GC-MS in Food Authentication and Quality Control: A Survey

Abstract

Gas Chromatography-Mass Spectrometry (GC-MS) is a cornerstone analytical technique in food authentication, providing precise identification of complex mixtures to ensure food quality and safety. The integration of Artificial Intelligence (AI) and deep learning into GC-MS data analysis marks a transformative shift, enhancing the accuracy and efficiency of food composition verification and contaminant detection. This survey explores the synergy between GC-MS and AI, highlighting advancements such as Model-Based Discriminant Analysis with Variable Selection (MBDA-VS) and deep learning models like ChromAlignNet, which optimize peak alignment and improve classification accuracy. The modified QuEChERS method exemplifies innovations in sample preparation, enhancing analyte recovery and reducing matrix interferences. Challenges in GC-MS data analysis, including user-dependent biases and matrix effects, underscore the need for robust computational methods and standardized protocols. Future directions emphasize the integration of emerging technologies to enhance model generalization and reduce false positives, while collaborative efforts aim to establish standardized methods for reliable food quality assessments. This survey underscores the critical role of GC-MS and AI in reinforcing the integrity of the global food supply chain, advocating for continued research and development to address emerging challenges in food authentication and safety.

1 Introduction

1.1 Importance of Food Authentication and Quality Control

Food authentication and quality control are critical for consumer safety and market regulation, particularly in the face of increasing food fraud and adulteration risks due to global trade complexities. Ensuring food authenticity and quality is vital for maintaining consumer trust and safeguarding public health, especially with the rise of food contaminants. Advanced analytical techniques, such as gas chromatography-mass spectrometry (GC-MS/MS) and innovative sample preparation methods like QuEChERS, are instrumental in identifying and mitigating food safety risks, thereby enhancing consumer confidence and health outcomes [1, 2, 3]. Regulatory bodies depend on stringent quality control measures to uphold food safety standards and prevent the distribution of substandard products, with economic losses from food fraud estimated in the billions annually. Consequently, advanced analytical techniques, including GC-MS, are essential for verifying food authenticity and quality, supporting regulatory compliance, and bolstering consumer trust.

1.2 Role of GC-MS in Food Authentication

Gas Chromatography-Mass Spectrometry (GC-MS) is a pivotal analytical technique for food authentication, recognized for its accuracy in verifying food authenticity and quality. The integration of Gas Chromatography coupled with tandem Mass Spectrometry (GC-MS/MS) enhances sensitivity and separation of food contaminants, making it increasingly favored in food safety laboratories.

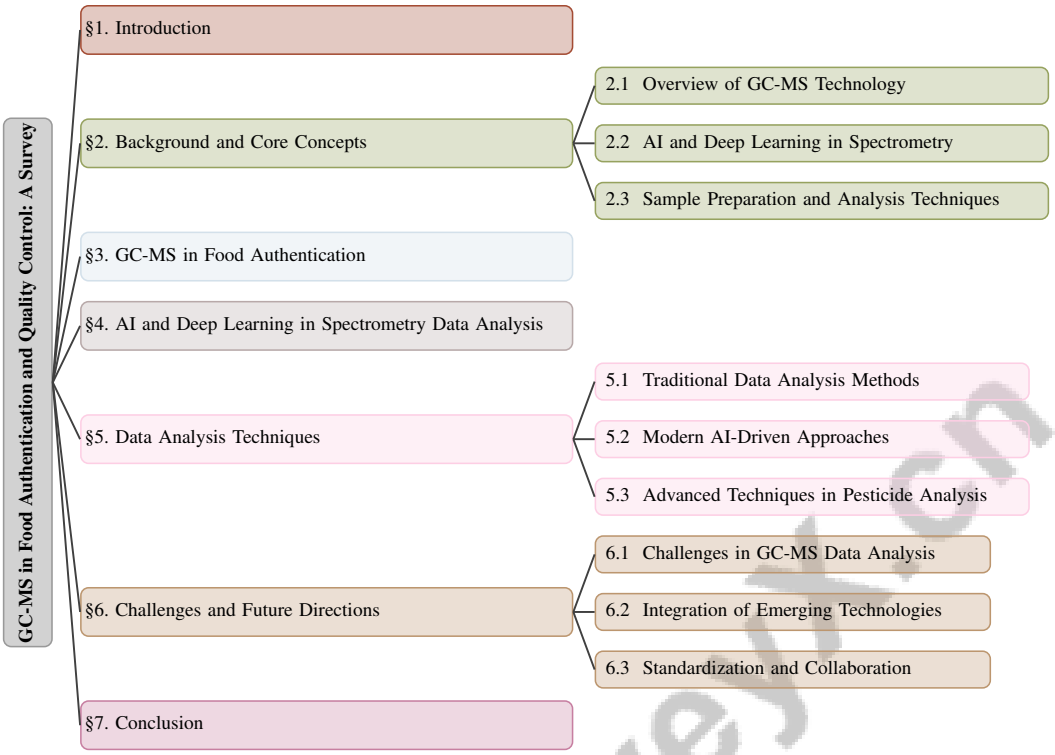


Figure 1: chapter structure

This effectiveness is further supported by advanced sample preparation techniques, such as QuEChERS and adsorption extraction, facilitating simultaneous extraction of multiple target analytes. As microorganic contaminants become more prevalent, these sophisticated approaches are vital for maintaining food safety and quality control [4, 1]. GC-MS combines the separation capabilities of gas chromatography with the detection power of mass spectrometry, enabling precise identification of complex mixtures and trace compounds in food matrices. Its specificity and sensitivity allow for the detection of adulterants and contaminants, such as pesticides and residues, which may compromise food safety. Moreover, GC-MS verifies geographical origin and varietal authenticity by analyzing distinct chemical profiles, enhancing food safety and quality control. The routine application of GC-MS not only ensures compliance with regulatory standards but also reinforces consumer trust by confirming the genuineness of food products.

