American Journal of Biomedicine and Pharmacy
https://biojournals.us/index.php/AJBP
ISSN: 2997-7177

Multivariate Data Analysis in Analytical Chemistry Using Artificial Intelligence Techniques (Article Review)

Ahmed Moussa Khalil
College of Pharmacy, University of Kirkuk, Kirkuk, Iraq

Received: 2025, 15, Jan
Accepted: 2026, 21, Feb
Published: 2026, 03, Mar

Annotation: In recent years, artificial intelligence (AI) and machine learning (ML) have revolutionized multivariate data analysis in analytical chemistry. Traditional chemometric analyses (e.g., PCA, PLSR) remain essential methods for dimension reduction and exploratory investigation. Nevertheless, AI methods like SVM (support vector machine), Random Forest model, Deep Learning models can enhance the prediction performance automatically extract features and process non-linear more complex datasets effectively. Applications are extremely diverse ranging from spectroscopy, chromatography, electrochemical sensors to pharmaceutical quality control and environmental monitoring. “Preprocessing and feature engineering are very important for detection quality of the model. This article aims to review the recent evolution of, applications for and synergy between AI-based and classical chemometric tools towards contemporary analytical processes.

Keywords: Analytical Chemistry, Artificial Intelligence, Data Analysis, Chemometrics.

Introduction

The rapid emergence of powerful analytical instruments in modern chemistry meant the generation of intricate data in current times. Imaging such as spectroscopy (UV-Vis, IR, Raman, NMR), chromatography (HPLC, GC) and hyphenated tech niques (GC - MS, LC - MS) offer us

data in several dimensions as they provide multivariate data on each point of a sample by many correlative variables [1,2]. This type of high-dimensional data are notoriously difficult for standard univariate or simple multivariate techniques, and more sophisticated approaches are necessary to obtain useful chemical information [2].

These problems are classically addressed in the field of chemometrics, a set of mathematical and statistical approaches whose goal is to extract information from chemical data using dimensionality reduction, pattern recognition and quantification prediction techniques [3]. The classical methods like PCA, partial least squares regression (PLSR), linear discriminant analysis (LDA) and cluster analysis still are the flagship antonyms in this domain of MCDS [4]. Indeed, these techniques deal with complex data sets through shrinking the most important information that help chemists to classify samples and estimate the concentration of analytes [3].

However, as information becomes more organized and complex, it has also exposed some limitations of classical chemometric tools when dealing with non-linear links between variables, large amount of data or measure errors due to noise/incomplete measures. Artificial Intelligence (AI) and Machine Learning (ML) is the answer to addressing these limitations as it enables us to generate data-driven models, which can learn complex patterns from data without human involvement [5]. Several modeling algorithms including support vector machines (SVM), random forests, and artificial neural networks (ANNs) have been proven less arbitrary over various analyses while enhancing predictive power1–3 that could instantaneously provide in-built feature extraction system to users and handle intricate inter-variables relationships within the datasets [2,5]. The objective of this review is to provide a brief summary about AI methods in multivariate data analysis in analytical chemistry, their potential application into practice and before all advantages and real disadvantages.

2. Comparisons between classical chemometric methods and AI techniques

For decades, classical chemometric techniques were the workhorse of multivariate data analysis in analytical chemistry. These are intended for data reduction, pattern recognition and generating meaning from complicated spectroscopic, hyphenated and chromatographic datasets. PCA is amongst the most well-established techniques whereby correlated variables are represented in terms of a few uncorrelated principal components, which account for the largest variance in the data. PCA is well-known for visualization, outlier identification and preliminary analysis [6,7], Table 1.

Partial Least Squares Regression (PLSR), another classical approach, models relationships between predictors and responses in the presence of collinear variables or when the number of predictors exceeds observations [8]. Furthermore, for supervised and unsupervised classification of chemical samples, Linear Discriminant Analysis (LDA) and cluster analysis have been extensively used. Albeit being highly interpretable, these approaches often fail to account for complex relationships or perform efficiently in large and noisy datasets, which are prevalent in contemporary analytical chemistry [9,10].

