Journal of Applied Pharmaceutical Research
Volume 13 Issue 3, Year of Publication 2025, Page 01 – 16
DOI: https://doi.org/10.69857/joapr.v13i3.861

Check for updates

Review Article

JOURNAL OF APPLIED PHARMACEUTICAL RESEARCH | *JOAPR*
www.japtronline.com *ISSN: 2348 – 0335*

A COMPREHENSIVE REVIEW ON THE APPLICATIONS OF CHEMOMETRICS IN ANALYTICAL CHEMISTRY

Projesh Saha*, Bibhas Pandit, Soma Pramanik, Bhupendra Shrestha

Article Information
Received: 31st March 2025
Revised: 15th May 2025
Accepted: 14th June 2025
Published: 30th June 2025

Keywords
Chemometrics, Data-driven Methodologies, Multivariate Analysis, Artificial Intelligence, Machine Learning in Chemometrics

ABSTRACT
Background: This article presents a review of the various applications of chemometrics in analytical chemistry. Chemometrics is essential to analytical chemistry because it provides sophisticated methods for extracting, analyzing, and interpreting chemical data. To maximize analytical procedures, enhance data reliability, and extract insights, this field combines statistical and mathematical techniques with chemical research. Chemometrics provides analytical chemists with the means to manage the massive datasets generated by contemporary analytical methods, such as spectroscopy and chromatography. **Methodology**: This review combines data from previous research articles that have elaborated and described the various applications of chemometrics in the analytical chemistry sector. **Results and Discussion**: The combination of chemometrics with artificial intelligence and machine learning offers more advanced analytical and predictive modelling possibilities. It is anticipated that these developments will transform analytical chemistry by enhancing researchers' ability to manage complex datasets and gain deeper insights from their investigations. This is especially important in industries where precise data interpretation is critical, such as pharmaceuticals, ecological surveillance, and food safety. **Conclusion**: Chemometrics is essential to contemporary analytical chemistry because it provides methods and instruments that enhance quality control, facilitate innovative research advancements, and improve data analysis capabilities. The accuracy and efficacy of chemical analyses are expected to continue to improve as this sector develops.

INTRODUCTION

Chemometrics is the science of using techniques from multivariate statistics, applied mathematics, and computer science to extract information from chemical systems through data-driven methodologies to solve issues in a variety of scientific domains [1]. In the early 1970s, Swedish researcher Svante Wold introduced the term "chemometrics" in his grant proposal, which outlined the application of statistical techniques to chemical data. "Kemometri," combines the words "kemo-" for chemistry and "-metri" for measure [2]. To evaluate and interpret complex chemical data, chemometrics, a crucial interdisciplinary field, combines computer science, statistics, and chemistry. It correlates chemical data with fundamental chemical events using statistical and mathematical methods,

*Department of Pharmaceutical Analysis, Himalayan Pharmacy Institute, Majhitar, East Sikkim, 737136, India

***For Correspondence:** **im.projeshsaha18@gmail.com**

improving comprehension and forecasting abilities in a range of analytical scenarios [3]. Nevertheless, the appearance of novel high-dimensional hyphenated instruments and different software has led to a breakthrough in chemometrics. These chromatographic instruments have enabled the accurate and precise determination of a variety of analytes in complex matrices [4]. It manages multivariate data produced by analytical tools, including those from multiple tools used to describe the same sample. It generates graphics that are easy to understand and capture all the variables considered simultaneously [5]. Various areas and principles that contribute to chemometrics are illustrated in Figure 1.

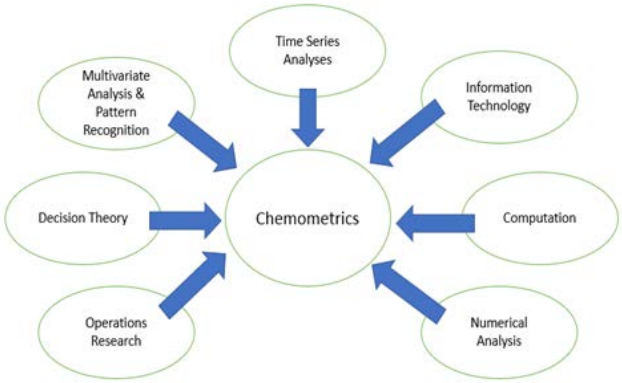

Figure 1: Various areas and principles that contribute to chemometrics [6]

The ability of chemometric principles to extract meaningful information from complex chemical systems is utilized in various areas, as shown in Figure 1. Time series analysis enhances current methods by measuring dynamic interactions of variables throughout time. As a branch of artificial intelligence (AI) and machine learning (ML), chemometrics is particularly well-suited for chemical analysis. Multivariate calibration techniques utilize multiple variables to determine concentrations, thereby reducing noise and accounting for interferents, in contrast to univariate techniques that focus on a single variable. To categorize substances and variables, find possible biomarkers, and lower the degree of dimensionality of complex systems, pattern recognition algorithms are employed [6].

Chemometrics is utilized in applications involving quality control to ensure product consistency and regulatory compliance in sectors such as food safety and pharmaceuticals. Manufacturers can preserve the quality of the products being developed by using chemometric techniques to track analytical measurements and identify deviations that could point to production problems [7]. Among the frequently employed process analyzers in the chemical industry are visible spectroscopy, nuclear magnetic resonance (NMR) spectroscopy, mid-infrared (MIR) spectroscopy, near-infrared spectroscopy (NIRS), and Fourier transform infrared (FTIR) spectroscopy. The process above analyzers provide a wealth of data, which has been analyzed and modelled using a variety of statistical tools, including PCA, modified soft independent modelling of the class analogy (SIMCA), canonical variable analysis (CVA), PCR, and partial least squares (PLS) [8]. Chemometrics has also been applied to the rapid examination of a gaseous effluent resulting from a heterogeneously catalyzed reaction, the online assessment of the carcinogenic toxicity of lubricant base oil, and the verification of meat products. To maximize process understanding, chemometric tools can be beneficial and powerful in obtaining important process information via analysis of easily accessible data [9].

Numerous methods are available in chemometrics for designing experiments, doing exploratory analysis, creating calibration and classification models, and validating them. Chemometrics has taken advantage of and drawn the interest of a large scientific community in the contemporary omics era [10]. Fundamental ideas in chemometrics are derived from a variety of statistical approaches that can be categorized into two primary types. The first one is descriptive chemometrics, which summarizes and visualizes data to comprehend the fundamental structure of chemical systems. This group includes methods like PCA and clustering. These techniques aid in locating connections and patterns in intricate datasets, directing the creation of additional experiments [11]. The second one is predictive chemometrics, which concentrates on creating models that, given input data, can predict results. Regression analysis, like multiple linear regression (MLR) and PLS, is often used to forecast analyte concentrations in complicated mixtures using spectral data from instruments [12].

Two categories of chemometric tools are qualitative and quantitative approaches. Classification and pattern recognition problems are the focus of the qualitative approach. These techniques are based on two theories that categorize procedures as either supervised or unsupervised [2]. Unsupervised approaches aim to reveal fundamental data structures without

introducing potential bias related to prior knowledge of group memberships. Conversely, the goal of supervised methods is to create the best possible group separation. As a result, they optimize the classification method's capacity to forecast the membership of a class of samples whose membership is unknown. Multiple chemometric methods are occasionally used because it is not necessarily a clear-cut decision [11]. Quantitative techniques are employed to determine the relationship that exists between detected signals and precise concentration values. Regression procedures examine a relationship (linear or nonlinear) between one or more independent variables and one (or more, but typically one) dependent variable [12]

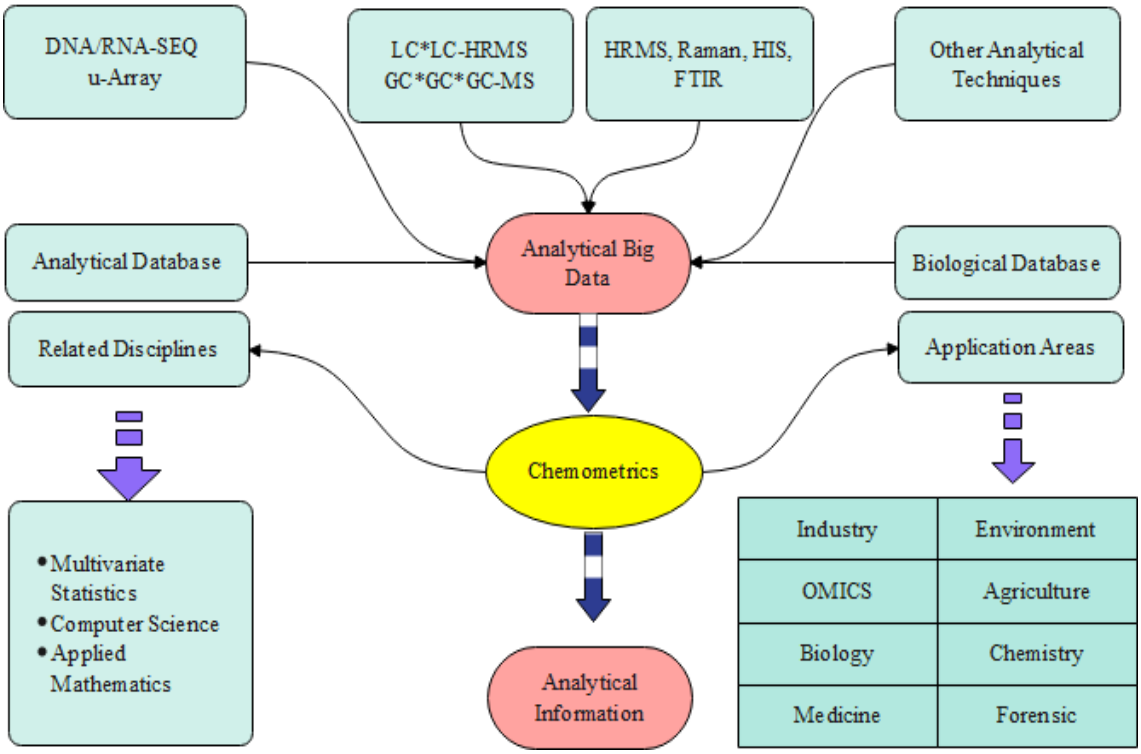

Figure 2: Chemometrics as an interdisciplinary field [12]