1.3 Integration of AI and Deep Learning

The integration of Artificial Intelligence (AI) and deep learning into Gas Chromatography-Mass Spectrometry (GC-MS) data analysis signifies a transformative advancement in food authentication and quality control. These technologies enhance GC-MS's analytical capabilities by improving data representation and model optimization, which are essential for accurate analysis of complex spectrometric data [5]. AI algorithms, particularly those leveraging deep learning, excel in managing the high-dimensional data generated by GC-MS, enabling the extraction of meaningful patterns indicative of food authenticity and quality. This application facilitates the identification of subtle variations in chemical compositions critical for distinguishing authentic from adulterated products. Furthermore, AI-driven models automate data processing and interpretation, significantly reducing the time and labor associated with traditional methods. By harnessing AI and deep learning, researchers can achieve greater accuracy in identifying contaminants and verifying food composition, thereby enhancing the reliability of GC-MS as a tool for food quality assurance. This integration not only aids in meeting stringent regulatory standards but also boosts consumer confidence in the safety and authenticity of food products.

1.4 Structure of the Survey

This survey is structured to comprehensively explore the role of Gas Chromatography-Mass Spectrometry (GC-MS) in food authentication and quality control, alongside the integration of Artificial Intelligence (AI) and deep learning techniques. It begins with an introduction that highlights the significance of food authentication and quality control, emphasizing the pivotal role of GC-MS and the transformative impact of AI and deep learning on spectrometry data analysis. Following this, the survey provides background on GC-MS technology, AI applications in spectrometry, and the intricacies of sample preparation and analysis techniques.

Subsequent sections focus on the application of GC-MS in food authentication, detailing its effectiveness in detecting contaminants, verifying food composition, and its advantages over alternative methods. The exploration of AI and deep learning in spectrometry data analysis includes discussions on automated region selection, model-based discriminant analysis, and deep learning applications for peak alignment.

The survey further contrasts traditional data analysis methods with modern AI-driven approaches and addresses advanced techniques in pesticide analysis. The penultimate section identifies challenges and future directions in utilizing GC-MS and AI for food authentication, discussing data analysis issues, the integration of emerging technologies, and the imperative for standardization and collaboration.

The conclusion synthesizes the critical roles of GC-MS and AI technologies in enhancing food authenticity and quality assurance, emphasizing the necessity of advanced sample preparation techniques, such as multi-residue analytical methods like QuEChERS and adsorption extraction, as effective tools for contaminant detection. It also identifies potential research avenues for improving analytical methods and addressing challenges posed by emerging food contaminants, reinforcing the ongoing need for innovation in food safety and quality control [4, 1].The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 Overview of GC-MS Technology

Gas Chromatography-Mass Spectrometry (GC-MS) is a fundamental analytical technique that combines gas chromatography and mass spectrometry to identify various substances within a sample. The process initiates with the separation of volatile compounds through gas chromatography, where the sample is volatilized and traverses a column with an inert carrier gas, typically helium or nitrogen. This separation is based on the compounds' volatility and interactions with the column's stationary phase [4]. Subsequently, the mass spectrometer ionizes, fragments, and detects the components according to their mass-to-charge ratios, producing a mass spectrum that uniquely identifies and quantifies the compounds present. This dual capability makes GC-MS a robust tool for analyzing complex mixtures, particularly in food authentication, where identifying trace compounds is crucial [4].

The efficacy of GC-MS is enhanced by advanced analytical methods that address challenges related to molecular fragmentation and data interpretation. Online metabolomics platforms improve preprocessing, statistical analyses, and data annotation, increasing result reliability and accuracy [6]. Standardized benchmarks for cross-linking experimental results facilitate comparison and evaluation across various analytical methods, advancing the field [7]. Sample preparation techniques, such as QuEChERS (Quick, Easy, Cheap, Effective, Rugged, and Safe) and solid-phase microextraction (SPME), are vital for optimizing GC-MS data analysis. These methods enhance sample extraction and purification efficiency, ensuring that analytes of interest are adequately represented in the final analysis [1]. Through its comprehensive capabilities, GC-MS remains indispensable in food authentication and quality control, delivering accurate results that support regulatory compliance and consumer safety.

2.2 AI and Deep Learning in Spectrometry

The integration of Artificial Intelligence (AI) and deep learning into spectrometry has revolutionized the analysis of complex datasets, especially in Gas Chromatography-Mass Spectrometry (GC-MS). These computational methods enhance the analytical process by efficiently extracting and interpreting

meaningful patterns from high-dimensional data, crucial for precise food authentication and quality control. Deep learning models effectively address challenges posed by sparse datasets, enabling robust data analysis even in the presence of limited or noisy data [5]. Incorporating variable selection into model-based discriminant analysis frameworks enhances classification accuracy in food authenticity assessments, improving precision by identifying the most relevant features from spectrometry data [3]. The xlmod ontology provides a comprehensive framework for documenting cross-linkers and their properties, facilitating improved data sharing and analysis across various spectrometry technologies, including GC-MS [7].