Machine learning (ML) and artificial intelligence (AI) supplement classical chemometrics by enabling models to learn from data non heuristically. SVMs have been successful to perform high-dimensional classification where classes are separated by optimal hyperplanes (using kernel functions) [9, 10]. Random Forests is an ensemble learning approach which aggregates a number of decision trees and has been tested as valid alternative modelling technique for classifiers on noisy or large data-sets with high-dimensional features [11, 12].

Artificial Neural Networks (ANNs) and Deep Learning techniques have also been applied to “learn” intricate non-linear patterns, capturing hierarchical representations from raw data [14,15,16]. Convolutional Neural Networks (CNNs) have been used on spectral and chromatographic data for automatic feature learning, increasing the predictive accuracy especially on challenging datasets where classical chemometric techniques fall short [6,16].

The necessity of data preprocessing is one of the main differences from classical chemometric and AI approaches. Conventional approaches usually involve baseline correction, normalization and scaling to improve noise and interpretability. The AI and deep learning models can, a lot of times learn strong or robust features from raw data or little preprocessed in timely-manner thus cutting dependence on any manual preprocessing [17].

Table 1. Comparison between Classical Chemometric Methods and Artificial Intelligence Techniques

Feature	Classical Chemometrics	AI / Machine Learning
Linearity assumption	Often assumed	Not required
Feature extraction	Manual preprocessing needed	Automatic in many models
Handling noisy data	Limited	Robust with sufficient training data
Interpretability	High	Lower (“black box”)
Scalability	Limited for very large datasets	Highly scalable
Example methods	PCA, PLSR, LDA	SVM, Random Forest, ANN/CNN

3. Applications in Analytical Techniques

Artificial intelligence (AI) and machine learning (ML) are transforming analytical chemistry by improving the analysis of complex multivariate data with greater speed and quality in a variety of analytical instruments. AI-enabled models simplify fast pattern recognition, prediction, machine automation and interpretative analysis for spectroscopy, chromatography and electrochemical sensing as well as environmental or pharmaceutical analysis [18], Table 2.

3.1. Spectroscopic Analysis

Spectroscopic data types, such as UV-Vis, IR, Raman, NMR and mass spectrometry generate high dimensional datasets where conventional chemometric algorithms may not be able to detect subtle similarities. AI techniques, such as neural networks and other sophisticated classifiers, may be able to identify the relevant spectral features and patterns for successful chemical identification. The Raman spectroscopy work shows that AI-boosted methods are much more effective (in comparison to classic chemometric modeling) with respect to identifying components within mixtures in a relatively specific manner [19]. Among these methods we find deep learning, a subset of machine learning which has also been published to work well for signal pre-processing and peak picking in vibrational, chromatographic and spectrometric datasets [20].

3.2. Chromatographic Techniques

In chromatography (for example HPLC, GC or LC-MS), peaks often overlay and the retention information is multidimensional, in particular when working on complex mixtures. AI within MS offers the opportunity of automated peak picking and quantification, and deconvolution thereby increasing throughput with a higher quality of data than manual or classical statistical methods. The recent studies shows the such machine learning that helps to interpret chromatography data and in its turn simplifies revisions, i.e., makes analyses more elementary and diminishes human errors in pharmaceutical industry or environmental protection [21]. Machine learning approach is also used in retention time prediction and feature alignment (important for this study to obtain quantitative analysis robustness) [18].

3.3. Electrochemical and Sensor-Based Systems

This is because electrochemical sensors produce time-dependent current or potential electrode signals, which are challenging to predict except by employing advanced models. Machine learning strategies facilitate enhanced analyte classification, signal denoising and concentration profile prediction. Reviews demonstrate great promise of ML for EC biosensors and sensors to achieve sensitive detection of analytes in complex matrices with high accuracy [22]. These are

used to account for the non-linearity response and noise in electrochemical measurements.