Chemometrics has a broad spectrum of applications in analysis, as illustrated in Figure 2.

Applications of chemometrics in analytical chemistry include the following:

1. ***Analysis of phytochemical profiles*:**

Chemometrics is increasingly recognized as a powerful tool in the investigation of phytochemical profiles. Researchers can more thoroughly examine big datasets by applying chemometric approaches, which show the connections between various substances and their possible health advantages [13]. To discover and measure the bioactive chemicals in plant materials responsible for their medicinal properties, chemometrics is essential. For example, chemometric techniques have been used in research studies involving Salvia species to link antioxidant activity with phenolic content, making it easier to identify phytochemicals that may have medicinal uses [14]. PCA is a popular multivariate analysis (MA) technique for visualizing correlations and variance in phytochemical data. Spectral data are correlated with chemical profiles or physiological activities using partial least squares (PLS). CA helps identify species or chemotypes exhibiting comparable biological characteristics by grouping related samples according to the corresponding phytochemical profiles [15].

Plant samples are categorized into predetermined groups using discriminant analysis techniques, such as Linear Discriminant Analysis (LDA), which helps ensure the quality and authenticity of herbal products. Data from many analytical methods, including Nuclear Magnetic Resonance (NMR), Gas Chromatography-Mass Spectrometry (GC-MS), & High-Performance Liquid Chromatography (HPLC), can be integrated using chemometric methodologies [16]. Despite its numerous benefits, chemometrics for phytochemical profiling has several drawbacks, including the complexity of the data, the potential for misunderstandings, and sample variability. In evaluating natural variation, plant-to-plant variability can account for a

significant portion of the total variability. All metabolic profiles in a research investigation involving 19 Arabidopsis thaliana accessions demonstrated that plant-to-plant diversity exceeded experimental batch variance and natural variation between accessions. Plant-to-plant variability accounted for half of the total variance, with the mean between-plant variance being 20% greater than the between-accession variance. Variance at every stage of an experimental design must be included to obtain a comprehensive picture of variability. Therefore, when using chemometric techniques, researchers must ensure that the experimental design is sound and that the data are validated [17].

2. *In Food Science Technology*:

Chemometrics has a wide range of constantly developing applications in food science and technology that tackle a variety of issues related to product creation, safety, quality control, and customer satisfaction [18]. Chemometric-based methods can be implemented in food authenticity, food quality evaluation, mathematical modelling, and process optimization in the food manufacturing industry [19].

By identifying pollutants in food products, chemometric approaches such as MA enhance the capacity for assessing food quality. Food composition can be understood with the use of techniques like PCA, which enables quality assurance and detection of adulteration. When paired with chemometrics, non-targeted metabolomic techniques enable thorough characterization of food metabolites. To improve the identification of genuine components in intricate food matrices, techniques like PLS Discriminant Analysis (PLS-DA) can be extensively used to categorize food products according to their phytochemical profiles [20]. Identifying chemical, biochemical, or biological markers linked to food safety and quality; determining the authenticity, quality, bio-functional, and nutritional aspects of food, existence of non-disclosed additives or contaminants in foods and facilitating evaluation of chemical & biological information obtained via instrumental analysis are all possible applications of chemometrics in food research and innovation [21]. Build reliable mathematical models to forecast the shelf life of processed foods and track the operations of food processing units; comprehend the relationships among sensory modalities and the physical, compositional, textural, and microstructural characteristics of food; optimize formulation & processing conditions in applications related to R&D problem-solving [22].

By assessing how different parameters affect extraction efficiency, techniques such as Response Surface Methodology (RSM) ensure that the highest yield of targeted chemicals is obtained at the lowest possible cost of materials and energy.

Chemometric modeling facilitates the understanding and control of microbial activity during food fermentation, thereby maximizing flavor and quality. Researchers can create models that forecast how fermentation processes will turn out by keeping an eye on the parameters of fermentation and the characteristics of the final product [23]. Exploratory (unsupervised statistical techniques like PCA and CA), classification (supervised statistical techniques like Discriminant Analysis, e.g., LDA, PLS-DA, etc., K-nearest neighbours, SIMCA, and Artificial Neural Networks, i.e. ANNs), regression, and calibration (MLR, PLS, PCR) are the three main chemometric approaches that are typically used when answering food R&D questions.

Even though the application of chemometrics in food research and development is becoming more and more popular, common instances of chemometric misuse include a lack of knowledge about chemometric tools, a lack of Design of Experiments (DoE) alongside the use of poorly designed DoE, and discrepancies in model calibration, validation, and update [24]. In one investigation, adulteration of Atlantic salmon using salmon trout muscle was detected using FTIR with PLS-R and exploratory PCA. PCA was able to cluster muscle samples according to the degree of adulteration [25].

Another study predicted microbiological spoilage of minced pork under different storage settings using spectroscopy-based sensors (FTIR), visual spectroscopy (VIS), & multispectral imaging (MSI) [26]. The results showed that FTIR sensors had the best prediction power. Green coffee beans were categorized by geographic origin using near-infrared reflectance spectroscopy (NIRS) and soft independent modeling of class analogy (SIMCA) [27]. Adulterants, including urea, melamine, and vegetable oils, are commonly detected in milk and dairy products using spectroscopic techniques such as FTIR, Raman spectroscopy, and NIRS in conjunction with chemometric approaches [28]. Chemometrics plays a crucial role in helping the food industry meet regulatory standards and support compliance with organizations such as the FDA and EMA. Research on food authenticity requires the ability to manage

massive datasets produced by sophisticated analytical techniques, which these technologies help with. By ensuring food authenticity and traceability, and facilitating fraud detection, chemometrics satisfies legal requirements for safe and properly labeled food products. Chemometrics improves both targeted and non-targeted analytical techniques used in food authenticity studies. While non-targeted approaches entail thorough chemical analysis, targeted methods examine predicted components. Despite the advantages, there are still obstacles in the adoption of techniques like NIRS. These include inadequate equipment calibration, a lack of standardized official standards, and limited method validation [29].

3. ***In Quality Assessment and Analysis of Medicines*****:**

Starting with medication formulation and extending to quality control throughout the production process, chemometrics enhances the effectiveness and reliability of pharmaceutical evaluations. Chemometric tools help analyze datasets and reveal sequences, correlations & structures that may be difficult to identify with traditional analytical techniques [30]. Blending spectroscopy with chemometrics leads to speedier quality checks that optimize production whilst assuring conformance regarding regulatory standards[31]. To create calibration models that link analytical signal data (like spectral absorbance) to concentration values for active pharmaceutical ingredients (APIs), chemometric approaches like PLS are used [32].

Through in-depth data analysis of chemical shifts, peak areas, and spectral attributes derived from methods such as NMR and HPLC, chemometric tools are used to assess the safety and potency of APIs. Impurities that could compromise the safety and efficacy of drugs must be identified and quantified in pharmaceutical manufacturing. During the pharmaceutical development process, these models aid in risk assessment & management [33].

By automating inspections, anticipating quality deviations, and streamlining production procedures, the incorporation of machine learning (ML) techniques, such as neural networks, significantly improves drug quality control. The pharmaceutical sector utilizes artificial neural networks (ANNs) for various tasks, including modeling and data analysis for pharmaceutical quality control. Drug release from matrix tablets can be predicted using neural networks, particularly those with a single hidden layer, which yield outcomes comparable to those of statistical analysis. Design of Experiments (DoE), which is a statistical tool employed in Quality by Design (QbD) approach to pharmaceutical research and ML methods like ANNs, are being utilized extensively to estimate stability, release of drugs, and manufacture of controlled release solid dosage formulations [34]. AI-powered quality control saves both money and time while increasing accuracy by analyzing data using software algorithms and ML frameworks to identify flaws or irregularities in drug manufacturing processes. However, several obstacles prevent AI and ML from being effectively utilized in drug quality monitoring. Since AI systems rely on high-quality data to generate trustworthy insights, data availability and quality are essential. Intricacy of incorporating ML approaches is further increased by initial cost outlays, data security issues, possible biases during algorithmic training & requirement for continuous monitoring [35].