Additionally, benchmarks for metabolomics data analysis standardize data processing and interpretation, crucial for reliable comparisons across studies involving technologies like LC-MS, FIA-MS, GC-MS, and NMR [6]. By leveraging AI and deep learning, researchers can automate complex data analysis processes, minimizing human error and significantly enhancing the throughput and accuracy of spectrometry-based food authentication. These technological advancements streamline analytical workflows and bolster regulatory compliance, instilling consumer confidence in food product authenticity and quality.

2.3 Sample Preparation and Analysis Techniques

Sample preparation for Gas Chromatography-Mass Spectrometry (GC-MS) is critical for ensuring result accuracy and reliability, particularly with complex food matrices. Effective sample preparation techniques are essential for detecting food contaminants and residues, which can be challenging due to intricate food compositions and various interfering substances [1]. QuEChERS (Quick, Easy, Cheap, Effective, Rugged, and Safe) has gained prominence for its efficiency in extracting a broad range of analytes from diverse matrices, simplifying the sample preparation process and enhancing GC-MS detection capabilities [8]. The Modified QuEChERS approach optimizes extraction and cleanup processes, crucial for accurate pesticide residue analysis in food and environmental samples, enhancing analyte recovery rates and reducing matrix interferences [8].

In addition to QuEChERS, solid-phase microextraction (SPME) is a valuable technique for optimizing sample preparation for GC-MS. Both methods improve the extraction and analysis of various contaminants, addressing the increasing complexity of food safety testing by allowing simultaneous extraction of multiple target analytes, thus enhancing overall sensitivity and robustness [4, 8, 1, 2]. SPME enables direct extraction of volatile and semi-volatile compounds from complex matrices, minimizing sample handling and reducing contamination risks. This technique is particularly beneficial for analyzing flavor compounds, pesticides, and trace elements in food products.

Furthermore, advanced data analysis workflows support sample preparation by facilitating the preprocessing and interpretation of complex metabolomics datasets, ensuring comprehensive and reproducible analysis [6]. The advancement and optimization of sample preparation techniques are essential for effectively addressing complexities encountered in GC-MS analysis of intricate food matrices, especially in light of the increasing prevalence of microorganic contaminants and the need for multi-residue analytical methods such as QuEChERS and adsorption extraction. These enhancements improve the efficiency, robustness, and sensitivity of contaminant detection in food safety laboratories, contributing to broader efforts in ensuring food safety and authenticity [4, 1, 6, 9].

In recent years, the application of Gas Chromatography-Mass Spectrometry (GC-MS) has gained prominence in the field of food authentication. This technique is particularly valuable for detecting contaminants and residues, as well as for verifying food composition. As illustrated in Figure 2, the figure categorizes the key roles of GC-MS, highlighting its methodological advancements and data analysis techniques that significantly enhance its effectiveness in ensuring food safety and authenticity. By integrating these aspects, GC-MS not only provides a robust framework for food analysis but also offers substantial advantages over other analytical techniques.

3 GC-MS in Food Authentication

3.1 Detection of Contaminants and Residues

Gas Chromatography-Mass Spectrometry (GC-MS) is crucial for detecting contaminants and residues in food, providing a robust analytical framework for ensuring food safety and authenticity. Its ability to separate, identify, and quantify trace levels of volatile and semi-volatile compounds facilitates

