administration (FDA) presented "Steps to bring the US into a new era of smarter food safety," incorporating the application of machine learning and artificial intelligence in the field of food safety (Sharpless and Yiannas, 2019). Machine learning is a data-driven pattern recognition approach, which aims to identify discriminative or generative models from a given dataset, utilizing statistical associations between features. The acquisition model can predict outputs, such as category labels, cluster categories, and continuous real values. Currently, there are several food safety detection approaches based on algorithms, such as hyperspectral imaging, which integrates spatial and spectral operations to provide valuable information on food characteristics in a nondestructive manner. However, hyperspectral images may contain irrelevant data (Jia et al., 2020). The combination of machine learning and biosensors could be a powerful tool for continuous learning in food safety monitoring and assessment. Machine learning possess huge potential in several aspects of food safety, including food adulteration detection, food quality prediction, and foodborne disease warning (Deng et al., 2021) (see Fig. 1).

Noteworthiness, there has been an increased awareness of the role of food safety in human health and well-being, which has facilitated the development of relevant technologies to control and assess food safety.

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail address: renshuyue2018@163.com (S. Ren).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.crfs.2024.100679
Received 22 November 2023; Received in revised form 10 January 2024; Accepted 11 January 2024
Available online 12 January 2024

Traditional analytical techniques based on large laboratory instruments, such as liquid chromatographs and atomic absorption spectrometers, can yield accurate results with excellent reproducibility. However, they are limited by a lengthy pre-experiment process, complex preparation, and technical operation. Biosensors, based on biomolecular recognition and downstream signal conversion, have the potential for food safety monitoring and assessment. Biosensors have been widely applied in several fields, including food safety, environmental monitoring, and disease diagnosis and treatment. Compared with traditional analytical methods, biosensors are more sensitive, simplified, fast, reliable, and efficient for food safety assessment and monitoring, with lesser demand for reagents. Biosensor technology possess the potential for real-time detection of food quality, which can facilitate the screening of suspicious samples. Currently, biosensors are widely employed in detecting agricultural and veterinary drug residues, illegal additives, foodborne pathogens, biotoxins, and other food contaminants (Jia et al., 2021) (see Fig. 2) .

In this review, we explore the role and potential applications of machine learning and biosensors in food safety. Firstly, the types of machine learning methods were highlighted, and the application of machine learning in food safety was discussed. Additionally, the combined application of biosensors in food safety was examined, and the limitations and possible solutions were discussed. Moreover, future trends in biosensor and machine learning research, along with potential applications in food safety, were highlighted. Overall, we believe that this review could serve as a reference for the application of biosensors and machine learning in food safety.

2. Machine learning

Machine learning, a fundamental component of artificial intelligence, enables algorithms and software applications to predict outcomes from provided data without explicit programming. Machine learning concentrates on extracting meaningful patterns from large datasets (referred to as "learning") through computational means. Subsequently, it applies this acquired knowledge to make predictions by processing additional data.

2.1. Classification

Machine learning can be classified in various ways based on different criteria. Based on the learning system, machine learning can be

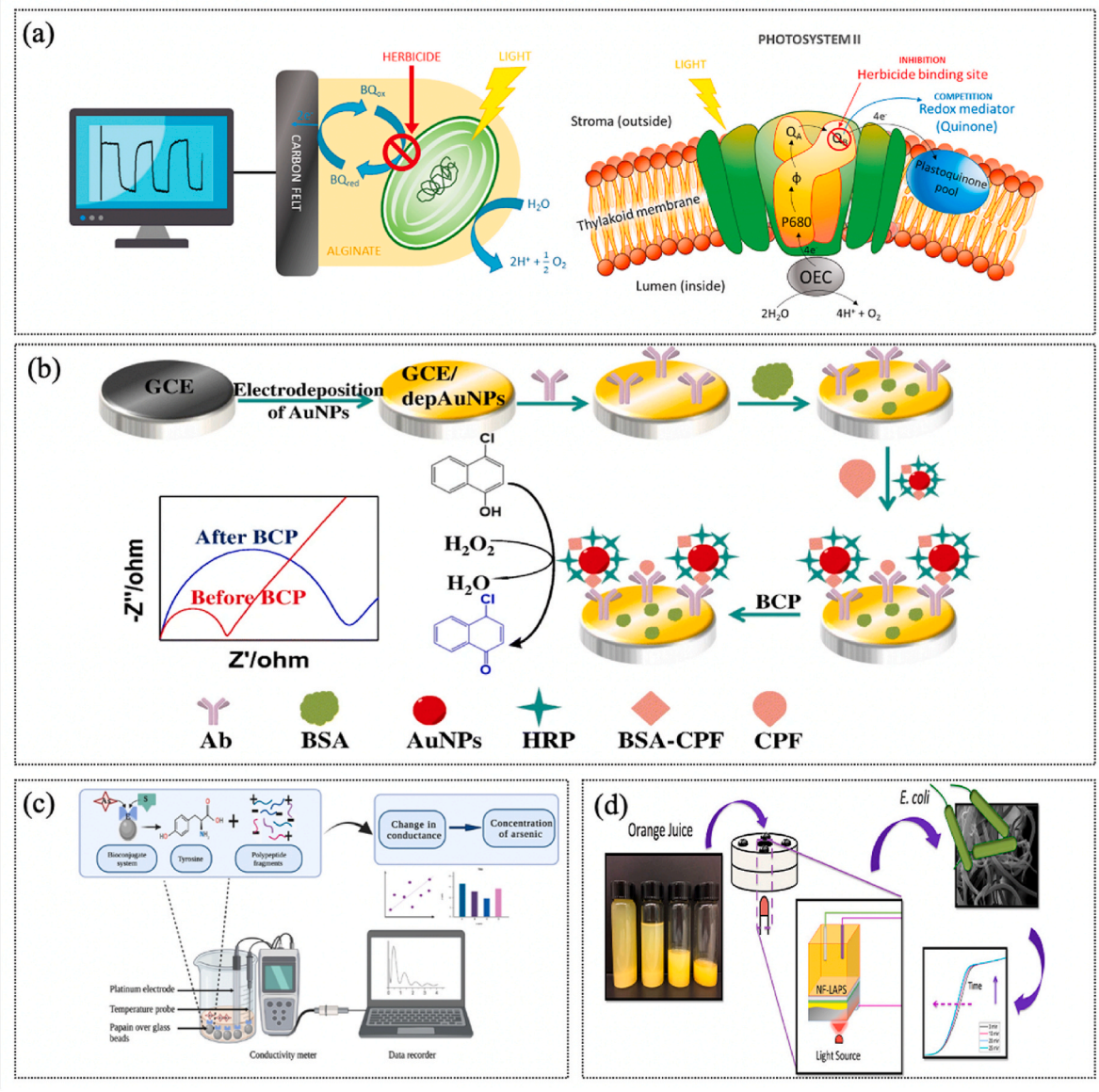

Fig. 1. Types of biosensors (I).