Chemometrics significantly enhances the evaluation of pharmaceutical quality; however, several issues remain. Advanced abilities in data analysis and model development are required due to the overwhelming magnitude and complexity of the data collected in pharmaceutical analysis. Moreover, the industry often lacks standardized chemometric techniques, which can lead to inconsistent data processing and interpretation [36]. The intricacy of international supply networks, uneven data governance practices, and the requirement to comply with stringent regulatory standards are among the obstacles that pharmaceutical companies encounter when it comes to data standardization [37]. Data integration initiatives are made more challenging by the various data standards, terminologies, and measurement units employed by multiple parties, including producers, wholesalers, and logistics providers [38]. Data standardization is severely hampered by fragmented and separated information repositories, along with legacy systems that are incompatible with one another, frequently as a consequence of mergers and acquisitions. Synchronizing data to fulfill the prerequisites imposed by regulatory authorities, including the FDA and EMA, provides an additional layer of complication to the standardization procedure [39]. Serious legal and financial repercussions may result from violating laws such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). Incomplete or inaccurate data can lead to poor research or therapeutic decisions [40]. Additionally, high-quality reference datasets are crucial for calibrating and validating chemometric

models. These datasets may contain errors that result in incorrect findings regarding medication quality. Further advancements in chemometrics would ensure a significant increase in the pharmaceutical industry's ability to operate smoothly and effectively generate high-quality medications [41].

4. *In Environmental Analysis*:

The application of chemometric approaches enriches our understanding of ecological systems, enabling improved monitoring, evaluation, and assessment of environmental quality. Incorporation of chemometrics with contemporary technologies will significantly enhance environmental surveillance and protection initiatives as approaches continue to advance [42].

Environmental remediation studies, impact assessments, and the determination of chromatographic and spectroscopic oil fingerprints are just a few of the applications that have utilized techniques such as pattern recognition, response surface methodology (RSM), and experimental design. Because environmental samples have complicated composition, real-world applications of quantitative approaches are less likely to succeed, whereas most research yields sufficient meaningful information using qualitative and semi-quantitative methods [43]. The most often used chemometric techniques are first-order MA techniques like MLR and PLS, alongside non-supervised pattern recognition techniques like PCA and HCA, supervised pattern recognition techniques like k-nearest neighbour (k-NN), SIMCA, and discriminant analysis (DA) [44].

The use of MIR (Mid-IR) spectrum techniques, in conjunction with PCR and PLS, was employed to assess soil pH, total nitrogen, organic and total carbon, phosphorus concentration, and other relevant factors. IR is a quick and environmentally friendly substitute for standard agrochemical analytical techniques [45]. After using proper data preprocessing methods, advanced multivariate models applied to infrared spectra across both ranges can produce precise predictions for the majority of soil properties [46].

Chemometrics can also be used to assess river and other water body pollution by evaluating and estimating the water and sediment. PLS, PCR & Self-Organizing Map-Artificial Neural Network (SOM-ANN) can be implemented for analyzing water pollution. When it comes to characterizing river water contamination and its origins, SOM-ANN outperforms traditional chemometric methods (PCA and PLS). But for source receptor modelling, a combination of PCA, CA, and SOM-ANN has proven to be a more effective method [47]. By refining analytical techniques, increasing resource efficiency, reducing waste, and monitoring environmental impacts, chemometrics plays a crucial role in advancing sustainability.

According to a recent investigation, the Eco-Scale score, which assesses the environmental sustainability of analytical methodologies, significantly increases when chemometrics is applied in analytical chemistry. Chemometrics reduces waste by facilitating more precise quality control & surveillance across the food distribution network [48]. Chemometrics is also used in the domain of environmental contamination to evaluate the enactment of sustainable development policy. Chemometrics applied to liquid chromatographic techniques can expedite analytical processes in pharmaceutical assessment, reducing solvent usage, time, & energy [49].

MA aids in evaluating the quality of air and water, streamlining monitoring systems, and identifying causes of pollution, all of which improve environmental management and safeguard public health [50]. Patterns that affect the fate and transportation of contaminants can be found using multivariate statistical models [51]. MA aids in evaluating how different elements, including ecological characteristics, climate, and pollution sources, affect the welfare of the environment[52]. In pollutant monitoring, several multivariate techniques are used, including discriminant analysis (DA), factor analysis (FA), PCA, and CA [53]. By interpreting data to reveal contamination trends, CA helps identify specific connections between sample locations. By removing spatial correlations between past environmental monitoring data, PCA lowers the complexity of the data and increases the effectiveness of pollution assessment [54].

By providing tools for evaluating the potential impacts of contaminants on ecosystems and human health, chemometrics facilitates environmental risk assessments. Chemometrics facilitates informed decisions about environmental management by identifying correlations between various contaminants and their associated health effects. By helping to calculate the likelihood of negative impacts from exposure to different pollutants, these models support public health campaigns and regulatory measures [55].

5. *In implementing Green Analytical Chemistry*:

A significant challenge in many sectors, particularly those that incorporate chemistry and chemometrics, is striking a balance between sustainability and analytical precision. Requirements for a high number of samples or successive measurements in conventional analytical techniques, like GC, HPLC, can result in a considerable amount of waste [49]. Furthermore, energy-intensive devices that must operate for prolonged periods are frequently required for precise, high-throughput analysis, particularly in fields such as chromatography, mass spectrometry, and NMR. The validity and accuracy of analytical data must be maintained while attempting to employ fewer materials or shorten the duration of the experiment [48].

Green analytical chemistry seeks to optimize analytical procedures and minimize or eliminate the use of hazardous solvents and reagents, thereby contributing to environmental protection. This objective is supported by the incorporation of chemometric approaches into green analytical chemistry, which facilitates improved data processing & deduction, resulting in more effective analytical procedures without sacrificing accuracy of results [56].

Methods like multivariate ranking approaches like PCA, CA and ANNs, are used to pick green materials or processes, cope with missing data, and group or classify information. Finding sustainable solutions for solvents, reagents, processes, or process conditions is primary application area. To find best conditions for sample preparation & analysis, for instance, RSM is often used, which minimizes consumption of reagents and solvents while optimizing analytical output [57].

Chemometrics has greatly enhanced analytical capabilities of spectroscopic techniques like UV-Vis, FTIR, & fluorescence spectroscopy while consuming least number of resources. These developments make it possible to detect and measure environmental pollutants more efficiently and sustainably [58]. Chemometric modelling is used in conjunction with techniques like Solid Phase Microextraction (SPME) & QuEChERS (Quick, Easy, Cheap, Effective, Rugged and Safe) to optimize extraction procedures [59]. Main parameters of QuEChERS, which is a sample preparation technique used to remove biocide contaminants off food items before to analysis, can be modified to increase extraction effectiveness [60]. Scientists can maximize QuEChERS parameters, including acetonitrile volume, acidity, and quantities of primary secondary amine (PSA) and C18 employed in the d-SPE stage, by using chemometric techniques [61]. To identify essential features and maximize extraction as well as cleanup studies, experimental designs, including Derringer's desire function (DF), central composite designs (CCD), and Plackett-Burman (P-B) screening models, are employed. The aforementioned experimental designs lower costs and reduce the number of tests needed while allowing simultaneous analysis of multiple parameters. These techniques lessen their influence on the environment by using few solvents and straightforward processes [62].

There are still difficulties, even though the integration of chemometrics into green analytical chemistry has several benefits. One issue that needs to be addressed is the possible loss of analytical precision when simplifying sample preparation. Educational initiatives are needed to give analysts the know-how to make efficient use of these cutting-edge methods. The significance of chemometrics in advancing sustainability within analytical chemistry may be strengthened in future with addition of ML & AI capabilities to further expedite data analysis procedures [63].

6. *In Drug Discovery*:

Specialized metabolites found in plants, fungi, animals & bacteria are known as natural products (NPs) and they have played a significant part in discovery and creation of novel medications. Chemometric approaches have been employed extensively in NP investigations recently, including chemotaxonomy, identifying geographical locations of medicinal plants as well as herbal goods, detecting bioactive NPs in medicinal plants, and quality control. Incorporating chemometrics into several phases of drug discovery reduces expenses and time while improving efficacy and efficiency of procedures including lead authentication, optimization, & validation. Chemometrics streamlines entire drug development pipeline by utilizing sophisticated statistical techniques, ML & data analysis utilities to help create safer and more potent pharmaceutical molecules [64].

In lead identification potential drug candidates are ranked according to their anticipated interactions with biological targets based on chemometrics. Data mining approaches can find chemicals that are likely to display desirable biological activity by building networks based upon chemical similarity. Prediction

capacities for lead detection are enhanced by sophisticated machine learning (ML) algorithms, such as deep learning (DL) and network propagation techniques. Employing known interactions & structural analogies, these algorithms effectively search through vast datasets to identify compounds exhibiting a high affinity towards target proteins [65].