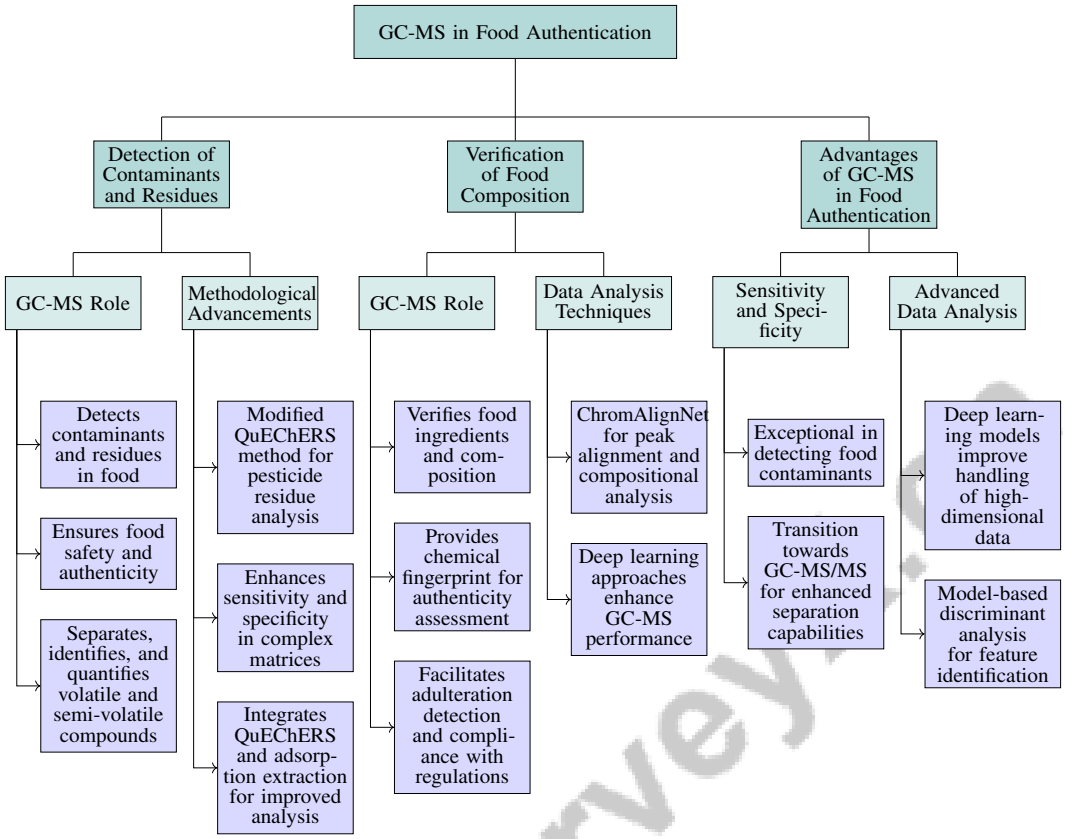


Figure 2: This figure illustrates the application of GC-MS in food authentication, highlighting its role in detecting contaminants and residues, verifying food composition, and offering significant advantages over other analytical techniques. The figure categorizes the key roles, methodological advancements, and data analysis techniques that enhance GC-MS's effectiveness in ensuring food safety and authenticity.

the detection of unwanted substances like pesticides, mycotoxins, and chemical residues that pose health risks. Recent advancements have led to the development of rapid, cost-effective methods for contaminant detection, enhancing GC-MS reliability across diverse food matrices [1].

The modified QuEChERS (Quick, Easy, Cheap, Effective, Rugged, and Safe) method represents a significant advancement, designed for the simultaneous analysis of multiple pesticide residues. This method optimizes extraction and cleanup processes, improving the sensitivity and specificity of GC-MS analyses in complex matrices such as soil [8]. By enhancing analyte recovery rates and minimizing matrix interferences, the modified QuEChERS method enables more accurate contaminant level assessments in food products.

Integrating advanced sample preparation techniques, such as QuEChERS and adsorption extraction, with GC-MS facilitates simultaneous extraction of multiple target analytes, enhancing sensitivity and selectivity in contaminant analysis. This approach aids regulatory compliance by ensuring adherence to stringent safety standards, addressing concerns over emerging contaminants in the food supply [2, 1, 4, 8, 3]. As demand for food safety and authenticity grows, GC-MS remains indispensable in detecting contaminants and residues, significantly contributing to public health protection and the integrity of the global food supply chain.

3.2 Verification of Food Composition

GC-MS plays a pivotal role in verifying food ingredients and composition, offering a detailed chemical fingerprint essential for authenticity assessments. This verification process involves accurately classifying food samples, a challenging task in high-dimensional datasets where variables often

exceed observations [3]. The capability of GC-MS to separate and identify complex mixtures allows precise determination of food composition, facilitating adulteration detection and compliance with labeling regulations.

Advanced data analysis techniques further enhance GC-MS performance in food composition verification. For instance, ChromAlignNet employs a deep learning approach that learns data patterns instead of relying solely on predefined mathematical rules, effectively managing the complexities of metabolomic data and improving peak alignment accuracy and compositional analysis reliability [2]. Such innovative methods ensure GC-MS remains a powerful tool for verifying food product authenticity, protecting consumer interests and maintaining food supply integrity.

3.3 Advantages of GC-MS in Food Authentication

GC-MS offers significant advantages for food authentication over other analytical techniques, primarily due to its exceptional sensitivity and specificity in detecting food contaminants. Recent developments highlight the transition towards Gas Chromatography coupled with tandem Mass Spectrometry (GC-MS/MS), enhancing separation capabilities for simultaneous identification of multiple contaminants. This shift is supported by innovative sample preparation methods, such as QuEChERS and adsorption extraction, which streamline analyte extraction, making the process faster, more cost-effective, and robust. These advancements underscore GC-MS's versatility in addressing food safety and quality concerns by enabling comprehensive detection of emerging food contaminants [4, 1].

A key benefit of GC-MS is its ability to separate and identify complex mixtures of volatile and semi-volatile compounds with high precision, crucial for ensuring food product authenticity and quality. Unlike other techniques, GC-MS provides a comprehensive chemical fingerprint of samples, facilitating the detection of subtle compositional differences indicative of adulteration or fraud.

The integration of advanced data analysis techniques further enhances GC-MS capabilities, allowing efficient extraction and interpretation of meaningful patterns from complex datasets. For example, deep learning models, as discussed in [5], improve the handling of high-dimensional data, enhancing authenticity assessment accuracy. Additionally, model-based discriminant analysis frameworks, highlighted by [3], boost classification accuracy by identifying the most relevant features from spectrometry data.

Moreover, GC-MS is highly effective in detecting contaminants and residues, providing a robust analytical framework for food safety. Its sensitivity and specificity allow for accurate quantification of trace levels of unwanted substances, such as pesticides and mycotoxins, critical for regulatory compliance and consumer protection [8]. The modified QuEChERS method, which optimizes extraction and cleanup processes, further enhances the detection capabilities of GC-MS, making it a preferred choice for food safety assessments [1].

4 AI and Deep Learning in Spectrometry Data Analysis

The integration of Artificial Intelligence (AI) and deep learning in spectrometry data analysis is revolutionizing the field by enhancing the precision and efficiency of data interpretation. This section delves into AI-driven automation of critical tasks in spectrometry, particularly the automated selection of regions of interest (ROI) in Gas Chromatography-Mass Spectrometry (GC-MS) data, highlighting the methodologies and benefits of these innovations.