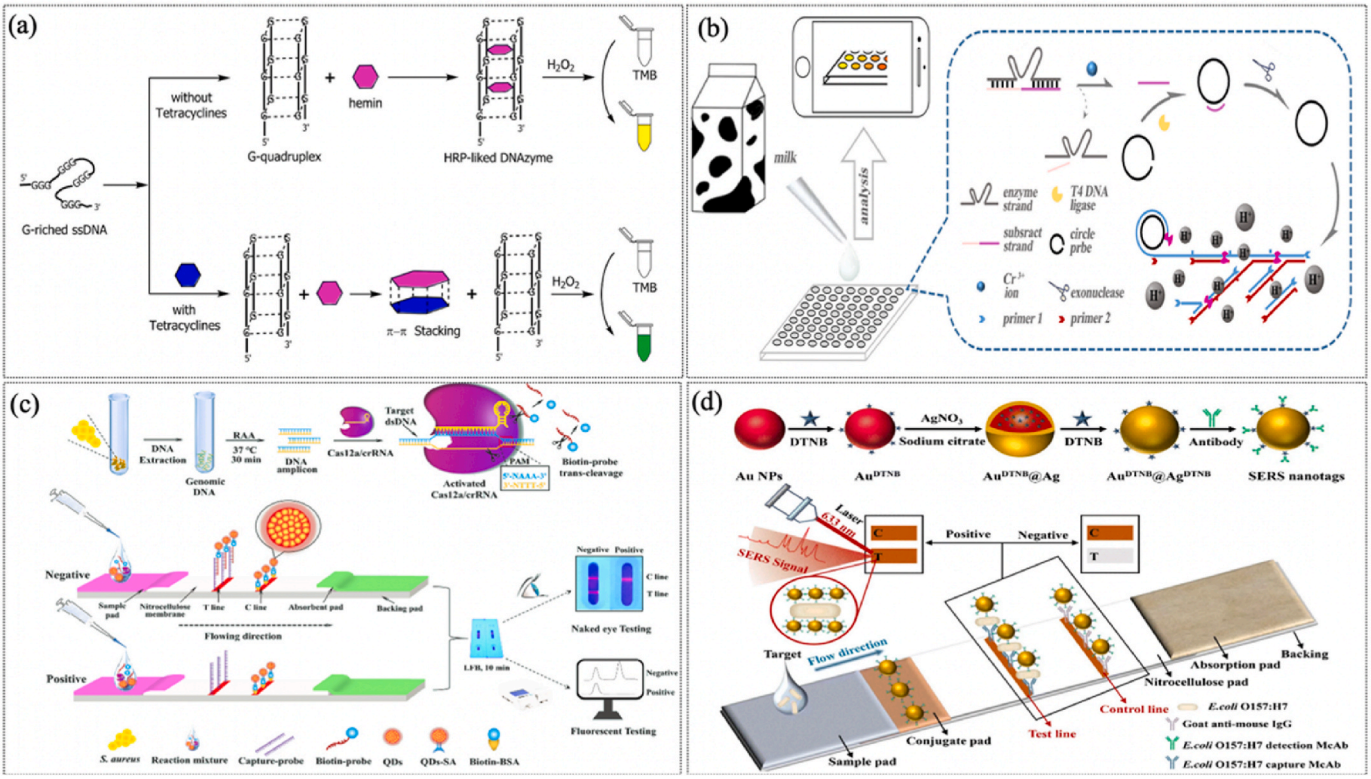

Fig. 2. Types of biosensors (II).

classified into four groups: unsupervised learning, supervised learning, semi-supervised learning, and reinforcement learning. In unsupervised learning, the labeled information about the training samples is unknown, and the purpose is to expose the inherent attributes and rules of the data via acquiring the unlabeled training samples. Unsupervised learning has been used extensively in organizing computing clusters, social network analysis, market segmentation, and astronomical data analysis.

However, supervised machine learning is the predominant approach in the field of natural science (Lai et al., 2018). In a given dataset, every piece of data and corresponding output values are known, establishing a specific relationship between the inputs and outputs (Lu, 2010). Computer algorithms can make predictions based on this specific relationship. The generated model can establish its own relationships through continuous learning and training on a known sample dataset, subsequently making predictions on a new dataset (Lu, 2010). Supervised learning contains regression and classification. When the target variable that we are trying to predict is a continuous variable, this algorithm is called regression. However, when target variable is predicted to be discrete values, this algorithm is called classification.

Semi-supervised learning combines aspects of both supervised and unsupervised learning, utilizing datasets that can be labeled or unlabeled. There are two common semi-supervised learning methods, including transductive learning and inductive learning. Reinforcement learning extracts information from the external environment rather than a dataset (Sutton and Barto, 1998), which is used for decision making. The goal is to learn the mapping from the environmental state to the behavior, then the selected behavior can obtain the maximum reward from the environment. The reward in reinforcement learning comes from the feedback in the environment.

2.2. Executive process

Machine learning executes four main tasks: classification, regression, clustering, and dimensionality reduction. Classification is a predictive problem where a dataset is assigned to discrete classes, while a regression problem arises when the target value is continuous. Additionally, clustering involves the spatial distribution and visualization of a dataset, and samples belonging to the same category are determined by comparing the distances between different samples. Dimensionality reduction is a technique for decreasing the dimensionality of variables to lower dimensional subspace (Cox, 1959).

Classification and regression are both part of supervised learning, which are used to develop prediction models to generated outputs based on any given input. However, the difference between them is that the target is discrete in classification and continuous in regression. Generally, the algorithms for both are similar, and the same learning algorithm can be used for classification and regression. The most frequently used learning algorithms for classification include support vector machine (SVM), stochastic gradient descent algorithm (SGD), Bayesian estimation, Ensemble, and K Nearest Neighbors. Moreover, SVM and SGD can also be used for regression.

Clustering also analyses the attributes of samples, which is similar to classification. The difference is that classification has already labeled specific target before prediction, whereas there are no labeled targets prior to clustering (Zubaroğlu and Atalay, 2021). Additionally, classification is a supervised learning technique, while clustering is an unsupervised learning technique. Clustering can only determine the attribute of a sample according to the distribution of the sample in space. The commonly used algorithms for clustering include K-means, Gaussian mixture model, and expectation-maximization.

Dimensionality reduction is another important field of machine learning, with several important applications in both supervised learning and unsupervised learning. Dimensionality reduction involves removing redundancies to represent data in lower-dimensional subspaces. The primary dimension reduction algorithm is principal component analysis (PCA), with some other algorithms, such as partial least squares regression and linear discriminant analysis (LDA), evolving from PCA.

2.3. Workflow

Machine learning process consists of several steps (Fig. 3). The initial step involves the preparation and acquisition of data to construct input for subsequent learning, serving as a determinant for the built model. This step is important because learning algorithms require large amount of data. The next step involved the development of the model using the training set, followed by the identification of the most appropriate algorithm and the validation of the model. Thereafter, the performance of the validated model should be tested using test data, and subsequently deployed to make predictions using new data. Finally, the model should be tuned to improve the performance of the algorithm, using more data, different features, or adjusted parameters. Additionally, the system should undergo evaluation for accuracy.

3. Main machine learning algorithms in food safety

3.1. Classic algorithms

Traditional and classical algorithms can deal with the problem of classification, clustering, and regression. However, most traditional and classical algorithms are not capable of handling complex-structured data, with multiple layers (Wang et al., 2021a).

K-means clustering (K-means) is an unsupervised clustering algorithm that automatically classify similar samples into categories (Fig. 4a). SVMs are a set of supervised learning methods used for classification, regression, and outlier detection (Fig. 4b) (Boser et al., 1992). Decision tree is a prediction model, which represents a mapping relationship between object attributes and values (Fig. 4c). (Breiman, 1996). The artificial neural network is a graphical computing model that mimics the functions of the human brain, where each neuron is interconnected to transmit information (Fig. 4d) (Kruse et al., 2016). Naïve Bayes adopts the "Attribute Conditional Independence Assumption," assuming that all attributes are independent of each other for known categories (Fig. 4e) (Devroye et al., 2014). The principle of K nearest neighbor is that the category of unmarked samples is determined by majority voting based on the class of k nearest neighbor training samples or the mean value of label values (Fig. 4f).

Suitable algorithms can significantly enhance the detection efficiency and accuracy of biosensors. For example, SVM can be used to analyzed nonlinear and high-dimensional data. Recently, a nano biosensor has been developed for the detection of four widely used antibiotics in the field of veterinary medicine, and SVM effectively determined the concentration of the antibiotics by analyzing acquired absorption spectrum (Gutiérrez et al., 2020). Moreover, various algorithms can be combined for food safety monitoring. For instance, smartphone-based lateral flow assay was applied to distinguish ambiguous concentrations of *Salmonella* spp. By analyzing test line images, using SVM and KNN (Min et al., 2021).