Chemometric methods can be used as a supplement to bioassay-guided separation and may help lower the quantity of biological material & extraction solvents utilized. Another use is the isolation of NPs from plants. This is in addition to the implementation of chemometric methods to facilitate the detection of NPs responsible for the physiological activity of plant extracts, as well as the development of predictive algorithms for these physiological activities. Chemometric methods can be used to predict a drug's ADMET (absorption, distribution, metabolism, excretion, and toxicity) characteristics. Chemometrics enhances assay data processing, facilitating the identification of "hits" that could potentially develop into leads. Researchers can increase hit rate and find more promising drug candidates by using consensus scoring, which combines different screening techniques [66].

Chemometrics is utilized not just to identify drug candidates but also to develop techniques and refine processes that increase the efficiency of drug discovery. RSM has been implemented in medication formulation and production processes to optimize experimental circumstances. To meet regulatory requirements and maximize efficiency, chemometrics helps validate and ensure the quality of analytical techniques used in drug research and manufacturing [67].

Chemometric approaches provide the tools necessary to enhance the separation of bioactive nanoparticles (NPs) derived from plants or alternative sources. Furthermore, fewer plant materials and solvents are needed for the extraction of chemometrically amplified bioactive NPs, saving time and money. Therefore, the usage of chemometric algorithms can become a prominent tool to optimise the detection & separation of bioactive NPs [68].

7. *In Disease Screening*:

Chemometrics plays a crucial role in disease screening by enhancing the sensitivity, specificity, and reliability of diagnostic tests. By analyzing biological samples such as blood, saliva, or urine, chemometric techniques enable researchers to identify potential biomarkers. Integration of data from many analytical methods, including mass spectrometry, chromatography, & spectroscopy, was made easier by chemometrics [69]. When paired with chemometrics, vibrational spectroscopy, including attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy, has the potential to be applied in a range of clinical contexts. The preferred technique for analyzing biological samples that have been tampered with is principal component analysis (PCA). PCA scores are readily used to identify specific biomarkers characteristic of the infection class and to investigate overall dataset variation and any grouping associated with the limit of detection. Numerous cancer forms can now be diagnosed using chemometric approaches. For example, differentially expressed proteins within colorectal cancer (CRC) tissues have been found using chemometrics to differentiate between malignant and healthy cells. Chemometric methods such as PCA alongside LDA (PCA-LDA) are used to evaluate how well these proteins detect the presence of cancer in tissues. Chemometric processing may successfully distinguish between malignant and non-cancerous samples when applied to urine profiles acquired using GC-MS [70].

Chemometric algorithms are also essential in the development of computerized deductions for cardiovascular magnetic resonance (CMR) info, which helps in the early detection of heart diseases in various populations. Chemometrics has been employed in cardiovascular ailments research, especially for analyzing quinary treatment [71]. Additionally, chemometric approaches are utilized to examine variation in elemental concentrations within blood serum and also arterial wall specimens in aortic occlusive disease (AO) victims [72]. ATR-FTIR spectroscopy, paired with chemometrics, additionally provides a unique tool for diagnosing myocardial fibrosis. Certain microbial infections can also be identified using chemometrics. Chemometrics, for instance, is used to analyze biological fluids and diagnose SARS-CoV-2 infection [73].

Moreover, the consequences of chronic kidney disease (CKD) can be examined by an integrated evaluation of metabolomic data using chemometrics. Complicated data structures containing several groups & variables, as well as repeated assessments for specific individuals, can be handled via Network-PCA (NetPCA). This methodology enables researchers to incorporate each of the groups into a

comprehensive global framework for assessing their metabolic characteristics. Similar subsets of chemicals that describe advantageous metabolic effects of kidney transplantation and haemodialysis have been highlighted using integrative techniques [74].

Developments in chemometric analysis and vibrational spectroscopy have shown promise as non-invasive methods for diabetes screening. New biomarkers or biochemical elements that aid in the recognition of diabetes are frequently identified using infrared spectroscopy, particularly Fourier transform infrared (FT-IR) spectroscopy. To diagnose diabetes, experts have explored the use of infrared spectroscopy to more accurately detect high glucose levels in body fluids, such as blood and saliva. Chemometric techniques, such as hierarchical clustering analysis (HCA) in conjunction with principal component analysis (PCA) and linear discriminant analysis (LDA), have been utilized to accurately distinguish between samples that have and do not have diabetes. By efficiently monitoring markers like blood sugar concentrations & metabolic alterations, these tactics improve early disease identification and management [75].

High-dimensional data, ubiquitous in biological research, are often handled by methods such as PCA and PLS-DA. These methods facilitate more straightforward sample interpretation and classification by reducing dimensionality while preserving the overall variance of the dataset. Predictive modeling capabilities are enhanced when machine learning (ML) is integrated into chemometric frameworks. To improve the sensitivity and specificity of diagnostic testing, algorithms like Random Forest and Support Vector Machines (SVM) are used to classify samples based on intricate data patterns [76].

Chemometrics has benefits. However, there are drawbacks to using it for illness screening. Biological data has a high degree of dimensionality and variability, which can make analysis more challenging and potentially result in overfitting or incorrect interpretation of the data if not appropriately handled. The development of chemometrics and its application in various research and disease domains may be hindered by the lack of universally recognized procedures for data collection, processing, and analysis. Chemometrics is transforming the discipline of disease testing by making diagnostic procedures more precise, effective, and economical [69].

8. *In Forensic Science*:

By offering instruments that facilitate the management and interpretation of complex datasets used in forensic work, chemometrics serves as a bridge between analytical chemistry and statistical analysis. To help create connections amongst accused, victims & crime scenes, strategies are designed to enhance decision-making in forensic settings [77].

Numerous chemometric techniques can simplify the dimensionality of complicated data, which could uncover or clarify underlying patterns. Forensic scientists can improve analytical techniques and improve chemical sample segregation & resolution by employing chemometric technologies. Examining multivariate datasets produced by sophisticated methods like spectroscopy & chromatography is a common task in forensics [78]. There are several chemometric contributions to reliable evidence analysis in forensic science. By addressing the requirement for rational strategies to evidence interpretation, it improves deductive reasoning & optimization of analytical methodologies [79].

Chemometric models can quantify similarities among specimens from a crime scene with a suspect, particularly in circumstances involving small amounts of evidence, such as soil or glass, which may result in more conclusive links. In forensic blood examination, spectroscopic techniques like ATR FT-IR spectroscopy assist in recognizing important biological indicators, including proteins, lipids, and carbohydrates [80]. Drugs of abuse have been the subject of extensive research regarding forensic use of chemometrics, especially for batch comparisons and impurity profiling. The qualitative & quantitative evaluation of amphetamines, heroin, opiates, cocaine & other medications, or a blend of these, is being carried out frequently and efficiently using these algorithms [81].

Chemometric algorithms integrated into Raman spectroscopy using SVM models can be employed to differentiate between medicinal product families and illegal medicines; as a result, this technology can also be utilized to identify counterfeit alternatives. When used in conjunction with pattern recognition techniques (like PCA, HCA, PLS-DA, and ANN), NIR & Raman spectroscopy can effectively separate batches of counterfeit pharmaceutical products from authentic batches [82]. Several chemometric methodologies are employed in toxicology to evaluate seized drugs, identify unknown compounds, and

investigate materials associated with legal conflicts. PCA, Hierarchical Clustering Analysis (HCA), discriminant analysis, pattern recognition, and classification algorithms are typical chemometric strategies in toxicology. These strategies are frequently employed to analyze complex data from analytical approaches, such as mass spectrometry, chromatography, and spectroscopy, to identify unknown toxic compounds, detect patterns in exposure data, and assess potential toxicity based on chemical profiles. Chemical profiling and forensic toxicology are also important uses of chemometrics in toxicology. For example, cocaine and similar chemicals can be found in untreated oral fluids and other bodily fluids using a miniature micro-NIR spectrometer in conjunction with PCA and PLS-DA [83].

Thus, chemometrics has evolved as a potent instrument in forensic science and is anticipated to develop further in forensic applications in the future, possibly combining with cutting-edge technologies like machine learning (ML) and artificial intelligence (AI). These developments may improve chemometric models, enabling more complex studies that increase precision and speed of forensic investigations [84]. The various chemometric methods implemented across different applications of chemometrics in analytical chemistry are summarized in Table 1.

CHALLENGES FOR CHEMOMETRICS IN ANALYTICAL CHEMISTRY

Throughout analytical chemistry, chemometrics faces challenges in preserving data integrity, transferring models from laboratories to production locations, and navigating complex information networks. Small sample numbers and improper use of complex modelling tools are two common flaws in modern research employing proficient methodologies [85]. When using machine learning (ML) in chemistry, repeatability, along with data availability, is a crucial consideration. To implement a single spectrum database across several instruments, calibration transfer necessitates merging chemometric methods with instrument technology. This entails establishing multivariate calibration, analyzing reference samples, and applying that calibration to a different instrument [86]. When it comes to assessing high-dimensional data, chemometrics encounters several obstacles, including difficulties with feature selection, grouping, or anomaly recognition. Use of traditional chemometric methods may be complicated with high dimensionality, which may result in models that are impacted by several factors and make interpretation more difficult [87]. High-dimensional data makes it challenging to estimate correlation matrices, a regular task in both supervised and unsupervised learning, and requires a minimum sample quantity to guarantee practical estimation. Many statistical methods lose their effectiveness in high-dimensional contexts, and new approaches are needed to manage the amount and complexities of the data. The evaluation of high-dimensional data is made more difficult by the small number of reliable techniques that are now accessible [88].