3.4. Environmental and Pharmaceutical Monitoring

AI and ML are also utilized in environmental chemistry to detect contaminants in water, soil, and air by fetching a quick detection and classification of chemicals. According to a recent review on water and wastewater analysis, AI-assisted methods have been able to enhance chemical analysis, monitoring of pharmaceuticals and personal care products (PPCP) by helping dealing with the multivariate in both spectral and chromatographic data [23]. AI models are employed for identifying degradation products, authenticity checks and method development in complex formulations matrices in drug quality control applications [18].

3.5. Tools for Real-Time Prediction and Sensor Integration

Data processing in vitro AI-enabled predictive tools coupled with advanced sensors enable real-time data analysis throughout the laboratory process. Such techniques, which enable the direct analysis of multi-dimensional data base that is produced by sensors or analytic instruments thereby providing speed and accuracy as for example required in environmental monitoring, medical diagnostics, etc. It is the incorporation of sensors with AI which is critical not only for developing portable point-of care analytical systems to make fast and reliable diagnosis possible [24].

Table 2. Representative Applications of AI in Analytical Chemistry

Technique	AI/ML Method	Application	Key Advantage	References
Spectroscopy	Deep learning, Neural networks	Spectral feature extraction & classification	Reduces manual preprocessing; higher accuracy	[18,19,20]
Chromatography	Machine learning classifiers	Peak deconvolution & retention prediction	Handles complex peak patterns	[18,21]
Electrochemical sensors	ML regression & classification	Analyte detection & quantification	Noise tolerance; improved prediction	[22]
Environmental monitoring	ML models	Pollutant detection & classification	Fast, robust analysis	[23]
Sensor integration	AI prediction tools	Real-time data interpretation	Immediate, automated insights	[24]

4. Preprocessing and Featurization of data in Multivariate Analysis

In the so-very interesting (note sarcasm) multivariate data analysis world that we are discussing here, there are preprocessing and feature engineering things that decide how well our analytics models in their varied forms can predict/reproduce/predict-and-compare/tell stories on/interpret, what they are dealing with – depending on modelling preference. Raw analytical datasets such as spectroscopic signals, chromatographic profiles or sensor responses, are often corrupted by noise, baseline effects or pollutants which mask the useful chemical information. Appropriate feature extraction will filter out this raw measurements and make them cleaner and more compatible in datasets, improving feasibility of subsequent analysis accordingly, also the performance and reliability of models [25,26].

It is of great importance as it removes unwanted variance, normalizes the intensities to be on an equal scale and normalizes features so that each variable should contribute equally to computational models. It is the role of feature engineering as an adjunct to preprocessing that can be used to automatically select or create indicative representations of data, reduce space dimensionality and improve the signal-to-noise ratio. As discussed in analytical literature, it is important to consider the optimal preprocessing combinations as different preprocessing methods interact with model algorithms in complex ways and there is no single “best” approach

that will work for all datasets [27, 28], Table 3.

4.1. Baseline and noise removal

One of the main purposes of preprocessing is to remove baseline drift and noise without altering the chemical signal. The use of baseline correction and smoothing operations (derivatives, wavelet transforms, scatter correction) allows the models to concentrate on meaningful variation as opposed to instrument artifacts or random noise [28,29]. A special attention to spectral pre-processing methods, such as baseline removing, normalization, filtering, and derivative calculation (all these techniques are essential for reliable spectral analysis in both MVA/MVD) is given in references [30].

4.2. Normalization and Scaling

Normalization- scaling of the data to a common range or distribution. These techniques serve to prevent that measurements on entirely different scales (e.g. absorbance intensities at different wavelengths) have excessive influence on the training of a model. Several preprocessing methods, such as mean centering, unit variance scaling or range scaling, are well known in literature of chemometrics as typical tools used for the standardisation of dataset to make them more comparable and to stabilize multivariate models [25,30].

