Artificial Intelligence and Machine Learning in Pharmaceutical Analytical Chemistry: Transforming Drug Quality Control and Process Optimization. A Review

Fayhaa Hassan Ali Ahmad [1*], Thura Z. Al-Sarraj [2]
[1] Department of Community Health Techniques, Mosul Medical Technical Institute, Northern Technical University, Mosul, Iraq
[2] Department of General Sciences, College of Basic Education, University of Mosul, Iraq

* Corresponding Author: **Fayhaa Hassan Ali Ahmad**

Article Info

E-ISSN: 2582-8940
ISSN (online): 2582-8940
Volume: 07
Issue: 02
Received: 07-03-2026
Accepted: 06-04-2026
Published: 05-05-2026
Page No: 113-117

Abstract
AI and ML in pharmaceutical analytical chemistry are transforming drug quality control and optimizing different aspects such as method, development, and manufacturing processes. This review aims to explore how AI and ML are changing four major UV-Vis spectrophotometry, High Performance Liquid Chromatography (HPLC), NIR spectroscopy, and Raman spectroscopy in pharmaceutical applications (2019-2025). It provides analytical solutions for quality control of the pharmaceutical industry and assesses the main ML algorithms: supervised models (ANN, SVM, RF, CNN), unsupervised methods (PCA), and hybrid chemometric-AI solutions. AI's role in conjunction with Process Analytical Technology (PAT), real-time release testing (RTRT), continuous manufacturing, and evolving drug solubility estimation, dissolution, and impurity detection is also evaluated. Some regulatory and interpretability challenges (e.g. compliance with ICH Q2(R2), Explainable AI (XAI), and validation of models) are discussed with possible solutions. The review proves that AI and ML in analytical methods for predicting quality control issues in the pharmaceutical industry outperform classical chemometric and AI solutions in terms of predictability and speed. However, data quality, model repeatability, and regulatory concerns are the biggest challenges in implementing these methods in the pharmaceutical industry.

DOI: https://doi.org/10.54660/IJMBHR.2026.7.2.113-117

Keywords: Artificial intelligence, Pharmaceutical quality control, Chemometrics, Process Analytical Technology, Spectroscopic methods

1. Introduction

The contemporary pharmaceutical sector must conform to increasing analytical demands on accuracy, selectivity, speed, sustainability, and cost-effectiveness. While traditional analytical methodologies are well-validated, they are limited in their ability to address modern analytical challenges. Traditional methods are reactive end product tests. The modern analytical market generates high dimensional, multivariate data sets. The emergence of AI and ML has transformed the landscape of pharmaceutical analysis, moving from static, deterministic methodologies to real-time, self-optimizing systems for pharmaceutical quality assurance. [1, 2, 3]

The basis of chemometrics in pharmaceutical spectroscopy has been multivariate statistical methods such as PLS, PCR, and MCR-ALS. Traditional chemometric models have linearity constraints, are poor at extracting features from data, and poorly model complicated analysis. Traditional chemometrics have poor performance in complex or non-linear analyses. AI and ML enhance the analysis capability of spectroscopy and chromatography through automated feature extraction and advanced non-linear calibration. [4, 5, 2]

This review will focus on the following: (1) the Moore's law of AI/ML for investigating and developing pharmaceutical analytical chemistry; (2) applications of AI/ML for UV-Vis spectroscopy, HPLC, NIR, and Raman; (3) applications of AI/ML in relation to the principles of PAT and continuous manufacturing; (4) the role of regulation and the challenges of explainability; (5) the research gaps and the future outlook for research.

2. Theoretical Foundations of AI/ML in Analytical Chemistry

2.1. From Chemometrics to Artificial Intelligence

Chemometrics is the basis of AI-driven pharmaceutical analysis and several techniques. PCA-PLS regression and MCR-ALS demonstrate the principles of dimensional and multivariate reduction, as well as data modelling and interpretation. All of which is foundational to modern ML algorithms. AI is a significant step forward from chemometrics. AI retains traditional linear models, but chemometrics lacks nonlinear, flexible, and large capacity models, which can readily accommodate disorganized large data sets [4,5].