Considering the additional advantages of SVM, it could also serve as a supporting method for data processing, which can expand the field of data processing. SVM could be applied in e-nose, which is a reliable instrument for inspecting the quality of food and agricultural products (Infante et al., 2008; Jiang and Wang, 2016; Xu et al., 2019). E-nose consists of an array of sensors and pattern recognition algorithms to probe odors, which is similar to the human olfactory system (Wang et al., 2021b). Apart from e-nose, KNN, naïve Bayes (NB), LDA, and adaptive resonance theory map (ARTMAP) have also been applied (Jha et al., 2019).

Electrochemical impedance spectroscopy (EIS) is used to detect pathogens in samples based on changes occurring at the electrode-solution. However, this strategy is not suitable for all kinds of EIS sensors and can lead to the inaccurate detection. Xu et al. (2020) developed a machine learning-based EIS biosensor for improved detection of *E. coli*. The model was programmed to automatically establish a quantitative relationship between bacterial concentration and multiple impedimetric parameters, using SVM and PCA.

Decision trees offer better interpretability but struggle to generalize data in overly complex trees; however, this limitation can be overcome by the random forest (RF) algorithm. RF is an ensemble learning algorithm, which involves generating multiple models, such as classifiers, to solve specific problems. RF algorithm constructs the decision tree

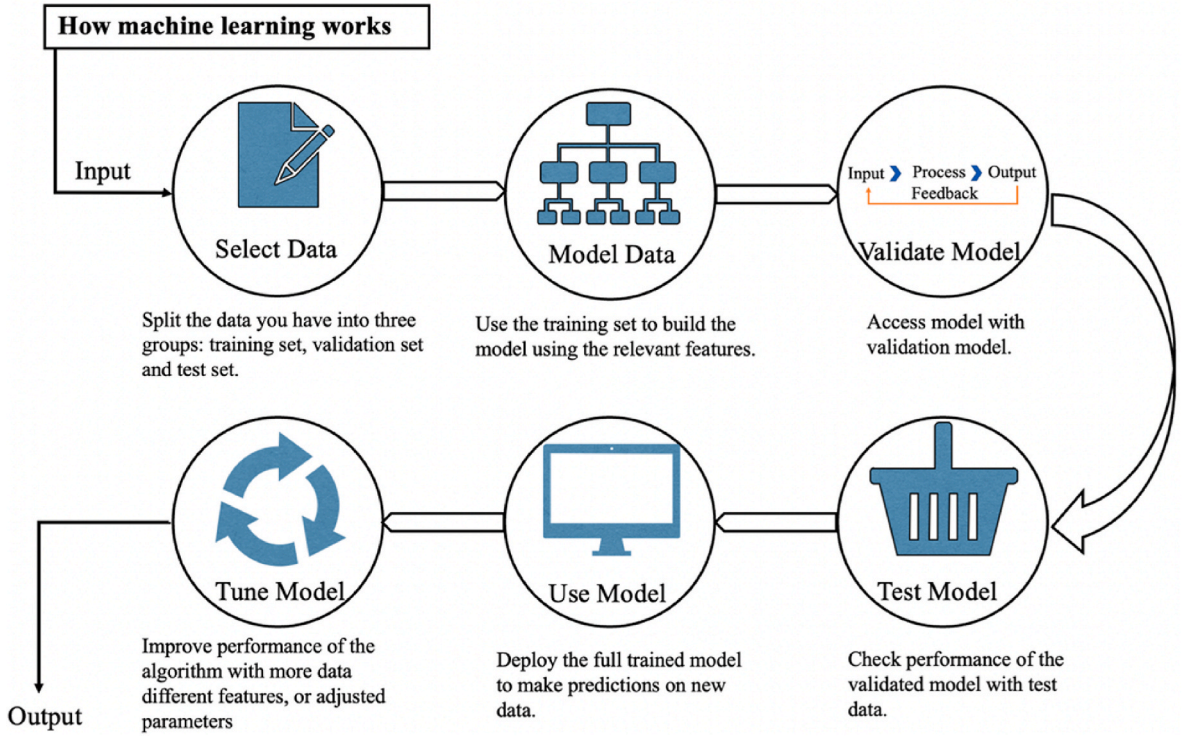

Fig. 3. The workflow of machine learning in processing data.

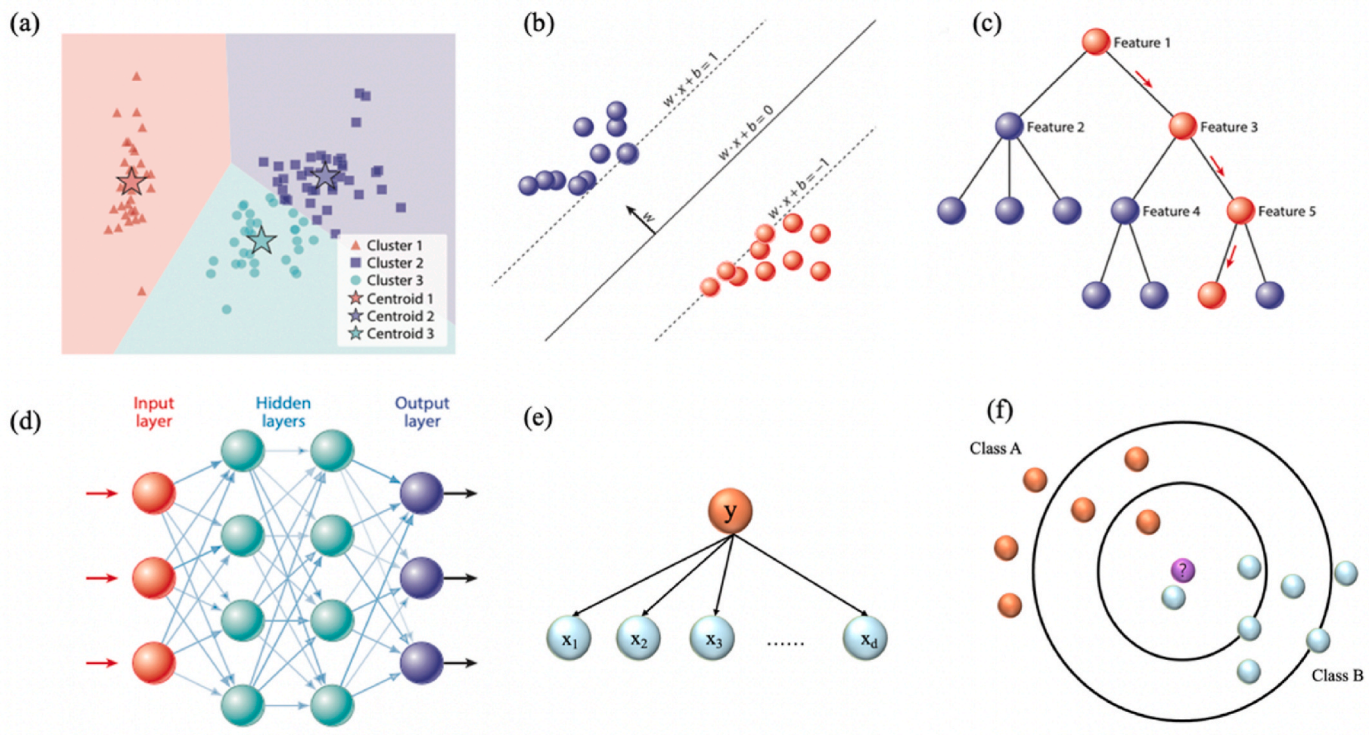

Fig. 4. Examples of machine learning models.

prediction set by randomly selecting a subset from the training data and aggregates the prediction results to reduce the variance. RF algorithm has been validated using complex and nonlinear data. Moreover, the training process of RF is rapid, with a low degree of overfitting (Alexander et al., 2014). Moreover, several studies have confirmed the robustness of RF and laser-induced breakdown spectroscopy techniques (Gazeli et al., 2020) or infrared spectroscopy analysis (Gazeli et al., 2020) for food classification.

