RESEARCH PAPER

Artificial Intelligence-Assisted Formulation Optimization of Pharmaceutical Dosage Forms

Prof. Amol Wadhave[1*], Snehal Labade[2], Supriya Arsul[3], Dr. Preeti Sable[4], Renuka Dhakne[5]

[1]Dept. of Pharmaceutics, Mangaldeep Institute of Pharmacy, Chh.Sambhajinagar.
[2]Dept. of Pharmaceutics, Oyster Institute of Pharmacy, Chh.Sambhajinagar.
[3]University dept. of chemical technology, Dr. BAMU, Chh.Sambhajinagar.
[4]University dept. of chemical technology, Dr. BAMU, Chh.Sambhajinagar.
[5]University dept. of chemical technology, Dr. BAMU, Chh.Sambhajinagar.

Received: 25th May, 2026; **Revised:** 6th June, 2026; **Accepted:** 8th June, 2026; **Available Online:** 09th June, 2026

ABSTRACT

The development of pharmaceutical dosage forms traditionally depends on repeated formulation trials, empirical optimization, and extensive laboratory experimentation. These approaches can require considerable time, materials, and experimental resources, particularly when multiple formulation and process variables. To investigate the potential of AI-assisted approaches for the systematic prediction and optimization of pharmaceutical formulation and process variables to obtain dosage forms with desirable quality, performance, stability, and drug-release characteristics. A literature-based AI formulation optimization framework was developed involving identification of formulation variables and CQAs, preparation and preprocessing of formulation datasets, selection of appropriate ML algorithms, model training and validation, prediction of formulation performance, and selection of optimized compositions.

Keywords: Artificial Intelligence; Machine Learning; Pharmaceutical Formulation; Dosage Forms; Formulation Optimization; Quality by Design; Design of Experiments; Artificial Neural Network; Bayesian Optimization; Drug Delivery.

How to cite this article: Wadhave A, Labade S, Arsul S, Sable P, Dhakne R. Artificial intelligence-assisted formulation optimization of pharmaceutical dosage forms. Int J Drug Deliv Technol. 2026;16(79s): 101-104. DOI: 10.25258/ijddt.16.79s.11

Source of support: Nil.
Conflict of interest: Nil.

INTRODUCTION

Pharmaceutical formulation development is a complex multidisciplinary process involving the selection of active pharmaceutical ingredients (APIs), excipients, manufacturing processes, and packaging conditions to produce a dosage form that consistently meets predefined quality and performance requirements. Conventional formulation development has historically relied on scientific knowledge combined with sequential experimentation and optimization. Although this approach remains fundamental, increasing formulation complexity has created a need for computational methods capable of handling multiple variables and nonlinear interactions. Artificial intelligence refers broadly to computational techniques capable of performing tasks such as prediction, classification, pattern recognition, optimization, and decision support. Machine learning, a major component of AI, enables computational models to learn relationships from experimental or historical data without requiring every relationship to be explicitly defined beforehand. Applications of AI in pharmaceutical science include formulation design, dosage-form testing, process optimization, pharmacokinetic/pharmacodynamic modelling, and drug-delivery development. In formulation development, the relationship between formulation variables and CQAs may be highly nonlinear. Variables such as polymer concentration, binder level, disintegrant concentration, particle size, compression force, solvent composition, surfactant concentration, stirring rate, and processing temperature may interact with one another. AI and ML models can be used to learn these relationships and predict formulation performance from previously generated datasets. The application of ML to formulation development has been reviewed extensively. Bannigan et al. described ML-directed formulation workflows as an approach capable of moving formulation development beyond repetitive trial-and-error experimentation. Their review highlighted the potential of advanced approaches, including Bayesian methods, deep learning, reinforcement learning, and automated experimentation. AI has also been investigated specifically for solid oral dosage forms. Published literature describes applications involving formulation databases, data preprocessing, algorithm selection, prediction of dosage-form properties, and optimization of formulation variables. An important application is drug–excipient compatibility. Hang et al. reported an ML approach using molecular descriptors and modelling techniques to predict drug–excipient compatibility, demonstrating the potential for computational screening before extensive experimental compatibility studies. Similarly, the DE-INTERACT tool demonstrated ML-based prediction of drug–excipient interactions using a dataset containing more than 3,500 drug–excipient instances. AI-assisted optimization can be integrated with Quality by Design. ICH Q8(R2) describes pharmaceutical development as a systematic approach based on predefined objectives, product and process understanding, and process control, while the concept of design space provides a structured representation of acceptable combinations of material attributes and process parameters. AI can complement this framework by