2.2. Core ML Algorithm Classes

Predictive relationships between input features (e.g. spectral absorbances, chromatographic retention times) and target variables (e.g. drug concentration, purity) can only be formed using supervised learning.[6]

- **Artificial Neural Networks (ANN):** Multi-layered structures with adaptive weights that are trained using backpropagation. Very useful in nonlinear spectral calibration and dissolution prediction. Showed better accuracy than PLS in handling non-linear spectral overlaps in multi-component pharmaceutical mixtures [4].
- **Support Vector Machines (SVM):** Kernel-based classifiers/regressors that map data to high-dimensional feature spaces, enabling complex boundary decisions. SVM with Gaussian kernels achieved MAPE of 6.2% in predicting dissolution profiles of nifedipine extended-release tablets[7]
- **Random Forests (RF):** Ensemble decision tree models with inherent resistance to overfitting and built-in feature importance metrics. Emerged as the predominant algorithmic family for drug release modelling since 2020[7]
- **Convolutional Neural Networks (CNN):** Deep learning architectures extracting spatial/spectral features through convolutional layers. Applied to hyperspectral imaging and visual quality inspection of tablets and capsules[8]
- **Gradient Boosting (XGBoost, LightGBM):** Ensemble boosting algorithms demonstrating superior performance on structured pharmaceutical datasets with limited training samples[9]

Unsupervised Learning methods including PCA and clustering algorithms enable pattern discovery and anomaly detection in pharmaceutical process data without labelled training sets.[6]

Explainable AI (XAI) frameworks — including SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) — address the "black-box" interpretability problem by providing auditable, feature-level explanations for AI-driven analytical decisions, a critical requirement for regulatory compliance.[10][6]

3. AI/ML in UV-Vis Spectrophotometric Analysis

UV-Vis spectrophotometry combined with chemometric models has long provided a green, cost-effective alternative to HPLC for pharmaceutical quality control. The integration of AI elevates this platform by addressing its primary limitation: reduced selectivity in multi-component mixtures with highly overlapping spectra.[4]

ANN-based calibration models have demonstrated superior predictive performance over classical PLS and PCR in resolving nonlinear spectral overlaps, particularly in ternary and quaternary pharmaceutical mixtures where collinearity challenges classical regression. Firefly Algorithm-optimized ANN (FA-ANN) models achieved RMSECV values significantly lower than those of PLS in multi-analyte pharmaceutical systems, establishing a clear performance hierarchy: MCR-ALS > FA-ANN > SVR > PLS > PCR for complex spectral resolution tasks.[5, 4]

Support Vector Regression (SVR) has been specifically validated for simultaneously determining Paracetamol, Tramadol, and Aceclofenac using UV-Vis spectrophotometry in aqueous media, achieving high MoGAPI greenness scores (81) while eliminating the need for hazardous organic solvents. Furthermore, ML-enhanced UV-Vis platforms integrated with Process Analytical Technology (PAT) architectures enable real-time in-process monitoring of active pharmaceutical ingredient (API) concentrations during tablet manufacturing, reducing end-product testing burden.[3, 4]

4. AI/ML in HPLC Method Development and Optimization

HPLC is the benchmark method for assessing the purity of pharmaceuticals, and the considerable time and costs associated with developing new methdods using traditional techniques is mainly due to the need to carry out numerous experiments to assess different mobile phases, different gradients, different columns, and different flow rates. With advances in AI/ML, this process has been accelerated in the following ways [11]:

- **Predictive chromatographic modeling:** From molecular descriptors and historical data, ML algorithms predict optimal parameters for HPLC such as retention time, resolution, and peak symmetry, and this reduces the number of analytical experiments by up to 60% when compared to traditional optimizing HPLC methods [11].
- **Automated method development:** An HPLC system with AI capabilities integrates and adjusts system parameters based on the methods of the previous experiments, and this allows for high-throughput method development for new APIs [11].
- **Deep learning for peak detection:** Based on the concept of convolutional neural networks (CNNs) applied to chromatographic data, deep learning models address the gap in developing stability-indicating

methods by enabling the detection of impurity peaks beyond traditional threshold levels of UV detection [12].

- **Automated ML (AutoML) platforms:** An example of such a platform is the PAH test case, where models built and optimized using this platform outperformed models manually built and optimized in about half the typical time [7].

Among the first tools that have implemented AI for method development in pharmaceuticals is a new deep learning tool for the prediction of HPLC and GC chromatographic parameters, Chrompredict 1.0, which is currently the best available on the market [12].

5. AI/ML in NIR and Raman Spectroscopy for PAT

Near-infrared (NIR) and Raman spectroscopy, deployed as in-line PAT tools, generate continuous high-dimensional spectral datasets ideally suited for ML-based modelling.[13]

5.1. NIR Spectroscopy

ML-backed chemometric models powered by NIR spectroscopy revolutionize the real-time identification of relevant CQAs: concentration of the active pharmaceutical ingredient, polymer miscibility, moisture concentration, and mean particle size, throughout the entirety of the manufacturing process. During hot-melt extrusion, process variable estimation was accomplished using NIR models, which also identified preliminary shift adjustments of the formulation, thus allowing control of the manufacturing process. The implementation of machine learning for NIR data, used especially in controlled feature selection, has proven to be an effective solution to the challenge of establishing reliable and uniformly applicable calibration models for various manufacturing locations.[13, 14]

5.2. Raman Spectroscopy

AI-based Raman spectroscopy is a revolutionary tool for pharmaceutical quality control encompassing the characterization of drug structures, identifying polymorphs and impurities, and the measurement of biopharmaceutical processes within the industry. LSTMs have been successfully applied to predicted important specifications of bioprocesses (cell density, metabolite concentrations, and product concentration) within a bioprocessing and predicted in real-time offering the potential for identifying adjustments to the bioprocess in an early manner. Raman spectroscopy is better than Near Infrared (NIR) spectroscopy for determining active pharmaceutical ingredient concentrations as low as 1% (w/w) in tablet compression feed frames, and is the preferred process analytical technology tool for low-dose pharmaceutical formulations [13, 15].