FUTURE OPPORTUNITIES FOR CHEMOMETRICS IN ANALYTICAL CHEMISTRY

It is anticipated that chemometrics combined with AI, ML & DL would enhance statistical analysis & boost spectroscopic methods' accuracy and effectiveness. AI systems are able to decipher intricate spectroscopic data, spot trends, and accurately forecast molecular behaviour, which helps advance material science & biochemistry research. The futuristic analytical skills and predictive precision in chemometrics are expected to be improved by the creation and use of sophisticated ML, DL, and AI algorithms [89]. Pattern identification in complicated datasets is accomplished by methods like PLS-DA, LDA, along with PCA. Newer techniques, such as artificial neural networks (ANNs) alongside support vector machines (SVMs), are developing into strong instruments for complex modeling, making it easier to decipher large, complex datasets and uncover hidden patterns. In situations requiring trace evidence, like dirt or glass, chemometric models can numerically assess how similar samples from the site of the crime are to a suspect, perhaps resulting in more conclusive links.

A novel approach to mass loss prediction in cement samples is presented via integration of ML and Hyperspectral Imaging (HSI), achieving good validation scores using PLS [90]. Chemometrics is particularly well-positioned to take advantage of upcoming prospects in high-dimensional data processing because to its improved AI integration, wide range of applications, and sophisticated analytical methods. Chemometrics has opportunity to generate strong models and derive significant insights due to intricacy of high-dimensional data. Demand for advanced chemometric software & tools is growing as sectors pursue digital transformation, suggesting an increasing tendency towards more digitization in the region [91].

Table 1: Comparison table summarizing chemometric methods and their applications.

Areas of Application	Chemometric methods that have been implemented	Ref
1. Analysis of Phytochemical Profiles	1. PCA: Visualizing correlations and variance in phytochemical data.	[15]
	2. PLS: Spectral data is correlated with chemical profile or physiological activity.	[13]
	3. CA: Identifying species or chemotypes exhibiting comparable biological characteristics.	[17]
	4. LDA: Plant samples are categorized into predetermined groups.	[17]
2. In food science technology	1. PCA: Food composition can be understood.	[20]
	2. PLS-DA: To improve identification of genuine components in intricate food matrices.	[20]
	3. SIMCA: Food items can be categorized.	[24]
3. In quality assessment and analysis of medicines	1. PCA: Identifying differences between batches as well as identifying quality deviations.	[32]
	2. PLS: Create calibration models that link analytical signal data to concentration values for APIs.	[33]
4. In environmental analysis	1. PLS & PCA: These techniques, in conjugation with NIR and Vis-NIR are effective in assessing soil pH, total nitrogen, organic and total carbon, and predicting a variety of soil characteristics.	[42]
	2. SOM-ANN: Can be used to assess river and other water body pollutions by evaluating and estimating its water and sediment.	[47]
5. In implementing green analytical chemistry	1. PCA, CA & ANNs: Facilitate extraction of valuable details regarding chemical systems and lessen chemical wastage, time-consuming sample preparation.	[59]
	2. RSM: Find the best conditions for sample preparation & analysis which aids in minimising consumption of reagents and solvents while optimizing analytical output.	[63]
6. In drug discovery	1. RSM: Implemented in medication formulation and production processes to optimize experimental circumstances.	[68]
7. In disease screening	1. PCA: PCA scores are readily used to determine particular biomarkers characteristic about infection class and to investigate overall dataset variation and any grouping linked with limit of detection.	[69]
	2. LDA: LDA along with PCA are used to evaluate how well specific proteins detect presence of cancer in tissues.	[70]
	3. HCA: HCA along with PCA-LDA have been used to accurately distinguish among samples that have and do not have diabetes.	[73]
8. In forensic science	1. PCA, HCA, PLS-DA, ANN: These methods in conjunction with NIR and Raman spectroscopy can be effectively used to separate batches of counterfeit pharmaceutical products from authentic batches.	[78]

CONCLUSION

As technology and data analysis techniques continue to develop, chemometrics is expected to become increasingly important in analytical chemistry. By adding non-linear interactions and making it easier to find intricate patterns in datasets that would be difficult to find with traditional techniques, AI improves classic chemometric models [92]. Artificial intelligence (AI) and machine learning (ML) hold tremendous potential in advancing chemometrics. One of the potential breakthroughs that AI and ML could bring to chemometrics is predictive modelling for chemical properties. Large datasets of known chemical characteristics can be used to create prediction models using machine learning (ML). For instance, by utilizing historical chemical data, supervised learning methods such as Random Forests and SVMs can predict molecular characteristics [93]. For optimal efficiency, process conditions can be continually adjusted through the use of reinforcement learning (RL). Additionally, DL models may be programmed to automatically evaluate data from various methods (such as NMR, IR, and GC-MS) and produce valuable insights. With the use of ML

techniques, chemical libraries may be virtually screened to determine compounds that have the best chance of succeeding [94]. By examining spectral data to identify adulteration and guarantee adherence to safety regulations, chemometric methods streamline quality control procedures. More reliable and adaptable analytical methods will probably be developed as chemometrics advances in the future. Using MA, more complex data modeling will be made possible by computerized methods, which will ultimately improve the extraction of important chemical information while reducing the time and expenses associated with conventional analysis. As the area develops, chemometrics is expected to play a key role in tackling current analytical procedure issues by providing creative solutions that maximize chemical insights and data analysis [95]. Although chemometrics has found its application in various sectors of analytical chemistry, there remain a few areas that continue to stay underexplored, like gene expression profiling, metabolic engineering, astronomical spectroscopy, planetary exploration, cultural heritage, and archaeometry (like Dating Techniques). Researchers should actively collaborate with domain experts (e.g., chemists, biologists, engineers) to ensure that chemometric models are correctly aligned with the real-world challenges of specific fields. Collaborative efforts help integrate domain knowledge into the modelling process, resulting in more accurate and relevant models. They should tailor chemometric techniques to the specific needs of different application areas, such as food quality control, pharmaceutical analysis, or environmental monitoring. They must explore how chemometrics can be applied to sustainability challenges, such as reducing waste, optimizing energy consumption, or enhancing the efficiency of chemical processes. For e.g., models that predict optimal reaction conditions in chemical manufacturing could lead to more sustainable and cost-effective processes.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

FINANCIAL ASSISTANCE

NIL

AUTHOR CONTRIBUTION

Bibhas Pandit conceptualized the article's original idea. Projesh Saha performed the literature review. Soma Pramanik contributed to the data compilation and manuscript preparation. All the activities were done under the guidance and supervision of Bhupendra Shrestha.

REFERENCES

[1] Wold S. Chemometrics and Intelligent Laboratory Systems. *Chemometrics and intelligent laboratory systems*, **30**, 109–15 (1995) https://doi.org/10.1016/0169-7439(95)00042-9.

[2] Riu J, Giussani B. Analytical chemistry meets art: The transformative role of chemometrics in cultural heritage preservation. *Chemometrics and Intelligent Laboratory Systems*, **247**, (2024) https://doi.org/10.1016/j.chemolab.2024.105095.

[3] Santos MC, Nascimento PAM, Guedes WN, Pereira-Filho ER, Filletti ÉR, Pereira FMV. Chemometrics in analytical chemistry - An overview of applications from 2014 to 2018. *Ecletica Quimica*, **44**, 11–25 (2019) https://doi.org/10.26850/1678-4618eqj.v44.2.11-25.

[4] Parastar H, Tauler R. Big (Bio)Chemical Data Mining Using Chemometric Methods: A Need for Chemists. *Angew Chem Int Ed Engl*, **61**, e201801134 (2022) https://doi.org/10.1002/anie.201801134.

[5] Harlina PW, Maritha V, Geng F, Nawaz A, Yuliana T, Subroto E, Dahlan HJ, Lembong E, Huda S. Comprehensive review on the application of omics analysis coupled with Chemometrics in gelatin authentication of food and pharmaceutical products. *Food Chem X*, **23**, (2024) https://doi.org/10.1016/j.fochx.2024.101710.

[6] Martynko E, Kirsanov D. Application of Chemometrics in Biosensing: A Brief Review. *Biosensors (Basel)*, **10**, (2020) https://doi.org/10.3390/bios10080100.

[7] Kharbach M, Alaoui Mansouri M, Taabouz M, Yu H. Current Application of Advancing Spectroscopy Techniques in Food Analysis: Data Handling with Chemometric Approaches. *Foods*, **12**, (2023) https://doi.org/10.3390/foods12142753.