KNN has been applied in the processing of hyperspectral imaging, including Vis-NIR (Khanal et al., 2021), fluorescence hyperspectral imaging (Lee et al., 2021), and mechanically-flexible electrical impedance tomography (Darma and Takei, 2021). Interestingly, KNN can facilitate the detection of *E. coli* and *Salmonella typhimurium* on the surface of food processing facilities by analyzing fluorescence hyperspectral imaging (Lee et al., 2021). Additionally, KNN has been performed in combination with other algorithms to achieve classification of samples or data. For instance, Schroeder et al. (2019)developed a robust array of 20 carbon nanotube-based chemical sensors with K nearest neighbor model and RF model for the classification of multi-class time series, which was efficient. The protocol successfully classified five independent test sets of cheese and wine samples.

3.2. Dimensionality reduction methods

In various research and application domains, substantial amounts of data with multiple variables are collected for analysis, informing decision-making processes. However, there is a need to reduce the dimensionality of these data and eliminate unnecessary data for efficient management. Dimensionality reduction entails reducing the variables requiring analysis while minimizing the loss of information from the original variables. Dimensionality reduction can consolidate closely related variables into fewer, unrelated variables. This allows for the use of fewer comprehensive indicators to represent various information within each variable. The most common methods for dimensionality reduction include PCA and LDA, and their differences are illustrated in Fig. 5.

PCA is applied to simplify multiple variables by reducing the dimension of high-dimensional variables while minimizing data loss. For instance, PCA was applied to SERS results to obtain characteristic variables, and SVM was employed to classify duck meat into four categories based on sulfapyridine and sulfadiazine concentrations (Ren et al., 2021). LDA is a supervised linear dimensionality reduction algorithm (Fulkerson, 1995). The central idea of LDA is to maximize interclass spacing and minimize intra-class distance. For instance, LDA has been used in combination with SVM to distinguish different antibiotics

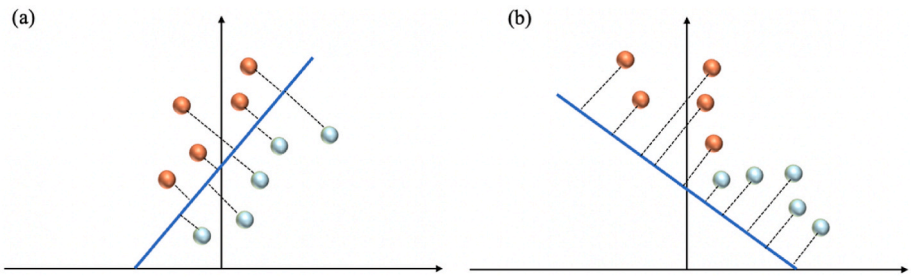

Fig. 5. Differences between principal component analysis (PCA) and linear discriminant analysis (LDA).

(Xu et al., 2020). Additionally, LDA can be used as a powerful tool for classification, particularly for data related to spectrum and e-nose signals. Employing chemometric approaches, feature vectors were extracted from a sensor array and utilized as inputs for PCA to detect formalin, hydrogen peroxide, and sodium hypochlorite in raw milk. However, LDA exhibited relatively low classification accuracy in this scenario (Tohidi et al., 2018).

3.3. Deep learning

In comparison to traditional methods, deep learning (DL) enables enhanced insights into complex data features at high levels of abstraction. Additionally, beyond the algorithms mentioned earlier, studies on DL have developed computational models with multiple processing layers to systematically analyze data from original inputs (Lecun et al., 2015). DL is based on artificial neural networks, and it performs better than other algorithms because of its deep architecture (Shorten et al., 2021).

Feedforward neural networks (FFNNs) are perhaps the simplest DL model, consisting of input layer, hidden layer, and output layer, with no cycle or loop inside the structure (Zhang et al., 2021). Additionally, FFNN is a useful tool for the analysis of data collected using colorimetric biosensors. Thankfully, FFNNs can address challenges, such as complex VOCs background signals and the intricate behavior of bacteria in the field of food safety. Additionally, an advanced deep feedforward neural network (DFFNN) with a learning rate scheduler, L^2 regularization, and shortcut connections have been developed (Jia et al., 2021). After training on the ΔR/ΔG/ΔB database, the network demonstrated excellent performance in identifying pathogens in single monocultures, multiple monocultures, and in cocktail culture, and was effective in distinguishing the pathogens from the background signal on cantaloupe, with accuracy of up to 93% and 91% under ambient and refrigerated conditions, respectively (Fig. 6a).

Apart from FFNN, two types of DL architectures are gaining considerable attention due to their applications in computer vision: CNN and recurrent neural networks (RNN). These algorithms are mostly applied in autonomous driving vehicles and medical care (Alawadi et al., 2019), and in resolving image processing problems (Russ, 2016). RNN has proven to be suitable for time series forecasting due to their ability to capture sequence or time series data (Kaviani and Sohn, 2021). The roles of RNN in environmental factor forecasting have been extensively discussed (Chen et al., 2018); however, food safety-related studies on RNN are limited. Therefore, this review focused on CNN due to its powerful ability to solve food safety-related issues.

Recently, the potentials of biosensors combined with CNN has gained considerable research interest. CNN plays a crucial part in image technology, biomedical technology, and industrial production because of its local receptive fields, weight sharing, pooling, and sparse connections (Patrício and Rieder, 2018). CNN consists of four layers: convolutional, pooling, active, and fully connected layers. The convolutional layer executes the transvection of input image (different data) and the filter matrix (a set of fixed weights) to enhance the original image and suppress noise interference. The process of convolution will automatically learn the features without manual selection of features, and it reflects the characteristics of local receptive field and weight sharing (Tian, 2020). The pooling layer executes the calculation of the average or maximum image region, greatly reduce the calculations without losing the main features of the image. The active layer is necessary to add a nonlinear relation to the result calculated above for most complex problems. The commonly used active functions include sigmoid function, Tanh function, ReLU function, and leaky ReLU function. The fully connected layer is similar to the traditional neuron network, and connects neurons in each layer.

Hu et al. (2021) developed a fluorescent biosensor to identify fluorescent bacteria, such as *S. typhimurium*, in milk based on DL via the faster R–CNN algorithm, and its minimum limit of detection was as low as 55 CFU/mL. Future studies should attempt focus on developing highly automated, accurate, sensitive, and rapid means of detection of pathogens, using portable fluorescence microscope equipment and more advanced DL algorithms.

As we know, the main challenge of colorimetric biosensor is the relatively low sensitivity. Interestingly, the sensitivity of colorimetric biosensors could be improved by constructing algorithm-reinforced biosensors for food safety monitoring and assessment. Utilizing polyacrylonitrile (PAN) and thin layer silica gel (SG), with p-aminophenylcyclic acid (SA) and naphthalene ethylenediamine hydrochloride as carriers and chromogenic agents, a nitrite color sensor named PAN-NSS was proposed. A combination of PAN-NSS, deep convolutional neural network (DCNN), and APP provides an efficient, highly sensitive, and fully integrated detection system for field detection (Guo et al., 2021).

While image analysis and decomposition can be performed with only a few accessible samples, insufficient training is insufficient to resolve more complex problems. Guo et al. (2020) developed a portable food-freshness prediction platform based on cross-reactive colorimetric barcode combinatorics and DCNNs for monitoring meat freshness, achieving an accuracy of 98.5% (Fig. 6b).

To deal with the problem of insufficient datasets, generative models, such as generative adversarial networks (GANs), can be used to generate data from the scratch (Creswell et al., 2018; Goodfellow et al., 2014; Guo et al., 2021). Yang et al. (2022) proposed a WGAN-ResNet method, which combines two DL networks, the Wasserstein generative adversarial network (WGAN) and the residual neural network (ResNet), to detect carbendazim based on terahertz spectroscopy. The WGAN and pretraining model technology were employed to solve the problem of insufficient learning samples for training the ResNet (Fig. 7). The WGAN was used for generating new datasets, while the pretraining model technology was applied to reduce the training parameters to avoid residual neural network overfitting. Overall, the results demonstrated that our proposed method achieves a 91.4% accuracy rate, which is better than those of SVM, KNN, NB, and ensemble learning.