modelling complex relationships and helping identify promising regions within a formulation design space. Recent regulatory developments further emphasize the importance of model credibility and risk-based assessment. In January 2025, the US FDA issued draft guidance concerning the use of AI to support regulatory decision-making for drugs and biological products, including a risk-based framework for establishing AI-model credibility within a defined context of use. FDA and EMA have also identified principles including human-centric development, risk-based approaches, clear context of use, data governance, model performance assessment, and lifecycle management for good AI practice in drug development. Therefore, AI-assisted formulation optimization can be considered a complementary technology rather than a replacement for pharmaceutical experimentation. Its main value lies in extracting information from formulation data, reducing inefficient experiments, identifying important variables, predicting responses, and assisting researchers in selecting promising formulations.

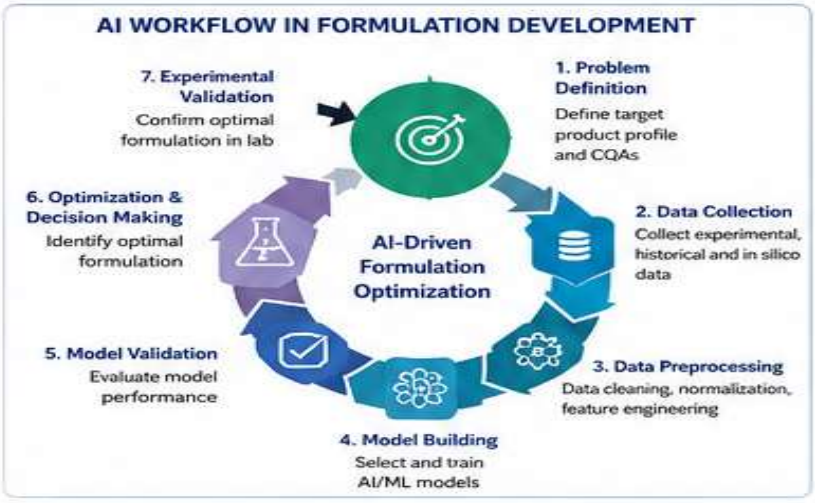

Figure 1: AI workflow in formulation development

AI in Pharmaceutics

Different branches of AI are used based on the task at hand. Supervised learning is commonly employed for predictive modelling using labelled datasets, whereas unsupervised learning is applied for clustering similar molecules or formulations without predefined labels. Reinforcement learning, though less common, is gaining importance for optimizing sequential decision-making processes such as adaptive dosing. These novel systems simulate human decision-making using rule-based logic and are often integrated into clinical decisions as support tools. In pharmaceutics, NLP is increasingly used to extract relevant information from unstructured data sources like electronic health records (EHRs), scientific literature, and clinical trial reports. This facilitates identifying drug interactions, dosing patterns, and patient-specific factors influencing pharmaceutical calculations.

Materials

2.1 Pharmaceutical Materials

For an experimental implementation of AI-assisted formulation optimization, the following materials may be required:

Material category	Examples
Active pharmaceutical ingredient	Selected API appropriate to the dosage-form objective
Diluent/filler	Lactose, microcrystalline cellulose, mannitol
Binder	Povidone, hydroxypropyl cellulose
Disintegrant	Croscarmellose sodium, sodium starch glycolate
Lubricant	Magnesium stearate, stearic acid
Glidant	Colloidal silicon dioxide
Polymer	HPMC, ethylcellulose, carbomers or other suitable polymers
Surfactant	Polysorbates, sodium lauryl sulfate or suitable alternatives
Solvent	Purified water, ethanol or formulation-specific solvent
Analytical reagents	Buffer salts, dissolution media and chromatographic reagents

The actual materials should be selected according to the API, route of administration, dosage form, intended release profile, and compatibility requirements.