[8] Taylan O, Cebi N, Tahsin Yilmaz M, Sagdic O, Bakhsh AA. Detection of lard in butter using Raman spectroscopy combined with chemometrics. *Food Chem*, **332**, (2020) https://doi.org/10.1016/j.foodchem.2020.127344.

[9] Elhamdaoui O, El Orche A, Cheikh A, Mojemmi B, Nejjari R, Bouatia M. Development of Fast Analytical Method for the Detection and Quantification of Honey Adulteration Using Vibrational Spectroscopy and Chemometrics Tools. *J Anal Methods Chem*, **2020**, (2020) https://doi.org/10.1155/2020/8816249.

[10] Peris-Díaz MD, Krężel A. A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry. *TrAC - Trends in Analytical Chemistry*, **135**, (2021) https://doi.org/10.1016/j.trac.2020.116157.

[11] Tortorella S, Cinti S. How Can Chemometrics Support the Development of Point of Need Devices? *Anal Chem*, **93**, 2713–22 (2021) https://doi.org/10.1021/acs.analchem.0c04151.

[12] James G, Witten D, Hastie T, Tibshirani R, Taylor J. Linear Model Selection and Regularization. 229–88 (2023) https://doi.org/10.1007/978-3-031-38747-0_6.

[13] Rai AK, Khan S, Kumar A, Dubey BK, Lal RK, Tiwari A, Trivedi PK, Elliott CT, Ratnasekhar C. Comprehensive Metabolomic Fingerprinting Combined with Chemometrics Identifies Species- and Variety-Specific Variation of Medicinal Herbs: An Ocimum Study. *Metabolites*, **13**, (2023) https://doi.org/10.3390/metabo13010122.

[14] Moshari-Nasirkandi A, Iaccarino N, Romano F, Graziani G, Alirezalu A, Alipour H, Amato J. Chemometrics-based analysis of the phytochemical profile and antioxidant activity of Salvia species from Iran. *Sci Rep*, **14**, (2024) https://doi.org/10.1038/s41598-024-68421-8.

[15] Pollo BJ, Teixeira CA, Belinato JR, Furlan MF, Cunha IC de M, Vaz CR, Volpato GV, Augusto F. Chemometrics, Comprehensive Two-Dimensional gas chromatography and "omics" sciences: Basic tools and recent applications. *TrAC - Trends in Analytical Chemistry*, **134**, (2021) https://doi.org/10.1016/j.trac.2020.116111

[16] de Falco B, Grauso L, Fiore A, Bonanomi G, Lanzotti V. Metabolomics and chemometrics of seven aromatic plants: Carob, eucalyptus, laurel, mint, myrtle, rosemary and strawberry tree. *Phytochemical Analysis*, **33**, 696–709 (2022) https://doi.org/10.1002/pca.3121.

[17] Abraham EJ, Kellogg JJ. Chemometric-Guided Approaches for Profiling and Authenticating Botanical Materials. *Front Nutr*, **8**, (2021) https://doi.org/10.3389/fnut.2021.780228.

[18] Aleixandre-Tudo JL, Castello-Cogollos L, Aleixandre JL, Aleixandre-Benavent R. Chemometrics in food science and technology: A bibliometric study. *Chemometrics and Intelligent Laboratory Systems*, **222**, (2022) https://doi.org/10.1016/j.chemolab.2022.104514.

[19] González-Domínguez R, Sayago A, Fernández-Recamales Á. An Overview on the Application of Chemometrics Tools in Food Authenticity and Traceability. *Foods*, **11**, (2022) https://doi.org/10.3390/foods11233940.

[20] Farag MA, Sheashea M, Zhao C, Maamoun AA. UV Fingerprinting Approaches for Quality Control Analyses of Food and Functional Food Coupled to Chemometrics: A Comprehensive Analysis of Novel Trends and Applications. *Foods*, **11**, (2022) https://doi.org/10.3390/foods11182867.

[21] Andre CM, Soukoulis C. Food quality assessed by chemometrics. *Foods*, **9**, (2020) https://doi.org/10.3390/foods9070897.

[22] Pelissari EMR, Covre KV, do Rosario DKA, de São José JFB. Application of chemometrics to assess the influence of ultrasound and chemical sanitizers on vegetables: Impact on natural microbiota, Salmonella Enteritidis and physicochemical nutritional quality. *LWT*, **148**, (2021) https://doi.org/10.1016/j.lwt.2021.111711.

[23] Vasconi M, Tirloni E, Stella S, Coppola C, Lopez A, Bellagamba F, Bernardi C, Moretti VM. Comparison of Chemical Composition and Safety Issues in Fish Roe Products: Application of Chemometrics to Chemical Data. https://doi.org/10.3390/foods9040540.

[24] Voccio R, Malegori C, Oliveri P, Branduani F, Arimondi M, Bernardi A, Luciano G, Cettolin M. Combining PLS-DA and SIMCA on NIR data for classifying raw materials for tyre industry: A hierarchical classification model. *Chemometrics and Intelligent Laboratory Systems*, **250**, (2024) https://doi.org/10.1016/j.chemolab.2024.105150.

[25] Fan Z, Jiawei Z, Jihao Z. Adulteration identification of astragalus polysaccharides by nir spectroscopy combined with simca and PLS-DA. *INMATEH - Agricultural Engineering*, **68**, 827–34 (2022) https://doi.org/10.35633/inmateh-68-82.

[26] Teixeira JLDP, Caramês ETDS, Baptista DP, Gigante ML, Pallone JAL. Adulteration Detection in Goat Dairy Beverage Through NIR Spectroscopy and DD-SIMCA. *Food Anal Methods*, **15**, 783–91 (2022) https://doi.org/10.1007/s12161-021-02151-9.

[27] De Angelis D, Summo C, Pasqualone A, Faccia M, Squeo G. Advancements in food authentication using soft independent modelling of class analogy (SIMCA): a review. *Food Quality and Safety*, (2024) https://doi.org/10.1093/fqsafe/fyae032.

[28] Grassi S, Tarapoulouzi M, D'alessandro A, Agriopoulou S, Strani L, Varzakas T. How Chemometrics Can Fight Milk Adulteration. (2022) https://doi.org/10.3390/foods.

[29] Squeo G, Cruz J, De Angelis D, Caponio F, Amigo JM. Considerations about the gap between research in near-infrared spectroscopy and official methods and recommendations of analysis in foods. *Curr Opin Food Sci*, **59**, (2024) https://doi.org/10.1016/j.cofs.2024.101203.

[30] Rajendran HK, Fakrudeen MAD, Chandrasekar R, Silvestri S, Sillanpää M, Padmanaban VC. A comprehensive review on analytical and equation derived multivariate chemometrics for the accurate interpretation of the degradation of aqueous contaminants. *Environ Technol Innov*, **28**, (2022) https://doi.org/10.1016/j.eti.2022.102827.

[31] Li Y, Shen Y, Yao C liang, Guo D an. Quality assessment of herbal medicines based on chemical fingerprints combined with chemometrics approach: A review. *J Pharm Biomed Anal*, **185**, (2020) https://doi.org/10.1016/j.jpba.2020.113215.

[32] Liu Z, Yang MQ, Zuo Y, Wang Y, Zhang J. Fraud Detection of Herbal Medicines Based on Modern Analytical Technologies Combine with Chemometrics Approach: A Review. *Crit Rev Anal Chem*, 52, 1606–23 (2022) https://doi.org/10.1080/10408347.2021.1905503.

[33] Lucini L, Ganugi P, Gajdoš Kljusuri J, Mastinu A, Xia L, Qi L. Integrative quantitative and qualitative analysis for the quality evaluation and monitoring of Danshen medicines from different

sources using HPLC-DAD and NIR combined with chemometrics. *Front Plant Sci*, **13**, 1–16 (2022) https://doi.org/10.3389/fpls.2022.93285.

[34] Vora LK, Gholap AD, Jetha K, Thakur RRS, Solanki HK, Chavda VP. Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design. *Pharmaceutics*, **15**, (2023) https://doi.org/10.3390/pharmaceutics15071916.

[35] Pantanowitz L, Hanna M, Pantanowitz J, Lennerz J, Henricks WH, Shen P, Quinn B, Bennet S, Rashidi HH. Regulatory Aspects of Artificial Intelligence and Machine Learning. *Modern Pathology*, **37**, (2024) https://doi.org/10.1016/j.modpat.2024.100609.

[36] Rebiai A, Seghir B Ben, Hemmami H, Zeghoud S, Amor I Ben, Kouadri I, Messaoudi M, Pasdaran A, Caruso G, Sharma S, Atanassova M, Pohl P. Quality Assessment of Medicinal Plants via Chemometric Exploration of Quantitative NMR Data: A Review. *Compounds*, **2**, 163–81 (2022) https://doi.org/10.3390/compounds2020012.

[37] Rehman A, Naz S, Razzak I. Leveraging Big Data Analytics in Healthcare Enhancement: Trends, Challenges and Opportunities. *Academic medicine: journal of the As- sociation of American Medical Colleges*, 1–24 (2020) https://doi.org/10.1007/s00530-020-00736-8.