6. AI/ML for Drug Property Prediction and Process Optimization

Beyond analytical measurement, ML models have demonstrated remarkable predictive power for physicochemical property estimation and formulation optimization:

- **Drug solubility prediction:** ANN, SVM, and RF models integrating molecular descriptors with Hansen Solubility Parameters achieved high accuracy (K-fold cross-validation score = 0.933) for drug solubility prediction in complex mixed-solvent micro-environments[16]
- **Dissolution profile prediction:** RF and XGBoost ensemble models achieved robust dissolution predictions for controlled-release formulations, with built-in SHAP-based feature importance metrics identifying critical formulation variables[7]
- **Continuous manufacturing:** Reinforcement Learning (RL) agents dynamically optimize coating uniformity, yield, and process conditions in real-time by learning from continuous process feedback — transforming pharmaceutical manufacturing from static operations to self-optimizing Industry 4.0-compliant systems[10]
- **Real-Time Release Testing (RTRT):** AI-enhanced NIR and Raman models enable batch release based on continuous in-process spectral data, eliminating the need for traditional end-product dissolution and assay testing[3]

7. Comparison of AI/ML Algorithms in Pharmaceutical Analytics

Algorithm	Type	Best Application	Accuracy	Interpretability	Data Requirement
PLS/PCR	Classical chemometric	Linear spectral overlap	Moderate	High	Low
ANN	Supervised DL	Nonlinear calibration	High	Low	Moderate
SVM	Supervised ML	Classification, regression	High	Moderate	Low–Moderate
Random Forest	Ensemble ML	Dissolution, formulation	Very High	Moderate (SHAP)	Moderate
CNN	Deep learning	Spectral imaging, visual QC	Very High	Low (XAI needed)	High
LSTM	Recurrent DL	Time-series PAT data	Very High	Low	High
XGBoost	Ensemble boosting	Structured tabular data	Very High	Moderate (SHAP)	Low–Moderate
MCR-ALS	Chemometric	Multi-component resolution	High	High	Low

8. Regulatory Considerations and Explainability

The deployment of AI/ML models in pharmaceutical quality control is subject to rigorous regulatory scrutiny. Key regulatory frameworks relevant to AI/ML analytical applications include:[1]

- **ICH Q2(R2):** Method validation guidelines requiring demonstrated accuracy, precision, linearity, and robustness — criteria that AI models must satisfy through statistically rigorous external validation, including independent test sets and stability-indicating protocols[1]
- **ICH Q8–Q10 (QbD framework):** AI-driven design of experiments (DoE) and response surface methodology (RSM) are increasingly recognized within QbD as tools for identifying design spaces and control strategies[1]
- **FDA's Process Analytical Technology (PAT) Guidance:** Explicitly encourages the use of Real-time

monitoring and management of processes with advanced analytic tools — AI/ML spectroscopic models [3]

The main barrier to the widespread regulatory acceptance of deep learning models is their "black box" nature. The SHAP and LIME applications are considered to be important aids for achieving regulatory acceptance of deep learning models as they provide interpretable and auditable explanations for AI-driven decisions [6, 10].

9. Challenges and Future Directions

Despite transformative potential, several challenges impede widespread AI/ML adoption in pharmaceutical analytical chemistry:[6]

1. Come up with some samples (or sites) for a training dataset, fix your AI, add data filtration for operators, and rely on AI learning while automating your system [6].
2. Scalability of AI-data transfers means that any on-site AI engineering will result in the AI being unusable at other sites. Changes in AI at other sites will need on-site calibration [14].
3. Validation of AI intelligence translates to a fast track for the AI's integration into the manufacturing and trading of psychotropic substances. [1]
4. A trade of the highest possible capacity vs. interpretable, low capacity models for drug development is a serious dilemma when it comes to extended controlled commercialization of psychiatric medicine. [5]

Future priorities include

- Creation of hybrid chemometry-AI models which combine the interpretability of PLS/PCR and the predictive ability of deep learning [4].
- Combination of digital twin technology and AI-driven PAT for fully automated and self-correcting pharmaceutical manufacturing production lines [6].
- Development of collaborative ICH-FDA consensus guidelines for the AI model validation in pharmaceutical analytical method development [1].
- Development of new APIs based on the expansion of AutoML and transfer learning to minimize data requirements and speed up methods development [7].