4.3. Dimension Reduction and Feature Selection

High-dimensional data is generally filled with a large amount of unimportant or redundant features. Techniques such as PCA for dimensionality reduction still capture a lot of data variations. Other types of feature selection algorithms attempt to clean up the data by pulling out the variables most relevant to the analysis. Such approaches have a profound value in model reduction and simplification and represent a more efficient way to ease computational spending while improving the predictive ability [31, 32]. Chemometrics surveys emphasize the importance of good feature selection for better interpretation of models and model generalization in dealing with some complex spectro- or chromatographic data [32].

4.4. Combining Preprocessing and Machine Learning

While hierarchical representations can be learned from raw data by modern machine learning models (in particular deep learning), the use of traditional preprocessing in advance of feature extraction is often beneficial to produce both more robust and interpretable results. For example, preprocessing methods such as smoothing and normalization had been found to improve clustering, classification or regression application results using AI model over the spectral data [33]. The investigation of the hybrid usage of window stable ary preprocessing, feature selection and machine learning classification shows better performance for Gaussian Raman spectrum analysis result showing the benefit for smart preprocessing to improve modeling efficiency [34].

Table 3. Common Data Preprocessing Methods and Their Applications

Preprocessing Method	Purpose	Typical Application	Reference
Baseline correction	Remove baseline drift	Spectroscopy data	[28]
Noise filtering & smoothing	Reduce random noise	Spectral & chromatographic data	[28,29]
Normalization	Standardize intensity scales	Multivariate models	[25,30]
Scaling	Adjust variable distributions	Chemometric & ML models	[25]
PCA	Dimensionality reduction	High-dimensional data	[31]
Feature selection	Identify informative variables	ML & classification	[32]
Window-based preprocessing	Improve ML performance	Raman & spectral analysis	[34]

5. Conclusions

The chemical analysis will benefit from the integration of classical chemometric techniques with AI methods in terms of efficiency, accuracy and automation. Classical models are interpretable and make a nice starting point to understand data, but AI/machine learning has immense predictive power, is good with nonlinearities, as well as being capable of handling big datasets in high dimensions. Good preprocessing on the data and feature engineering are crucial to get a good model performance. Finally, the fusion of AI with classical chemometrics, improves analytical precision and response rates, second develops modern chemical analysis tool in research and applied chemistry.

References

1. Saha P, Pandit B, Pramanik S, Shrestha B. A comprehensive review on the applications of chemometrics in analytical chemistry. J Appl Pharm Res [Internet]. 2024 [cited 2026 Feb 22];13(3). Available from: https://doi.org/10.69857/joapr.v13i3.861
2. Kowalski BR, Corcoran O, Wold S. Use of Chemometrics in Analytical Chemistry. Talanta. 2024;274:125949. Available from: https://www.sciencedirect.com/science/article/pii/S003991402400328X
3. Musa H, Ajibola VO, Omoniyi KI. Integrating Artificial Intelligence in Analytical Chemistry: Enhancing Spectroscopic Analysis and Predictive Modelling. Adv J Sci Eng Technol. 2025; 5: 1-9. Available from: https://aspjournals.net/ajset/index.php/ajset/article/view/150
4. Palit R, Dutta S, Basak S. Multi-Way Data Analysis Nowadays: Taking Advanced Chemometric Tools to Everyday Analytical Chemistry Applications. F Sensors. 2026;14(2):37. Available from: https://www.mdpi.com/2227-9040/14/2/37
5. IBM. Principal Component Analysis (PCA). IBM Think [Internet]. [cited 2026 Feb 22]. Available from: https://www.ibm.com/ae-ar/think/topics/principal-component-analysis.
6. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065):20150202.
7. Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscip Rev Comput Stat. 2010;2(4):433-59.
8. Forina M, Casolino MC, Lanteri S. Multivariate data analysis as a tool in analytical chemistry. Anal Chim Acta. 2023;1201:338046.
9. Jiën P, Yang Y, Gao Y. Support vector machines in chemometrics and analytical chemistry. TrAC Trends Anal Chem. 2020;123:115766.
10. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20(3):273-97.
11. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32.
12. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85-117.
13. Ballabio D, Consonni V. Classification tools in chemometrics. Part 1: linear models. Anal Methods. 2013;5(16):3790-8.
14. Ballabio D, Consonni V. Classification tools in chemometrics. Part 2: nonlinear models. Anal Methods. 2013;5(34):6244-51.
15. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.
16. Li H, Wang Y, Xu J, Liu Y. Artificial neural network modeling for prediction of chromatographic retention times. J Chromatogr A. 2021;1646:462066.
17. Rinnan Å, van den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem. 2009;28(10):1201-22.