4.1 Automated Region Selection in GC-MS Data

AI techniques significantly improve the automation of ROI selection in GC-MS data, addressing the challenges of multivariate datasets. Manual ROI selection is traditionally labor-intensive and susceptible to human bias, potentially affecting analysis accuracy. AI-driven approaches, such as ChromAlignNet, minimize user input and adaptively learn from data patterns, enhancing alignment accuracy in GC-MS data [4, 2]. ChromAlignNet's adaptive learning ensures precise peak identification and alignment, thus improving the quality of spectrometric analyses.

The Basis-Projected Layer (BPL) method further transforms sparse GC-MS datasets into dense representations by projecting data onto an N-sphere, enhancing deep learning model performance in high-dimensional spectrometry data and streamlining workflows by reducing manual intervention [5].

4.2 Model-Based Discriminant Analysis for Food Authenticity

Model-Based Discriminant Analysis (MBDA) offers a robust framework for classifying authentic food samples by integrating variable selection within the model-fitting process. This approach is crucial for identifying informative features from spectrometry data, thereby enhancing the precision of authenticity assessments. The Model-Based Discriminant Analysis with Variable Selection (MBDA-VS) leverages both labeled and unlabeled data, improving classification outcomes in high-dimensional datasets where traditional methods may falter [3].

The efficacy of MBDA-VS in food authenticity is enhanced by advanced data representation techniques like the BPL method, which reformulates data into higher-dimensional representations while preserving original data space properties. This enhances the model's ability to learn from sparse data, a common challenge in spectrometry analyses, improving the robustness of discriminant analysis models [5].

Moreover, pseudo-F-ratio analysis facilitates automated ROI selection without predefined regions or extensive user input, allowing MBDA to effectively utilize multivariate information in GC-MS data for more accurate discrimination of authentic food samples [4]. By integrating these model-based approaches, researchers achieve higher accuracy in food authenticity classification, supporting regulatory compliance and reinforcing consumer trust. Recent advancements in MBDA and related techniques underscore their potential to improve food authentication processes by effectively utilizing labeled and unlabeled data and identifying key variables for accurate labeling. The integration of advanced analytical techniques, such as GC-MS/MS and innovative sample preparation methods, further enhances the detection and analysis of food contaminants, solidifying the efficacy of model-based approaches in ensuring food quality and safety [4, 1, 5, 3].

4.3 Deep Learning for Peak Alignment

Deep learning techniques have transformed chromatographic peak alignment in GC-MS data, significantly enhancing the accuracy and reliability of spectrometric analyses. ChromAlignNet, a deep learning framework for peak alignment, leverages neural networks to learn directly from data patterns, accommodating the inherent variability in chromatographic data [2]. This adaptability allows ChromAlignNet to effectively manage diverse GC-MS datasets, enhancing alignment accuracy by uncovering complex relationships within the data, even when retention times vary across samples.

ChromAlignNet outperforms traditional mathematical methods in aligning peaks, requiring no user input for reference chromatograms, thus streamlining the biomarker discovery process [5, 2, 6, 4, 3]. By automating the alignment process, ChromAlignNet not only enhances peak alignment precision but also reduces errors associated with manual methods, facilitating faster and more consistent results.

Incorporating advanced deep learning models such as ChromAlignNet into the peak alignment process significantly improves the analysis of complex and high-dimensional GC-MS data, addressing challenges related to aligning varying retention times across chromatograms. This model effectively handles the intricacies of GC-MS datasets, which often contain noise and sparse representations, enabling more accurate biomarker discovery and analysis without extensive user input [5, 2, 6, 4, 9]. Accurate peak alignment is critical for subsequent data analysis steps, such as compound identification and quantification, essential for food authentication and quality control. Through deep learning applications, researchers can achieve more reliable and reproducible analyses, ultimately supporting broader objectives of ensuring food safety and authenticity.

5 Data Analysis Techniques

A thorough grasp of both traditional and modern methodologies is pivotal for advancing Gas Chromatography-Mass Spectrometry (GC-MS) analysis. Table 1 provides a comprehensive overview of the modern AI-driven approaches and advanced techniques employed in Gas Chromatography-Mass Spectrometry (GC-MS) analysis. Additionally, Table 2 provides a detailed comparison of

Category	Feature	Method
Modern AI-Driven Approaches	Feature Optimization	MBDA-VS[3]
	AI-Driven Data Processing	CAN[2]
	Data Representation Enhancement	BPL[5]
Advanced Techniques in Pesticide Analysis	Extraction Efficiency	FRMV[4]

Table 1: Table summarizing modern AI-driven approaches and advanced techniques in pesticide analysis for Gas Chromatography-Mass Spectrometry (GC-MS). The table categorizes methods into AI-driven approaches and advanced techniques, detailing specific features and methodologies, including MBDA-VS for feature optimization and FRMV for extraction efficiency. These methods demonstrate the integration of AI in enhancing data processing and representation, as well as the advancement of pesticide analysis techniques.

traditional, modern AI-driven, and advanced techniques employed in Gas Chromatography-Mass Spectrometry (GC-MS) data analysis, highlighting their respective data processing methods, performance focuses, and key advantages. This section delves into foundational traditional methods, recognizing their critical role and limitations, particularly in detecting low-concentration analytes. Subsequently, modern AI-driven approaches are highlighted for their potential to address these challenges.