10. Conclusion

This review details how AI and ML have become instrumental in pharmaceutical analytical chemistry. Taking into account the existing and emerging AI/ML applications in pharmaceutical analytical chemistry, we will assess how they enable a paradigm shift in the measurement, monitoring, and assurance of drug quality. Chemometric methods (PLS, PCR, and MCR-ALS) and their effective, interpretable, and regulatory-compliant analytical solutions provide a starting point for the AI/ML methods and architectures (ANN, SVM, RF, CNN, LSTM) in the realms of nonlinear, high dimensional, and real-time analytics. The marriage of AI to the PAT, RTRT, and continuous manufacturing frameworks places pharmaceutical analytical chemistry at the edge of completely automated and fully anticipatory quality assurance.

The gaps remaining, such as interpretability, data standardization, and harmonization in regulations and quality standards, can only be bridged through teamwork across various disciplines with the common end of delivering safer and higher-quality pharmaceutical products to the patients around the globe.

References

1. European Medicines Agency. Review of artificial intelligence and machine learning applications in the medicine's lifecycle [Internet]. European Medicines Agency; [cited 2026 May 1[3]. Available from: European Medicines Agency
2. Jiang Z, Shi H, Yang B, Yin X. Artificial intelligence-enabled analytical technologies for pharmaceutical quality control and process optimization. Chem Eng Technol. 2026;49(4):e202400312. doi:10.1002/ceat.70213
3. Patel R, Mehta K, Shah D. Role of artificial intelligence and machine learning in pharmaceutical formulation: recent advances and future perspectives. Biotech Asia. 2026;23(1):45-61.
4. Workman J Jr, Lavine B. Recent research in chemometrics and AI for spectroscopy. Part I: Foundations, definitions, and algorithms. Spectroscopy. 2025;40(11):12-21.
5. Kumar S, Singh A, Sharma P. A comprehensive review of advancing pharmaceutical analysis technology through chemometric approaches. Biotech Asia. 2025;22(4):88-102.
6. Hassan WA, El-Kady EH, Fouad MA. Artificial intelligence in bioanalytical chemistry: a review of algorithms, applications, and future directions. J Anal Chem. 2024;79(12):1423-1438. doi:10.1134/S1061934824120089
7. Agarwal M, Tiwari R, Verma N. Applications and tools of artificial intelligence in analytical chemistry. Int J Sci Res Anal. 2025;4(3):215-228.
8. Xu Z, Chen L, Wang Y, Zhang H, Li M, Liu J, *et al*. Artificial intelligence-guided Raman spectroscopy in biomedicine and pharmaceutical quality control. Front Pharmacol. 2025;16:1521847. doi:10.3389/fphar.2025.1521847
9. Feudale RN, Woody NA, Tan H, Myles AJ, Brown SD, Ferré J. Transfer of multivariate calibration models: a review. Chemom Intell Lab Syst. 2002;64(2):181-192. doi:10.1016/S0169-7439(02)00048-8
10. Pasquini C. In-line Raman and NIR spectroscopy for real-time pharmaceutical process monitoring [Internet]. AZO Optics; [cited 2026 May 1[3]. Available from: AZO Optics
11. Rashid M, Khan MA, Alam F. AI-driven spectroscopic data interpretation in pharmaceutical quality control: a systematic overview. J Emerg Technol Innov Res. 2025;12(7):58-71.
12. Ding Y, Li Z, Wang X, Chen G. Development of deep learning software (Chrompredict 1.0) to improve HPLC and GC chromatographic parameter prediction. J Chromatogr A. 2024;1712:464523. doi:10.1016/j.chroma.2024.464523

13. Tripathi A, Gupta S, Singh RK. Predictive modeling of drug release kinetics using supervised machine learning algorithms. Int J Drug Dev Technol. 2024;16(3 Suppl):118. doi:10.25258/ijddt.16.3S.118
14. Chen W, Liu Y, Zhang F, Wang Q. Prediction of drug solubility in complex micro-environments using machine learning integrated with Hansen solubility parameters. Sustain Chem Eng. 2025;3(4):1000005. doi:10.1016/j.sce.2025.1000005
15. Mouchlis VD, Afantitis A, Serra A, Fratello M, Papadiamantis AG, Aidinis V, *et al*. Advances in de novo drug design: from conventional to machine learning methods. Int J Mol Sci. 2021;22(4):1676. doi:10.3390/ijms22041676

How to Cite This Article
Ahmad FHA, Al-Sarraj TZ. Artificial intelligence and machine learning in pharmaceutical analytical chemistry: transforming drug quality control and process optimization. A review. Int J Med All Body Health Res. 2026;7(2):113-117. doi:10.54660/IJMBHR.2026.7.2.113-117.