18. Cardoso Rial R. AI in analytical chemistry: Advancements, challenges, and future directions. Talanta. 2024;274:125949. doi:10.1016/j.talanta.2024.125949
19. Pan L, Zhang P, Daengngam C, Chongcheawchamnan M. A review of artificial intelligence methods combined with Raman spectroscopy to identify the composition of substances. arXiv [Preprint]. 2021 Apr 5.
20. M Mousavizadegan, A Firoozbakhtian, M Hosseini, H Ju. Machine learning in analytical chemistry: From synthesis of nanostructures to their applications in luminescence sensing. TrAC Trends Anal Chem. 2023. doi:10.1016/j.trac.2023.117216
21. Kalpana D., Shishir K. Automation in Analytical Chemistry: The Role of AI in Chromatography. Innovative Academic Sciences. 2024; 16(3): 1-9.
22. Araújo CE et al. Aprendizado de máquina aplicado na eletroquímica. 2024. International Journal of Applied Pharmaceutics.
23. Applications of artificial intelligence for chemical analysis and monitoring of pharmaceutical and personal care products in water and wastewater: A review. Chemosphere. 2024;368:143692. doi:10.1016/j.chemosphere.2024.143692
24. Mukesh K, Nandi A, Yadav RL, Das Gupta G, Sharma K. AI-Enhanced Prediction Tools and Sensor Integration in Advanced Analytical Chemistry Techniques. Curr Anal Chem. 2025;21. doi:10.2174/0115734110373957250516113853
25. Rinnan Å, van den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem. 2009;28(10):1201–1222. doi:10.1016/j.trac.2009.07.007
26. Kumar N, Bansal A, Sarma GS, Rawal RK. Chemometrics tools used in analytical chemistry: an overview. Talanta. 2014;123:186–199. doi:10.1016/j.talanta.2014.02.003
27. Rato TJ, Reis MS. An optimization approach coupling preprocessing with model regression for enhanced chemometrics. Ind Eng Chem Res. 2019;58(47):21736–21747. doi:10.1021/acs.iecr.9b04583
28. Yan C. A review on spectral data preprocessing techniques for machine learning and quantitative analysis. iScience. 2025; DOI:10.1016/j.isci.2025.112759.
29. Bogomolov A, et al. Multisensor systems review including preprocessing of spectra. J Anal Chem. 2022;77(3):277–294.
30. Chemometrics and Data Analysis – Data preprocessing in analytical data analysis (overview). (General entry on feature scaling and normalization).
31. Chemometrics | Analytical Chemistry – Explorations of feature selection and its importance in multivariate models. ACS Anal Chem. (related feature selection importance).
32. Karisa M P., Benjamin K., Luke C M., Jamin C H., Robert E S. Review of chemometric analysis techniques for comprehensive two dimensional separations data – Discusses preprocessing in chromatography data analysis. J Chromatogr A. 2012;1255:3–11. doi:10.1016/j.chroma.2012.05.050
33. Mehdi E., Hamed T., Eiko T., Rainer K., José C., Eberhard H. Effective spectral preprocessing methods enhance accuracy – Example of smoothing and preprocessing to improve model predictions. Appl Sci. 2025; 7: 1-8. DOI:10.1007/s42452-025-07580-3
34. Yaju Z, Wei L., Yinsheng Z., Minmin T. Haiyan W. Enhanced data preprocessing with novel window function... – Combines preprocessing and feature selection for spectral classification. Spectrochim Acta A Mol Biomol Spectrosc. 2024; 323:124913. doi: 10.1016/j.saa.2024.124913. Epub 2024 Jul 31.