[38] Casian T, Nagy B, Kovács B, Galata DL, Hirsch E, Farkas A. Challenges and Opportunities of Implementing Data Fusion in Process Analytical Technology—A Review. *Molecules*, **27**, (2022) https://doi.org/10.3390/molecules27154846.

[39] Tayefi M, Ngo P, Chomutare T, Dalianis H, Salvi E, Budrionis A, Godtliebsen F. Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdiscip Rev Comput Stat*, **13**, (2021) https://doi.org/10.1002/wics.1549.

[40] Nelly Tochi Nwosu. Reducing operational costs in healthcare through advanced BI tools and data integration. *World Journal of Advanced Research and Reviews*, **22**, 1144–56 (2024) https://doi.org/10.30574/wjarr.2024.22.3.1774.

[41] Liu Z, Yang MQ, Zuo Y, Wang Y, Zhang J. Fraud Detection of Herbal Medicines Based on Modern Analytical Technologies Combine with Chemometrics Approach: A Review. *Crit Rev Anal Chem*, **52**, 1606–23 (2022) https://doi.org/10.1080/10408347.2021.1905503.

[42] Dupont MF, Elbourne A, Cozzolino D, Chapman J, Truong VK, Crawford RJ, Latham K. Chemometrics for environmental monitoring: A review. *Analytical Methods*, 12, 4597–620 (2020) https://doi.org/10.1039/d0ay01389g.

[43] Barra I, Kharbach M, Qannari EM, Hanafi M, Cherrah Y, Bouklouze A. Predicting cetane number in diesel fuels using FTIR spectroscopy and PLS regression. *Vib Spectrosc*, **111**, (2020) https://doi.org/10.1016/j.vibspec.2020.103157.

[44] Wilcox TM, McKelvey KS, Young MK, Engkjer C, Lance RF, Lahr A, Eby LA, Schwartz MK. Parallel, targeted analysis of environmental samples via high-throughput quantitative PCR. *Environmental DNA*, **2**, 544–53 (2020) https://doi.org/10.1002/edn3.80.

[45] Barra I, Haefele SM, Sakrabani R, Kebede F. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. *TrAC - Trends in Analytical Chemistry*, **135**, (2021) https://doi.org/10.1016/j.trac.2020.116166.

[46] Krzebietke S, Daszykowski M, Czarnik-Matusewicz H, Stanimirova I, Pieszczek L, Sienkiewicz S, Wierzbowska J. Monitoring the concentrations of Cd, Cu, Pb, Ni, Cr, Zn, Mn and Fe in cultivated Haplic Luvisol soils using near-infrared reflectance spectroscopy and chemometrics. *Talanta*, **251**, (2023) https://doi.org/10.1016/j.talanta.2022.123749.

[47] Athamena A, Gaagai A, Aouissi HA, Burlakovs J, Bencedira S, Zekker I, Krauklis AE. Chemometrics of the Environment: Hydrochemical Characterization of Groundwater in Lioua Plain (North Africa) Using Time Series and Multivariate Statistical Analysis. *Sustainability (Switzerland)*, **15**, (2023) https://doi.org/10.3390/su15010020.

[48] Fuente-Ballesteros A, Ares AM, Bernal J. Paving the way towards green contaminant analysis: Strategies and considerations for sustainable analytical chemistry. *Green Analytical Chemistry*, **12**, (2025) https://doi.org/10.1016/j.greeac.2025.100221.

[49] Aboushady D, Samir L, Masoud A, Elshoura Y, Mohamed A, Hanafi RS, El Deeb S. Chemometric Approaches for Sustainable Pharmaceutical Analysis Using Liquid Chromatography. *Chemistry (Easton)*, **7**, 11 (2025) https://doi.org/10.3390/chemistry7010011.

[50] Zhao X, Cheng K, Zhou W, Cao Y, Yang SH. Multivariate Statistical Analysis for the Detection of Air Pollution Episodes in Chemical Industry Parks. *Int J Environ Res Public Health*, **19**, (2022) https://doi.org/10.3390/ijerph19127201.

[51] Chapman J, Truong VK, Elbourne A, Gangadoo S, Cheeseman S, Rajapaksha P, Latham K, Crawford RJ, Cozzolino D. Combining Chemometrics and Sensors: Toward New Applications in Monitoring and Environmental Analysis. *Chem Rev*, **120**, 6048–69 (2020) https://doi.org/10.1021/acs.chemrev.9b00616.

[52] Reitz A, Hemric E, Hall KK. Evaluation of a multivariate analysis modeling approach identifying sources and patterns of nonpoint fecal pollution in a mixed use watershed. *J Environ Manage*, **277**, (2021) https://doi.org/10.1016/j.jenvman.2020.111413.

[53] Muniz DHF, Oliveira-Filho EC. Multivariate Statistical Analysis for Water Quality Assessment: A Review of Research Published between 2001 and 2020. *Hydrology*, **10**, (2023) https://doi.org/10.3390/hydrology10100196.

[54] Machado M, Reisen VA, Santos JM, Reis Junior NC, Frère S, Bondon P, Ispány M, Aranda Cotta HH. Use of multivariate time series techniques to estimate the impact of particulate matter on the perceived annoyance. *Atmos Environ*, **222**, (2020) https://doi.org/10.1016/j.atmosenv.2019.117080.

[55] Oliveira LG, Araújo KC, Barreto MC, Bastos MEPA, Lemos SG, Fragoso WD. Applications of chemometrics in oil spill studies. *Microchemical Journal*, **166**, (2021) https://doi.org/10.1016/j.microc.2021.106216.

[56] Martinez-Mayorga K, Madariaga-Mazon A, Medina-Franco JL, Maggiora G. The impact of chemoinformatics on drug discovery in the pharmaceutical industry. *Expert Opin Drug Discov*, **15**, 293–306 (2020) https://doi.org/10.1080/17460441.2020.1696307.

[57] Kalinowska K, Bystrzanowska M, Tobiszewski M. Chemometrics approaches to green analytical chemistry procedure development. *Curr Opin Green Sustain Chem*, **30**, (2021) https://doi.org/10.1016/j.cogsc.2021.100498.

[58] Cornejo-Báez AA, Peña-Rodríguez LM, Álvarez-Zapata R, Vázquez-Hernández M, Sánchez-Medina A. Chemometrics: a complementary tool to guide the isolation of pharmacologically active natural products. *Drug Discov Today*, **25**, 27–37 (2020) https://doi.org/10.1016/j.drudis.2019.09.016.

[59] Gullifa G, Barone L, Papa E, Giuffrida A, Materazzi S, Risoluti R. Portable NIR spectroscopy: the route to green analytical chemistry. *Front Chem*, **11**, (2023) https://doi.org/10.3389/fchem.2023.1214825.

[60] Zhu X, Li W, Wu R, Liu P, Hu X, Xu L, Xiong Z, Wen Y, Ai S. Rapid detection of chlorpyrifos pesticide residue in tea using surface-enhanced Raman spectroscopy combined with chemometrics. *Spectrochim Acta A Mol Biomol Spectrosc*, **250**, (2021) https://doi.org/10.1016/j.saa.2020.119366.

[61] Shikweni M, Tavengwa NT, Mokgehle TM. Characterization and chemometric based optimization of bioactive metabolites in Hypoxis hemerocallidea with the aid of UPLC-QqQ-MS/MS. *Chemical Papers*, **78**, 2223–33 (2024) https://doi.org/10.1007/s11696-023-03232-1.

[62] Narenderan ST, Meyyanathan SN, Karri VVSR, Babu B, Chintamaneni P. Chemometrics Assisted QuEChERS Extraction Method for the Residual Analysis of Organophosphate Insecticides: Application to Their Dissipation Kinetics in Open Field Ecosystem. *Analytical Chemistry Letters*, **10**, 798–810 (2020) https://doi.org/10.1080/22297928.2021.1876575.

[63] Jurjeva J, Koel M. Implementing greening into design in analytical chemistry. *Talanta Open*, **6**, (2022) https://doi.org/10.1016/j.talo.2022.100136.

[64] Joshi PB. Navigating with chemometrics and machine learning in chemistry. *Artif Intell Rev*, **56**, 9089–114 (2023) https://doi.org/10.1007/s10462-023-10391-w.

[65] dos Santos DP, Sena MM, Almeida MR, Mazali IO, Olivieri AC, Villa JEL. Unraveling surface-enhanced Raman spectroscopy results through chemometrics and machine learning: principles, progress, and trends. *Anal Bioanal Chem*, **415**, 3945–66 (2023) https://doi.org/10.1007/s00216-023-04620-y.

[66] Benjelloun M, Miyah Y, Evrendilek GA, Lalami AEO, Demir I, Atmaca B, Ssouni S, Lairini S, Bouslamti R. Synergistic effect of coupling ozonation/adsorption system for toxic dye efficient removal: chemometric optimization by Box–Behnken response surface methodology. *Desalination Water Treat*, **306**, 220–35 (2023) https://doi.org/10.5004/dwt.2023.29821.