5.1 Traditional Data Analysis Methods

Traditional GC-MS data analysis methods have been instrumental in interpreting complex spectrometric data, offering insights into food composition through manual inspection of chromatograms and mass spectra. These methods depend on retention time and spectral matching with established libraries, demanding significant manual effort and expertise, which affects accuracy and reliability. Tools like Workflow4Metabolomics 3.0 enhance reproducibility by providing user-friendly features and comprehensive resources for metabolomics technologies [4, 6, 9].

A significant limitation of traditional methods is their performance at low analyte concentrations, where spectral quality may merge with baseline noise, leading to detection dropouts [4]. This highlights the need for advanced techniques to improve signal detection and analyte identification reliability.

Despite these challenges, traditional methods remain crucial, especially when integrated with modern computational tools like Workflow4Metabolomics 3.0, enhancing accuracy in detecting low-concentration analytes. This synergy is vital for exploring complex biological questions and ensuring transparency and reproducibility in research through complete analyses with permanent DOIs [4, 6, 9]. The interplay between traditional and modern techniques is essential for advancing GC-MS data analysis.

5.2 Modern AI-Driven Approaches

AI-driven methods significantly advance data analysis for food quality control, leveraging computational power and deep learning to enhance spectrometric analysis accuracy and efficiency. These approaches effectively analyze high-dimensional data, facilitating pattern identification crucial for food authentication. Techniques like model-based discriminant analysis and deep learning algorithms, such as ChromAlignNet, outperform traditional methods in classifying and aligning complex food data. Additionally, advancements in multi-residue analytical methods, including QuEChERS, enhance contaminant detection, supporting rigorous food safety measures [1, 2, 3].

ChromAlignNet, designed for GC-MS peak alignment, learns from data patterns, adapting to chromatographic variability, improving precision, and reducing manual errors [2]. This adaptability is crucial for accurate compound identification and quantification, essential for verifying food composition and detecting contaminants.

The Basis-Projected Layer (BPL) method exemplifies AI integration, transforming sparse datasets into dense representations to enhance deep learning performance. By projecting data onto an N-sphere, BPL improves model capability in identifying GC-MS regions of interest, streamlining workflows and reducing manual intervention [5]. This supports automation in data processing, increasing analysis throughput.

Model-Based Discriminant Analysis with Variable Selection (MBDA-VS) further illustrates AI efficacy in food quality control, enhancing classification accuracy by incorporating variable selection during model fitting [3]. This ensures informative feature utilization for authenticity assessments, reinforcing regulatory compliance and consumer trust.

AI-driven approaches revolutionize food quality control data analysis, employing advanced techniques like deep learning for GC-MS peak alignment to enhance accuracy and efficiency. Methods like GC-MS/MS and innovative preparation techniques, including QuEChERS, enable multiple analyte extraction and superior contaminant separation, providing robust solutions for detecting emerging contaminants and improving food safety [1, 2]. These advancements bolster food authentication reliability and support efforts to ensure food safety and integrity.

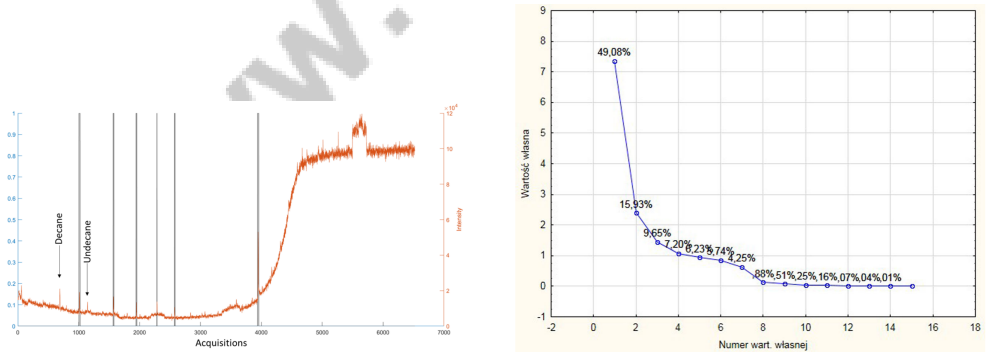
5.3 Advanced Techniques in Pesticide Analysis

Advanced GC-MS techniques have significantly improved pesticide residue detection and quantification in food products. Comprehensive sample preparation strategies are essential to mitigate matrix effects, ensuring effective analyte isolation and concentration for enhanced GC-MS sensitivity and reliability [1].

The modified QuEChERS technique is highly effective for routine pesticide analysis, efficiently extracting residues from complex matrices. It optimizes extraction and cleanup processes to improve recovery rates and minimize matrix interferences, enhancing GC-MS sensitivity and specificity, and streamlining high-throughput testing workflows [8].

Tandem mass spectrometry (GC-MS/MS) adds accuracy in pesticide analysis through multiple reaction monitoring (MRM) modes, enabling selective detection of residues at trace levels, essential for meeting regulatory standards and ensuring consumer safety through accurate residue identification and quantification [1].