The combination of biosensors and machine learning provides a new strategy for food safety. The combination of machine learning and biosensors supports the rapid and highly sensitive detection of food contaminants (Table 1). Despite some practical challenges, biosensors combined with machine learning could become a general trend in the field of food safety in the future.

4. Challenges and future trends

Biosensor technology serves as a potent alternative to traditional laboratory methods for food analysis, offering potential applications in monitoring food bioprocesses. For example, the e-nose can be tailored for detecting various parameters, replacing the multi-instrumentation employed in laboratories (Nayak et al., 2020). Furthermore, enhancements in biosensor stability and reproducibility are imperative to meet the stringent demands of the food bioprocess industry. Moreover, future research should concentrate on enhancing biosensors for trace detection and extracting additional features from complex samples. Collection and analysis of food signals can facilitate the detection of hazard factors and promote the optimization of food processing (Lv et al., 2018). Nevertheless, the integration of machine learning into biosensing can address certain limitations mentioned earlier (Jiménez-Sanchidrián and Ruiz, 2016). While machine learning has been successfully employed in biosensing, its potential in spectrum-based biosensors remains underexplored. The precision of machine learning depends largely on how quickly and accurately test data can be obtained, especially as the number of substances to be tested in food continues to increase. Besides choosing the most suitable algorithm and well-defined inputs and outputs, machine learning requires high-quality training data to achieve accurate prediction results (Baker et al., 2018).

The success of machine learning is dependent on the availability of

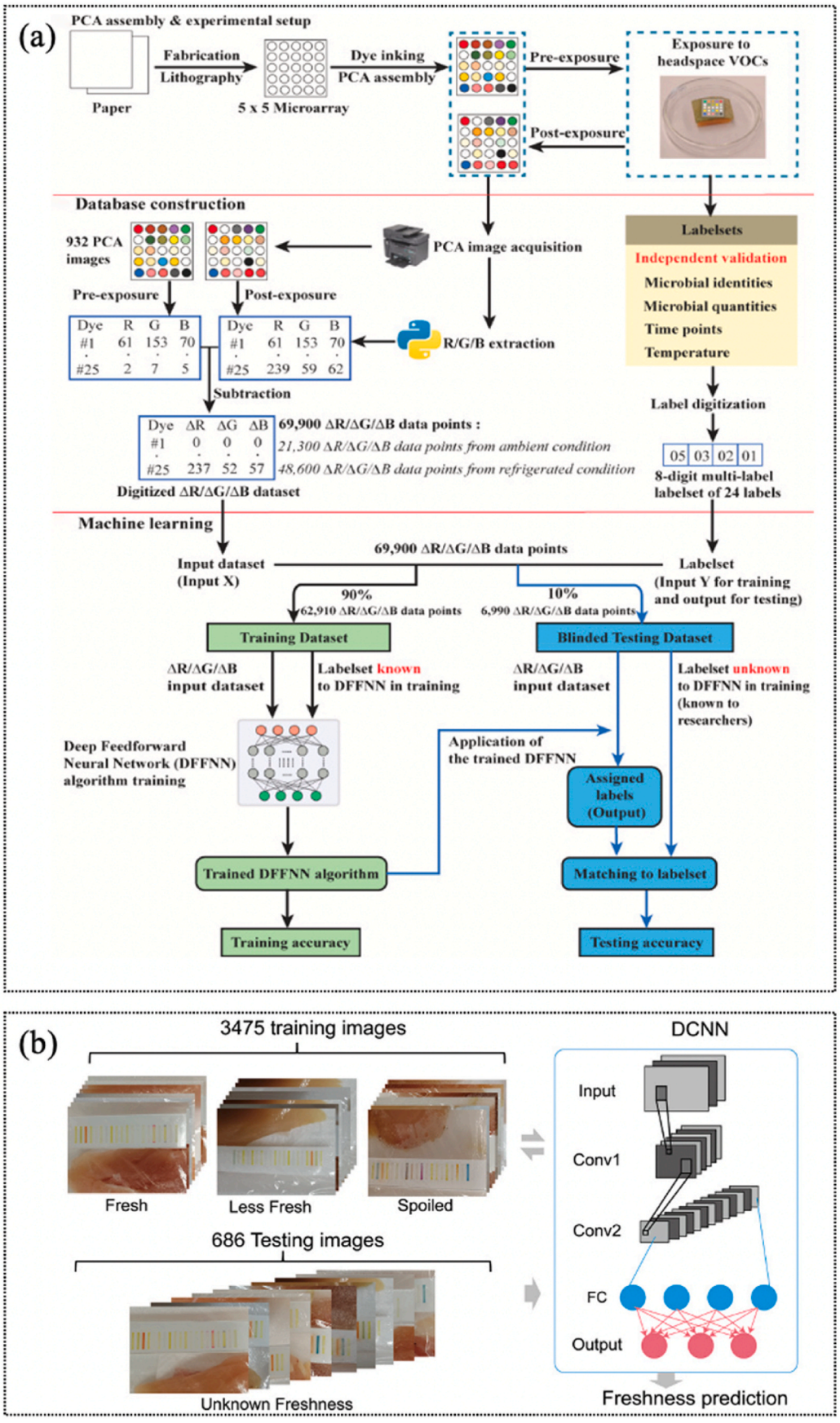

Fig. 6. Application of Feedforward neural networks (FFNNs).

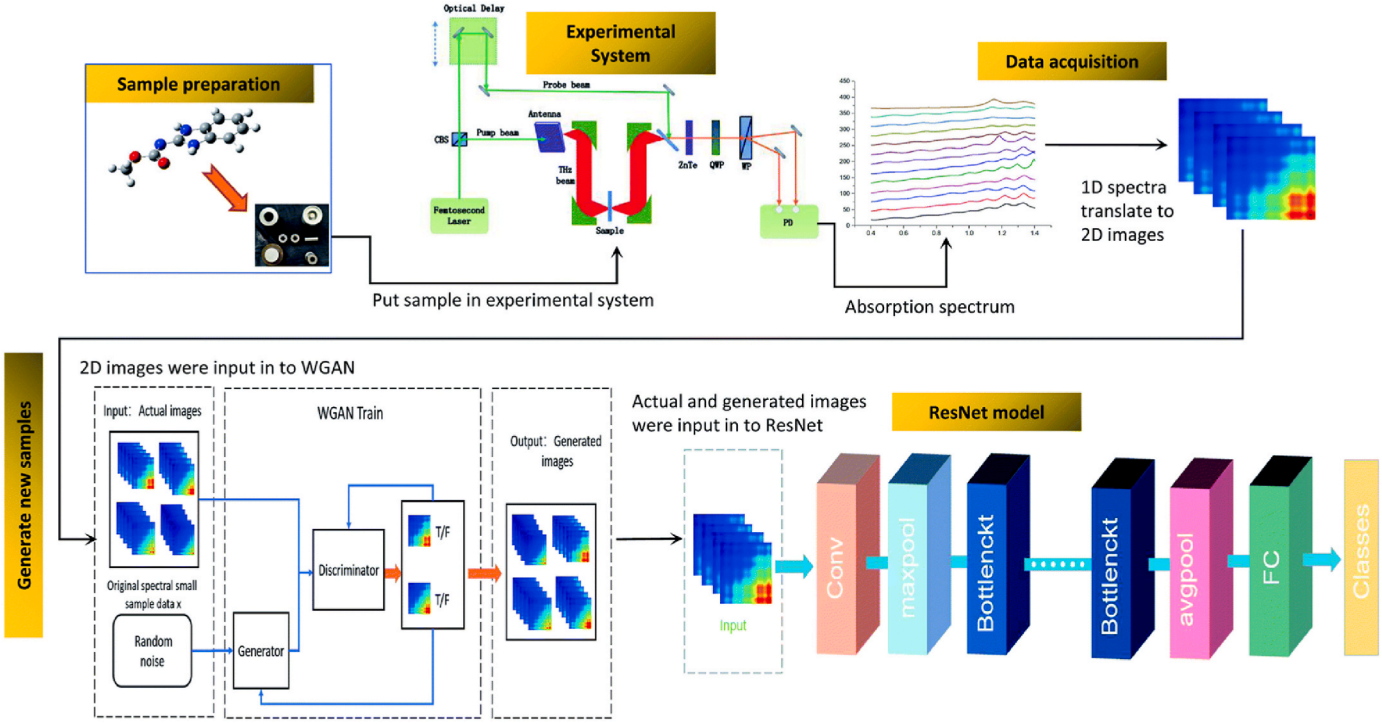

Fig. 7. Flow chart of the detection of carbendazim using WGAN-ResNet based on THz spectroscopy.