[67] Zulkifli B, Fakri F, Odigie J, Nnabuife L, Isitua CC, Chiari W. Chemometric-empowered spectroscopic techniques in pharmaceutical fields: A bibliometric analysis and updated review. *Narra X*, **1**, (2023) https://doi.org/10.52225/narrax.v1i1.80.

[68] Olasupo SB, Uzairu A, Shallangwa G, Uba S. QSAR modeling, molecular docking and ADMET/pharmacokinetic studies: a chemometrics approach to search for novel inhibitors of norepinephrine transporter as potent antipsychotic drugs. *Journal of the Iranian Chemical Society*, **17**, 1953–66 (2020) https://doi.org/10.1007/s13738-020-01902-5.

[69] Passos JOS, dos Santos Alves MV, Morais CLM, Martin FL, Cavalcante AF, Lemos TMAM, Moura S, Freitas DLD, Mariz JVM, Carvalho JL, Lima KMG, Pegado R. Spectrochemical analysis in blood plasma combined with subsequent chemometrics for fibromyalgia detection. *Sci Rep*, **10**, (2020) https://doi.org/10.1038/s41598-020-68781-x.

[70] Martin FL, Dickinson AW, Saba T, Bongers T, Singh MN, Bury D. ATR-FTIR Spectroscopy with Chemometrics for Analysis of Saliva Samples Obtained in a Lung-Cancer-Screening Programme: Application of Swabs as a Paradigm for High Throughput in a Clinical Setting. *J Pers Med*, **13**, (2023) https://doi.org/10.3390/jpm13071039.

[71] Elsonbaty A, Serag A, Abdulwahab S, Hassan WS, Eissa MS. Analysis of quinary therapy targeting multiple cardiovascular diseases using UV spectrophotometry and chemometric tools. *Spectrochim Acta A Mol Biomol Spectrosc*, **238**, (2020) https://doi.org/10.1016/j.saa.2020.118415.

[72] Calvo-Gomez O, Calvo H, Cedillo-Barrón L, Vivanco-Cid H, Alvarado-Orozco JM, Fernandez-Benavides DA, Arriaga-Pizano L, Ferat-Osorio E, Anda-Garay JC, López-Macias C, López MG. Potential of ATR-FTIR-Chemometrics in Covid-19: Disease Recognition. *ACS Omega*, **7**, 30756–67 (2022) https://doi.org/10.1021/acsomega.2c01374.

[73] Pushpa SR, Sukumaran RK, Savithri S. Robustness of FTIR-Based Ultrarapid COVID-19 Diagnosis Using PLS-DA. *ACS Omega*, **7**, 47357–71 (2022) https://doi.org/10.1021/acsomega.2c06786.

[74] Boccard J, Schvartz D, Codesido S, Hanafi M, Gagnebin Y, Ponte B, Jourdan F, Rudaz S. Gaining Insights Into Metabolic Networks Using Chemometrics and Bioinformatics: Chronic

Kidney Disease as a Clinical Model. *Front Mol Biosci*, **8**, (2021) https://doi.org/10.3389/fmolb.2021.682559.

[75] Senger RS, Sayed Issa A, Agnor B, Talty J, Hollis A, Robertson JL. Disease-Associated Multimolecular Signature in the Urine of Patients with Lyme Disease Detected Using Raman Spectroscopy and Chemometrics. *Appl Spectrosc*, **76**, 284–99 (2022) https://doi.org/10.1177/00037028211061769.

[76] Risoluti R, Caprari P, Gullifa G, Massimi S, Maffei L, Sorrentino F, Carcassi E, Materazzi S. An Innovative Multilevel Test for Hemoglobinopathies: TGA/Chemometrics Simultaneously Identifies and Classifies Sickle Cell Disease From Thalassemia. *Front Mol Biosci*, **7**, (2020) https://doi.org/10.3389/fmolb.2020.00141.

[77] Cano-Trujillo C, García-Ruiz C, Ortega-Ojeda FE, Romolo F, Montalvo G. Forensic analysis of biological fluid stains on substrates by spectroscopic approaches and chemometrics: A review. *Anal Chim Acta*, **1282**, (2023) https://doi.org/10.1016/j.aca.2023.341841.

[78] Singh S, Shakeel H, Sharma R. Overview of chemometrics in forensic toxicology. *Egypt J Forensic Sci*, **13**, (2023) https://doi.org/10.1186/s41935-023-00371-0.

[79] Huhtala S, Nordgaard A, Ahrens B, Alberink I, Korpinsalo T, Bovens M. Chemometrics in Forensic Chemistry – Part III: Quality assessment and interpretation of chemometric output. *Forensic Sci Int*, **348**, (2023) https://doi.org/10.1016/j.forsciint.2023.111612.

[80] Zhang J, Jiang H, Duan B, Liu F. A rapid and nondestructive approach for forensic identification of cigarette inner liner papers using shift-excitation Raman difference spectroscopy and chemometrics. *J Forensic Sci*, **66**, 2180–9 (2021) https://doi.org/10.1111/1556-4029.14798.

[81] Lavine BK. Chemometrics in forensic science. *J Chemom*, **35**, (2021) https://doi.org/10.1002/cem.3322.

[82] Takamura A, Ozawa T. Recent Advances of Vibrational Spectroscopy and Chemometrics for Forensic Biological Analysis. *Analyst*, 1–35 (2021) https://doi.org/10.1039/D1AN01637G.

[83] Eliaerts J, Meert N, Dardenne P, Baeten V, Pierna JAF, van Durme F, de Wael K, Samyn N. Comparison of spectroscopic techniques combined with chemometrics for cocaine powder analysis. *J Anal Toxicol*, **44**, 851–60 (2020) https://doi.org/10.1093/jat/bkaa101.

[84] Sauzier G, Van Bronswijk W, Lewis SW. Chemometrics in Forensic Science: Approaches and Applications. *Forensic Chem*, 16–47 (2022) https://doi.org/10.1039/D1AN00082A.

[85] Pfeiffer P, Filzmoser P. Robust statistical methods for high-dimensional data, with applications in tribology. *Anal Chim Acta*, **1279**, (2023) https://doi.org/10.1016/j.aca.2023.341762.

[86] Merone GM, Tartaglia A, Locatelli M, D'Ovidio C, Rosato E, de Grazia U, Santavenere F, Rossi S, Savini F. Analytical Chemistry in the 21st Century: Challenges, Solutions, and Future Perspectives of Complex Matrices Quantitative Analyses in Biological/Clinical Field. *Analytica*, **1**, 44–59 (2020) https://doi.org/10.3390/analytica1010006.

[87] Wang J. Biostatistical Challenges in High-Dimensional Data Analysis: Strategies and Innovations. *Computational Molecular Biology*, (2024) https://doi.org/10.5376/cmb.2024.14.0019.

[88] Auddy A, Xia D, Yuan M. Tensor Methods in High Dimensional Data Analysis: Opportunities and Challenges. *A Journal of the IMA*, 1–30 (2024) https://doi.org/10.48550/arXiv.2405.18412.

[89] Mishra P, Roger JM, Jouan-Rimbaud-Bouveresse D, Biancolillo A, Marini F, Nordon A, Rutledge DN. Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC - Trends in Analytical Chemistry*, **137**, (2021) https://doi.org/10.1016/j.trac.2021.116206.

[90] Biancolillo A, D'Archivio AA, Marini F, Vitale R. Novel Applications of Chemometrics in Analytical Chemistry and Chemical Process Industry. *Frontiers (Boulder)*, 6–41 (2022) https://doi.org/10.3389/978-2-88976-297-2.

[91] Bărbulescu A, Barbe L. Challenges and Opportunities in the Application of Chemometrics in the Pharmaceutical and Food Science Industries. *J Chem*, **2022**, (2022) https://doi.org/10.1155/2022/9823497.

[92] Caratti A, Squara S, Bicchi C, Liberto E, Vincenti M, Reichenbach SE, Tao Q, Geschwender D, Alladio E, Cordero C. Boosting comprehensive two-dimensional chromatography with artificial intelligence: Application to food-omics. *TrAC - Trends in Analytical Chemistry*, **174**, (2024) https://doi.org/10.1016/j.trac.2024.117669.

[93] Ayres LB, Gomez FJV, Linton JR, Silva MF, Garcia CD. Taking the leap between analytical chemistry and artificial intelligence: A tutorial review. *Anal Chim Acta*, **1161**, (2021) https://doi.org/10.1016/j.aca.2021.338403.

[94] Jia W, Georgouli K, Martinez-Del Rincon J, Koidis A. Challenges in the Use of AI-Driven Non-Destructive Spectroscopic Tools for Rapid Food Analysis. *Foods*, **13**, (2024) https://doi.org/10.3390/foods13060846.

[95] Massei A, Cavallini N, Savorani F, Falco N, Fissore D. Exploring NIR spectroscopy data: A practical chemometric tutorial for analyzing freeze-dried pharmaceutical formulations. *Chemometrics and Intelligent Laboratory Systems*, **257**, (2025) https://doi.org/10.1016/j.chemolab.2024.105291.