Integrating advanced preparation techniques and GC-MS/MS marks significant pesticide analysis advancements. Transitioning to tandem mass spectrometry, coupled with methodologies like QuEChERS, enhances analytical performance by improving contaminant separation. These methods facilitate simultaneous multi-analyte extraction, ensuring fast, cost-effective, and robust analyses that yield reliable results, support regulatory compliance, and safeguard public health against increasing contaminants. High-resolution mass spectrometry integration further enhances comprehensive contaminant detection, reinforcing food safety measures [4, 1].



(a) Decane and Undecane Detection in a Gas Chromatography-Mass Spectrometry (GC-MS) Experiment[4]

(b) The image is a scatter plot graph with a title in Polish.[8]

Figure 3: Examples of Advanced Techniques in Pesticide Analysis

As illustrated in Figure 3, advanced data analysis techniques are crucial for accurately identifying and quantifying chemical compounds in pesticide analysis. The first image presents a GC-MS chromatogram, highlighting decane and undecane detection, with the x-axis indicating acquisitions and the y-axis representing mass spectrum intensity. The orange line shows concentration levels, while vertical markers denote retention times, essential for compound separation and identification. The second image, a scatter plot with axes labeled in Polish, depicts eigenvalue trends, providing

insights into data structure and variability. These figures exemplify advanced analytical techniques, offering a glimpse into chemical detection and data interpretation processes [4, 8].

Feature	Traditional Data Analysis Methods	Modern AI-Driven Approaches	Advanced Techniques in Pesticide Analysis
Data Processing Method	Manual Inspection	Deep Learning	Tandem Mass Spectrometry
Performance Focus	Low Analyte Concentrations	High-dimensional Data	Pesticide Residue Detection
Key Advantage	Reproducibility Enhancement	Accuracy And Efficiency	Sensitivity And Specificity

Table 2: A comparative analysis of data processing methods in Gas Chromatography-Mass Spectrometry (GC-MS) showcases the evolution from traditional manual inspection techniques to modern AI-driven approaches and advanced pesticide analysis techniques. The table highlights key features such as data processing methods, performance focus, and advantages of each approach, emphasizing the transition towards more accurate, efficient, and sensitive methodologies.

6 Challenges and Future Directions

6.1 Challenges in GC-MS Data Analysis

Gas Chromatography-Mass Spectrometry (GC-MS) data analysis faces challenges that impede accurate analyte identification and quantification in complex food matrices. A significant issue is the reliance on user-defined parameters, which can introduce artifacts and inconsistencies [4]. The complexity of food matrices demands high sensitivity and selectivity to discern target compounds amidst numerous interferences [1]. Co-eluting compounds further complicate detection, necessitating advanced techniques for improved analyte resolution. In pesticide detection, matrix effects impact residue analysis accuracy, causing variable recovery rates and quantification errors [8]. Robust sample preparation and analysis techniques are essential to address these issues.

Retention time (RT) variability across chromatograms complicates the alignment of identical analytes, requiring sophisticated alignment algorithms for accurate peak identification [2]. Additionally, the sparsity of GC-MS data challenges deep learning models, potentially misrepresenting data geometry [5]. Advanced computational methods are needed to capture the underlying data structure effectively.

Variable selection in model-based approaches can be compromised by highly correlated variables, affecting authenticity assessments [3]. Improved feature selection techniques are necessary to handle multicollinearity. Moreover, existing benchmarks often lack diversity, leading to misleading conclusions about model effectiveness [9]. Developing comprehensive benchmarks that reflect GC-MS data analysis complexities is crucial for evaluating and comparing analytical methods.

6.2 Integration of Emerging Technologies

Integrating emerging technologies into food authentication practices offers significant potential for improving data analysis and model development. A key focus is reducing false positives in GC-MS data analysis to enhance food authentication reliability [2]. Developing robust models that accurately distinguish between authentic and adulterated products is essential for improving authenticity assessments.

Models must generalize across varying data complexities to be effective in diverse food matrices. Future research should enhance this adaptability, ensuring analytical techniques remain effective across a wide range of food products [2]. Incorporating advanced computational methods, such as deep learning and machine learning, into GC-MS data analysis underscores the transformative potential of these technologies. These methodologies improve the extraction and interpretation of intricate data patterns, significantly enhancing accuracy and efficiency. Addressing challenges like signal fragmentation and noise, and employing flexible chemometric techniques such as Multivariate Curve Resolution and Parallel Factor Analysis, facilitates precise data analysis while minimizing artifacts. Platforms like Workflow4Metabolomics 3.0 streamline the analysis process by integrating multiple metabolomics technologies, providing user-friendly tools for data preprocessing, statistical analysis, and annotation [2, 1, 6, 4, 9]. This integration enhances analytical workflows, reduces human error, and increases throughput in food authentication processes.

Advanced gas chromatography coupled with tandem mass spectrometry (GC–MS/MS) and innovative data analysis methods, such as model-based discriminant analysis and deep learning algorithms,

effectively enhance food safety. These technologies enable more reliable detection of contaminants and improved classification of food samples, ensuring the integrity and safety of food products through faster, more accurate, and comprehensive analytical solutions [1, 2, 3]. As these technologies evolve, their application in food authentication is expected to yield significant improvements in detecting adulteration and verifying food authenticity.

6.3 Standardization and Collaboration

Advancing food authentication and quality control through GC-MS and AI-driven methods requires concerted efforts toward standardization and collaboration within the scientific community. Developing standardized methods for pesticide analysis is crucial for enhancing accuracy and reliability in food quality assessments. Establishing universally accepted protocols ensures consistency in analytical results, facilitating regulatory compliance and consumer protection [8].