Table 1
Application of machine learning in food detection.

target analytes	mechanism	algorithm	application	ref.
Kanamycin, Ampicillin, Oxytetracycline and Sulfadimethoxine	optical	SVM	detection of antibiotics in the milk	Gutiérrez et al. (2020)
Salmonella spp.	optical	SVM, KNN	detection Salmonella spp. In raw meat, egg products, and milk	Min et al. (2021)
doxycycline (DOX), tetracycline, oxytetracydine (OTC), and metacydine (MTC)	optical	SVM, LDA	detection and identification of tetracyclines in river water and milk	Xu et al. (2020)
indigo	optical	RF	determine indigo in cream	Zhang et al. (2020)
honey adulteration	optical	RF	detection of honey adulteration	Calle et al. (2023)
aflatoxin	optical	RF	detection of aflatoxin-polluted corn kernels	Cheng and Stasiewicz (2021)
α-naphthalene acetic acid (NAA)	electrochemical	ANN	detection of α-naphthalene acetic acid (NAA) residues in food	Zhu et al. (2021a,b)
aflatoxin B1 and fumonisins	electrochemical	ANN	aflatoxin B1 and fumonisins in maize	Leggieri et al. (2021)
benzoic acid	electrochemical	ANN	benzoic acid in cola-type carbonated beverages	Yang et al. (2021)
pesticide residue	optical	SVM, RF, ANN	determination of pesticide residue in food	Khanal et al. (2021)
xanthine (XT) and hypoxanthine (HX)	electrochemical	ANN	determination of XT and HX in fish	Zhu et al. (2021c)
food odor and microbial population	electrochemical	k-NN, LDA, SVM/SVR, MLP	beef quality monitoring	Wijaya et al. (2018)
detection of the sulfapyridine and sulfadimidine	optical	PCA, SVM	detection of the sulfapyridine and sulfadimidine remained in duck meat	Ning et al. (2020)
food odor	optical	DCNN	assessment of food freshness	Guo et al. (2020)
Salmonella typhimurium	optical	CNN	detection of *Salmonella typhimurium* in milk	Hu et al. (2021)
identifying pathogens	optical	DFFNN	identifying pathogens	Jia et al. (2021)

large amount of training data. Machine learning may struggle with small datasets, particularly for high-dimensional datasets; the model tends to "remember" each data point as a unique case rather than "learning" from the data, resulting in overfitting and inefficient training (von Rueden et al., 2021). Additionally, with the proliferation of application-specific analytical algorithms, there is an urgent need for rigorous evaluation and benchmarking of data (Quainoo et al., 2017).

Besides exploring and enhancing new classification algorithms, efforts should be directed towards improvements in data management and collection (Farrell et al., 2018). Additionally, both organic and inorganic compounds have been utilized in the fabrication of biosensors for assessing and monitoring food safety. Notably, it is important to collect the spectral data of each food substance and continually update databases of spectral data of food materials for a more accurate and robust forecasting. Additionally, data storage and management is important to facilitate the global application of machine learning and biosensing, especially data ownership and privacy. Therefore, further research is necessary to develop rules regarding data usage and privacy.

Presently, machine learning has been applied in food safety monitoring and assessment. Machine learning could improve the tracking of diverse analytes and detection of interactions in complex biological environments by improving the diversity and functionality of small sensing devices (Bertani et al., 2020).

Furthermore, organic and inorganic compounds have been used in

the development of biosensors for food safety assessment and monitoring. However, the mechanisms of synthesizing these compounds are yet to be fully explained, with simulation and data-driven methods employed as alternatives to the experimental trial-and-error. Machine learning models can effectively predict the conditions for the formation of new organic or inorganic products. Machine learning provides a new perspective for the designing of next generation of materials (Hou et al., 2018); moreover, machine learning could be applied for the monitoring of clinical diagnosis and the environment.

5. Conclusion

In conclusion, the amalgamation of machine learning with biosensors has been successfully employed in the realm of food safety, yielding significant results. Hence, the integration of machine learning into other biosensor detection methods holds a promising future. In the long run, it is possible to envision a combination of machine learning and biosensors technologies in the large-scale and end-to-end predictive modeling systems.

Financial interest

All authors declare that they have no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Consent for publication

All the authors read and agreed to publish this article.

CRediT authorship contribution statement

Zixuan Zhou: Conceptualization, Writing – original draft, Writing – review & editing. **Daoming Tian:** Conceptualization, Writing – original draft, Writing – review & editing. **Yingao Yang:** Conceptualization, Writing – review & editing. **Han Cui:** Conceptualization, Writing – review & editing. **Yanchun Li:** Conceptualization, Writing – original draft. **Shuyue Ren:** Supervision. **Tie Han:** Supervision. **Zhixian Gao:** Supervision, All authors have read and agreed to the published version of the Manuscript.

Declaration of competing interest

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

Data availability

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

Acknowledgements

The authors thank the Research Institute Fund for Young Talents of AMMS (AMMS-QNPY-2021-007), National Natural Science Foundation of China (82003465) and (82173573), Tianjin Institute of Environmental Medicine and Occupational Medicine independent research project (ZZKY2021206) for funding this research project.

References

Alawadi, S., Fernández-Delgado, M., Mera, D., Barro, S., 2019. Polynomial kernel discriminant analysis for 2D visualization of classification problems. Neural Comput. Appl. 31, 3515–3531. https://doi.org/10.1007/s00521-017-3290-3.

Alexander, D.C., Zikic, D., Zhang, J., Zhang, H., Criminisi, A., 2014. Image quality transfer via random forest regression: applications in diffusion MRI. Med. Image Comput. Comput. Assist. Interv. 17, 225–232. https://doi.org/10.1007/978-3-319-10443-0_29.

Baker, R.E., Peña, J.M., Jayamohan, J., Jérusalem, A., 2018. Mechanistic models versus machine learning, a fight worth fighting for the biological community? Biol. Lett. 14, 20170660 https://doi.org/10.1098/rsbl.2017.0660.

Bertani, F.R., Businaro, L., Gambacorta, L., Mencattini, A., Brenda, D., di Giuseppe, D., de Ninno, A., Solfrizzo, M., Martinelli, E., Gerardino, A., 2020. Optical detection of aflatoxins B in grained almonds using fluorescence spectroscopy and machine learning algorithms. Food Control 112. https://doi.org/10.1016/j.foodcont.2019.107073.

Boser, B.E., Guyon, I.M., Vladimir, V.N., 1992. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory 144–152. https://doi.org/10.1145/130385.130401.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1007/bf00058655.

Calle, J.L.P., Punta-Sánchez, I., González-de-Peredo, A.V., Ruiz-Rodríguez, A., Ferreiro-González, M., Palma, M., 2023. Rapid and automated method for detecting and quantifying adulterations in high-quality honey using vis-NIRs in combination with machine learning. Foods (Basel, Switzerland) 12 (13), 2491. https://doi.org/10.3390/foods12132491.