The integration of emerging technologies, such as deep learning into GC-MS data analysis, highlights the necessity for collaborative efforts to enhance computational models. Future research should explore improvements to techniques like the Basis-Projected Layer (BPL), broadening its applicability to other types of sparse data and fostering collaboration in advancing deep learning methodologies [5]. Such collaborations could lead to the development of more robust models capable of addressing the complexities of high-dimensional spectrometric data.

Expanding existing frameworks, such as the xlmod ontology, is vital for improving data reporting standards in structural biology. Incorporating additional cross-linkers and enhancing data documentation can facilitate better data sharing and analysis across various spectrometry technologies, including GC-MS [7].

Unresolved issues remain regarding detecting emerging contaminants and developing universally applicable methods for diverse food types. Addressing these challenges requires a unified effort from researchers, regulatory bodies, and industry stakeholders to create comprehensive solutions that adapt to the evolving landscape of food safety concerns [1].

Refining algorithms to enhance sensitivity, particularly at low concentrations, remains an area of ongoing research. Future studies could focus on optimizing these algorithms and exploring their application in other analytical contexts, broadening the scope of GC-MS applications [4].

The future of food authentication and quality control will significantly depend on enhancing collaborative initiatives and developing standardized methodologies that adapt to rapid technological advancements and the growing complexities of food safety challenges. This includes implementing advanced analytical techniques, such as GC–MS/MS, which offers superior sensitivity and separation capabilities for detecting food contaminants. Innovative sample preparation methods, like QuEChERS and adsorption extraction, are emerging as effective tools for simultaneously extracting multiple target analytes, improving the efficiency and accuracy of food safety assessments. As the landscape of food contaminants evolves, integrating these methods with high-resolution mass spectrometry and other sensitive detection technologies is crucial for comprehensive monitoring of food quality and safety [1, 6, 3].

7 Conclusion

The survey highlights the pivotal contributions of Gas Chromatography-Mass Spectrometry (GC-MS) and Artificial Intelligence (AI) in advancing food authentication and quality assurance. By integrating GC-MS with AI and deep learning, the precision and efficiency in verifying food composition and detecting contaminants are significantly enhanced. AI-driven approaches, such as Model-Based Discriminant Analysis with Variable Selection (MBDA-VS), demonstrate superior performance over conventional classification methods, particularly in handling complex, high-dimensional datasets.

Advancements in sample preparation, including the refined QuEChERS method, alongside sophisticated deep learning models like ChromAlignNet for peak alignment, augment the analytical power of GC-MS. These developments not only improve sensitivity and specificity but also optimize analytical workflows, ensuring compliance with regulatory standards and safeguarding consumer interests.

Emerging technologies present new opportunities for research and innovation in this field. Efforts to enhance model adaptability across varied food matrices and reduce false positive rates in GC-

MS analyses are crucial for refining food authentication methodologies. Moreover, the creation of standardized benchmarks, akin to those used in natural language processing, could guide future advancements in spectrometry data analysis, fostering more robust and reliable food safety practices.

References

[1] ehensive strategy for sample pre.

[2] Mike Li and X. Rosalind Wang. Peak alignment of gas chromatography-mass spectrometry data with deep learning, 2019.

[3] Thomas Brendan Murphy, Nema Dean, and Adrian E. Raftery. Variable selection and updating in model-based discriminant analysis for high dimensional data with food authenticity applications, 2010.

[4] Ryland T. Giebelhaus, Michael D. Sorochan Armstrong, A. Paulina de la Mata, and James J. Harynuk. Untargeted region of interest selection for gc-ms data using a pseudo f-ratio moving window (ψfrmv), 2022.

[5] Yu Tang Chang and Shih Fang Chen. Design of an basis-projected layer for sparse datasets in deep learning training using gc-ms spectra as a case study, 2024.

[6] Yann Guitton, Marie Tremblay-Franco, Gildas Le Corguillé, Jean-Francois Martin, Mélanie Pétéra, Pierrick Roger-Mele, Alexis Delabrière, Sophie Goulitquer, Misharl Monsoor, Christophe Duperier, et al. Create, run, share, publish, and reference your lc–ms, fia–ms, gc–ms, and nmr data analysis workflows with the workflow4metabolomics 3.0 galaxy online infrastructure for metabolomics. *The International Journal of Biochemistry & Cell Biology*, 93:89–101, 2017.

[7] Gerhard Mayer. Xlmod: Cross-linking and chromatography derivatization reagents ontology, 2020.

[8] Bożena Łozowicka, Ewa Rutkowska, and Magdalena Jankowska. Influence of quechers modifications on recovery and matrix effect during the multi-residue pesticide analysis in soil by gc/ms/ms and gc/ecd/npd. *Environmental Science and Pollution Research*, 24(8):7124–7138, 2017.

[9] Yann Guitton, Marie Tremblay-Franco, Gildas Le Corguillé, Jean-François Martin, Mélanie Pétéra, Pierrick Roger-Mele, Alexis Delabrière, Sophie Goulitquer, Misharl Monsoor, Christophe Duperier, Cécile Canlet, Remi Servien, Patrick Tardivel, Christophe Caron, Franck Giacomoni, and Etienne Thévenot. Create, run, share, publish, and reference your lc-ms, fia-ms, gc-ms, and nmr data analysis workflows with the workflow4metabolomics 3.0 galaxy online infrastructure for metabolomics, 2017.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.