Chen, Y., Cheng, Q., Cheng, Y., Yang, H., Yu, H., 2018. Applications of recurrent neural networks in environmental factor forecasting: a review. Neural Comput. 30, 2855–2881. https://doi.org/10.1162/neco_a_01134.

Cheng, X., Stasiewicz, M.J., 2021. Evaluation of the Impact of Skewness, Clustering, and Probe Sampling Plan on Aflatoxin Detection in Corn. *Risk analysis : an official publication of the Society for Risk*. Analysis 41 (11), 2065–2080. https://doi.org/10.1111/risa.13721.

Cox, D.R., 1959. The regression analysis of binary sequences. J. Roy. Stat. Soc. 20, 215–232. https://doi.org/10.1111/j.2517-6161.1959.tb00334.x.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A.A., 2018. Generative adversarial networks: an overview. IEEE Signal Process. Mag. 35, 53–65. https://doi.org/10.1109/MSP.2017.2765202.

Darma, P.N., Takei, M., 2021. High-speed and accurate meat composition imaging by mechanically-flexible electrical impedance tomography with k-nearest neighbor and fuzzy k-means machine learning approaches. IEEE Access 9, 1–10. https://doi.org/10.1109/ACCESS.2021.3064315.

Deng, X., Cao, S., Horn, A.L., 2021. Emerging applications of machine learning in food safety. Annu. Rev. Food Sci. Technol. 12, 22. https://doi.org/10.1146/annurev-food-071720-024112, 1-22.26.

Devroye, L., Gyorfi, L., Lugosi, G., 2014. A Probabilistic Theory of Pattern Recognition. Igarss, 2014.

Farrell, F., Soyer, O.S., Quince, C., 2018. Machine Learning Based Prediction of Functional Capabilities in Metagenomically Assembled Microbial Genomes. https://doi.org/10.1101/307157 bioRxiv.

Fulkerson, B., 1995. Machine learning, neural and statistical cassification. Technometrics 37, 459. https://doi.org/10.2307/1269742.

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

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems. https://doi.org/10.3156/jsoft.29.5_177_2.

Guo, L., Wang, T., Wu, Z., Wang, J., Wang, M., Cui, Z., Ji, S., Cai, J., Xu, C., Chen, X., 2020. Portable food-freshness prediction platform based on colorimetric barcode combinatorics and deep convolutional neural networks. Adv. Mater. 32, 2004805 https://doi.org/10.1002/adma.202004805.

Guo, K., Yang, Z., Yu, C.H., Buehler, M.J., 2021. Artificial intelligence and machine learning in design of mechanical materials. Mater. Horiz. 8, 1153–1172. https://doi.org/10.1039/d0mh01451f.

Gutiérrez, P., Godoy, S.E., Torres, S., Oyarzún, P., Sanhueza, I., Díaz-García, V., Contreras-Trigo, B., Coelho, P., 2020. Improved antibiotic detection in raw milk using machine learning tools over the absorption spectra of a problem-specific nanobiosensor. Sensors 20, 4552. https://doi.org/10.3390/s20164552.

Hou, X., Mu, L., Chen, F., Hu, X., 2018. Emerging investigator series: design of hydrogel nanocomposites for the detection and removal of pollutants: from nanosheets, network structures, and biocompatibility to machine-learning-assisted design. Environ. Sci.: Nano 5, 2216–2240. https://doi.org/10.1039/c8en00552d.

Hu, Q., Wang, S., Duan, H., Liu, Y., 2021. A fluorescent biosensor for sensitive detection of salmonella typhimurium using low-gradient magnetic field and deep learning via faster region-based convolutional neural network. Biosensors 11, 447. https://doi.org/10.3390/bios11110447.

Infante, R., Farcuh, M., Meneses, C., 2008. Monitoring the sensorial quality and aroma through an electronic nose in peaches during cold storage. J. Sci. Food Agric. 88 https://doi.org/10.1002/jsfa.3316.

Jha, S.K., Yadava, R.D.S., Hayashi, K., Patel, N., 2019. Recognition and sensing of organic compounds using analytical methods, chemical sensors, and pattern recognition approaches. Chemometr. Intell. Lab. Syst. 185, 18–31. https://doi.org/10.1016/j.chemolab.2018.12.008.

Jia, B., Wang, W., Ni, X., Lawrence, K.C., Zhuang, H., Yoon, S.C., Gao, Z., 2020. Essential processing methods of hyperspectral images of agricultural and food products. Chemometr. Intell. Lab. Syst. 198, 103936 https://doi.org/10.1016/j.chemolab.2020.103936.

Jia, X.X., Li, S., Han, D.P., Chen, R. peng, Yao, Z.Y., Ning, B. an, Gao, Z.X., Fan, Z.C., 2021. Development and perspectives of rapid detection technology in food and environment. Crit. Rev. Food Sci. Nutr. 1–20 https://doi.org/10.1080/10408398.2021.1878101.

Jia, Z., Luo, Y., Wang, D., Dinh, Q.N., Lin, S., Sharma, A., Block, E.M., Yang, M., Gu, T., Pearlstein, A.J., Yu, H., Zhang, B., 2021. Nondestructive multiplex detection of foodborne pathogens with background microflora and symbiosis using a paper chromogenic array and advanced neural network. Biosens. Bioelectron. 183, 113209 https://doi.org/10.1016/j.bios.2021.113209.

Jiang, S., Wang, J., 2016. Internal quality detection of Chinese pecans (Carya cathayensis) during storage using electronic nose responses combined with physicochemical methods. Postharvest Biol. Technol. 118 https://doi.org/10.1016/j.postharvbio.2016.03.016.

Jiménez-Sanchidrián, C., Ruiz, J.R., 2016. Use of Raman spectroscopy for analyzing edible vegetable oils. Appl. Spectrosc. Rev. 51, 417–430. https://doi.org/10.1080/05704928.2016.1141292.

Kaviani, S., Sohn, I., 2021. Study of scale-free structures in feed-forward neural networks against backdoor attacks. ICT Express 7, 265–268. https://doi.org/10.1016/j.icte.2020.11.004.

Khanal, B., Pokhrel, P., Khanal, B., Giri, B., 2021. Machine-learning-assisted analysis of colorimetric assays on paper analytical devices. ACS Omega 6, 33837–33845. https://doi.org/10.1021/acsomega.1c05086.

Kruse, R., Borgelt, C., Braune, C., Mostaghim, S., Steinbrecher, M., 2016. Computational Intelligence: A Methodological Introduction, second ed. Springer. https://doi.org/10.1007/978-1-4471-7296-3.

Lai, K., Twine, N., O'Brien, A., Guo, Y., Bauer, D., 2018. Artificial intelligence and machine learning in bioinformatics. ABC Bioinfo 1, 272–286. https://doi.org/10.1016/B978-0-12-809633-8.20325-7.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lee, A., Park, S., Yoo, J., Kang, J., Lim, J., Seo, Y., Kim, B., Kim, G., 2021. Detecting bacterial biofilms using fluorescence hyperspectral imaging and various discriminant analyses. Sensors 21, 2213. https://doi.org/10.3390/s21062213.

Leggieri, C.M., Mazzoni, M., Fodil, S., Moschini, M., Bertuzzi, T., Prandini, A., Battilani, P., 2021. An electronic nose supported by an artificial neural network for the rapid detection of aflatoxin B1 and fumonisins in maize. Food Control 123, 107722. https://doi.org/10.1016/j.foodcont.2020.107722.

Lu, Z.Q.J., 2010. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Journal of the Royal Statistical Society Series a-Statistics in Society.

Lv, M., Liu, Y., Geng, J., Kou, X., Xin, Z., Yang, D., 2018. Engineering nanomaterials-based biosensors for food safety detection. Biosens. Bioelectron. 106, 112–128. https://doi.org/10.1016/j.bios.2018.01.049.

Marvin, H.J.P., Janssen, E.M., Bouzembrak, Y., Hendriksen, P.J.M., Staats, M., 2017. Big data in food safety: an overview. Crit. Rev. Food Sci. Nutr. 57, 2286–2295. https://doi.org/10.1080/10408398.2016.1257481.

Min, H.J., Mina, H.A., Deering, A.J., Bae, E., 2021. Development of a smartphone-based lateral-flow imaging system using machine-learning classifiers for detection of Salmonella spp. J. Microbiol. Methods 188, 106288. https://doi.org/10.1016/j.mimet.2021.106288.

Nayak, J., Vakula, K., Dinesh, P., Naik, B., Pelusi, D., 2020. Intelligent food processing: journey from artificial neural network to deep learning. Comput. Sci. Rev. 38, 100297 https://doi.org/10.1016/j.cosrev.2020.100297.

Ning, X., Mu-Hua, L., Hai-Chao, Y., Shuang-Gen, H., Xiao, W., Jin-Hui, Z., Jian, C., Ting, W., Wei, H., Yi-Xin, S., 2020. Classification of sulfadimidine and sulfapyridine in duck meat by surface enhanced Raman spectroscopy combined with principal component analysis and support vector machine. Anal. Lett. 53, 1513–1524. https://doi.org/10.1080/00032719.2019.1710524.

Patrício, D.I., Rieder, R., 2018. Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review. Comput. Electron. Agric. 153, 69–81. https://doi.org/10.1016/j.compag.2018.08.001.

Quainoo, S., Coolen, J.P.M., van Hijum, S.A.F.T., Huynen, M.A., Melchers, W.J.G., van Schaik, W., Wertheim, H.F.L., 2017. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin. Microbiol. Rev. 30, 1015–1063. https://doi.org/10.1128/CMR.00016-17.

Ren, S., Li, Q., Jingyi, W., Fan, B., Bai, J., Peng, Y., Li, S., Han, D., Wu, J., Wang, J., Qin, K., Zhao, Z., Ning, B., Han, T., Zhou, H., Gao, Z., 2021. Development of a fast and ultrasensitive black phosphorus-based colorimetric/photothermal dual-readout immunochromatography for determination of norfloxacin in tap water and river water. J. Hazard Mater. 402, 123781 https://doi.org/10.1016/j.jhazmat.2020.123781.

Russ, J.C., 2016. The Image Processing Handbook. CRC Press. https://doi.org/10.1201/b10720.

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

Sharpless, N., Yiannas, F., 2019. Statement from acting FDA commissioner Ned sharpless. M.D., and Deputy Commissioner Frank Yiannas on Steps to Usher the U.S. Into a New Era of Smarter Food Safety. FDA statement.

Shorten, C., Khoshgoftaar, T.M., Furht, B., 2021. Deep learning applications for COVID-19. J. Big Data 8, 18. https://doi.org/10.1186/s40537-020-00392-9.

Sutton, R.S., Barto, A.G., 1998. Reinforcement learning: an introduction. IEEE Trans. Neural Network. 9, 1054. https://doi.org/10.1109/tnn.1998.712192, 1054.

Tian, Y., 2020. Artificial intelligence image recognition method based on convolutional neural network algorithm. IEEE Access 8, 125731–125744. https://doi.org/10.1109/ACCESS.2020.3006097.

Tohidi, M., Ghasemi-Varnamkhasti, M., Ghafarinia, V., Bonyadian, M., Mohtasebi, S.S., 2018. Development of a metal oxide semiconductor-based artificial nose as a fast, reliable and non-expensive analytical technique for aroma profiling of milk adulteration. Int. Dairy J. 77, 38–46. https://doi.org/10.1016/j.idairyj.2017.09.003.

von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Walczak, M., Pfrommer, J., Pick, A., Ramamurthy, R., Garcke, J., Bauckhage, C., Schuecker, J., 2021. Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Trans. Knowl. Data Eng. 1. https://doi.org/10.1109/tkde.2021.3079836, 1.

Wang, H., Wu, T., Li, M., Tao, Y., 2021a. Recent advances in nanomaterials for colorimetric cancer detection. J. Mater. Chem. B 9, 921–938. https://doi.org/10.1039/d0tb02163f.

Wang, P., Fan, E., Peng, W., 2021b. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern Recogn. Lett. 141, 61–67. https://doi.org/10.1016/j.patrec.2020.07.042.

Wijaya, D.R., Sarno, R., Zulaika, E., 2018. Electronic nose dataset for beef quality monitoring in uncontrolled ambient conditions. Data in brief 21, 2414–2420. https://doi.org/10.1016/j.dib.2018.11.091.

Xu, M., Wang, J., Gu, S., 2019. Rapid identification of tea quality by E-nose and computer vision combining with a synergetic data fusion strategy. J. Food Eng. 241 https://doi.org/10.1016/j.jfoodeng.2018.07.020.

Xu, Y., Li, C., Jiang, Y., Guo, M., Yang, Y., Yang, Y., Yu, H., 2020. Electrochemical impedance spectroscopic detection of *E.coli* with Machine Learning. J. Electrochem. Soc. 167, 047508 https://doi.org/10.1149/1945-7111/ab732f.

Xu, Z., Wang, Z., Liu, M., Yan, B., Ren, X., Gao, Z., 2020. Machine learning assisted dual-channel carbon quantum dots-based fluorescence sensor array for detection of tetracyclines. Spectrochim. Acta Mol. Biomol. Spectrosc. 232, 118147 https://doi.org/10.1016/j.saa.2020.118147.

Yang, M., Liu, X., Luo, Y., Pearlstein, A.J., Wang, S., Dillow, H., Reed, K., Jia, Z., Sharma, A., Zhou, B., Pearlstein, D., Yu, H., Zhang, B., 2021. Machine learning-enabled non-destructive paper chromogenic array detection of multiplexed viable pathogens on food. Nature Food 2, 110–117. https://doi.org/10.1038/s43016-021-00229-5.

Yang, R., Li, Y., Qin, B., Zhao, D., Gan, Y., Zheng, J., 2022. Pesticide detection combining the wasserstein generative adversarial network and the residual neural network based on terahertz spectroscopy. RSC Adv. 12, 1769–1776. https://doi.org/10.1039/D1RA06905E.

Zhang, S., Tan, Z., Liu, J., Xu, Z., Du, Z., 2020. Determination of the food dye indigotine in cream by near-infrared spectroscopy technology combined with random forest model. *Spectrochimica acta*. Part A, Molecular and biomolecular spectroscopy 227, 117551. https://doi.org/10.1016/j.saa.2019.117551.

Zhang, K., Wang, J., Liu, T., Luo, Y., Loh, X.J., Chen, X., 2021. Machine learning-reinforced noninvasive biosensors for healthcare. Adv. Healthcare Mater. 10, e2100734 https://doi.org/10.1002/adhm.202100734.

Zhu, D., Liu, B., Wei, G., 2021a. Two-dimensional material-based colorimetric biosensors: a review. Biosensors 11, 259. https://doi.org/10.3390/bios11080259.

Zhu, X., Lin, L., Wu, R., Zhu, Y., Sheng, Y., Nie, P., Liu, P., Xu, L., Wen, Y., 2021b. Portable wireless intelligent sensing of ultra-trace phytoregulator α-naphthalene acetic acid using self-assembled phosphorene/Ti3C2-MXene nanohybrid with high ambient stability on laser induced porous graphene as nanozyme flexible electrode. Biosens. Bioelectron. 179, 113062 https://doi.org/10.1016/j.bios.2021.113062.

Zhu, Y., Liu, P., Xue, T., Xu, J., Qiu, D., Sheng, Y., Li, W., Lu, X., Ge, Y., Wen, Y., 2021c. Facile and rapid one-step mass production of flexible 3D porous graphene nanozyme electrode via direct laser-writing for intelligent evaluation of fish freshness. Microchem. J. 162, 105855 https://doi.org/10.1016/j.microc.2020.105855.

Zubaroğlu, A., Atalay, V., 2021. Data stream clustering: a review. Artif. Intell. Rev. 54, 1201–1236. https://doi.org/10.1007/s10462-020-09874.