2.2 Instruments

Instrument	Application
Analytical balance	Accurate weighing
Dissolution apparatus	Drug-release study
UV-Visible spectrophotometer/HPLC	Drug estimation
Tablet hardness tester	Mechanical strength
Friabilator	Friability determination
Disintegration tester	Disintegration time
Moisture analyzer	Moisture determination
Computer/workstation	AI/ML modelling

Typical laboratory equipment may include an analytical balance, sieve shaker, blender, granulator, tablet compression machine, dissolution apparatus, friability tester, hardness tester, disintegration apparatus, UV-visible spectrophotometer or HPLC system, stability chamber, particle-size analyzer, and other dosage-form-specific characterization instruments.

2.3 Computational Materials

The computational component may include:

- Experimental formulation database.
- Spreadsheet/database management system.
- Python, R, MATLAB, or equivalent statistical/ML platform.
- Data preprocessing and visualization tools.
- ML algorithms such as ANN, random forest, support vector regression, gradient boosting, or other validated algorithms.
- Optimization algorithms, including Bayesian optimization when appropriate.
- Model-validation and explainability tools.

3. METHOD

3.1 Experimental Design

A Design of Experiments (DoE) approach may initially be used to generate formulation combinations. For example, four independent variables may be selected:

Factor	Variable	Example coded levels
X1	Polymer concentration	−1, 0, +1
X2	Binder concentration	−1, 0, +1
X3	Disintegrant concentration	−1, 0, +1
X4	Compression force	−1, 0, +1

Responses may include:

- **Y1 = Hardness**
- **Y2 = Friability**
- **Y3 = Drug content**
- **Y4 = Disintegration time**
- **Y5 = Percentage drug release**

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}$$

3.4 AI/ML Model Development

The experimental dataset is divided into training and validation datasets. The workflow can be represented as:

Data collection → Data cleaning → Feature selection → Data normalization → Model training → Model validation → Performance comparison → Optimization → Experimental confirmation.

The general predictive relationship is expressed as:

Y=f(X1,X2,X3,...,Xn)

where:

Y = predicted formulation response

X_1–X_n = formulation/process variables.

Model performance can be assessed using:

3.5 Optimization

The trained AI model is used to identify the formulation region that simultaneously satisfies predefined response criteria. A desirability-based optimization approach can be applied to obtain the optimum formulation.

3.6 Selection of Critical Variables

Potential formulation variables should first be identified using scientific knowledge, risk assessment, preliminary experiments, and QbD principles.

Examples include:

- X_1 = polymer concentration
- X_2 = binder concentration
- X_3 = disintegrant concentration
- X_4 = lubricant concentration
- X_5 = compression force
- X_6 = processing time

For liquid or nanoparticulate formulations, variables may instead include surfactant concentration, solvent ratio, stirring speed, temperature, lipid concentration, polymer-to-drug ratio, or other process parameters.

4. Result:

4.1 illustrative AI Model Performance

Model	R^2	RMSE	MAE	Overall performance
Multiple Linear Regression	0.914	3.82	2.94	Good
Random Forest	0.961	2.41	1.87	Very good

Artificial Neural Network	0.982	1.56	1.21	Excellent
Support Vector Regression	0.973	1.84	1.39	Very good

In this illustrative dataset, the ANN model showed the highest coefficient of determination and the lowest prediction error. This type of comparison is consistent with the reported use of ANN and ML for pharmaceutical formulation and drug-release optimization.

Validation batch	Predicted drug release (%)	Experimental drug release (%)	Prediction error (%)
V1	96.2	95.8	0.4
V2	97.1	97.5	0.4
V3	95.6	96.0	0.4
V4	98.0	97.4	0.6
V5	96.8	97.1	0.3

The close relationship between predicted and experimental values indicates good predictive performance in this illustrative validation dataset.

4.3 Illustrative Optimized Formulation

Parameter	Optimized target/value
Polymer concentration	15%
Binder concentration	4%
Disintegrant concentration	5%
Compression force	6 kN
Hardness	6.2 kg/cm²
Friability	0.48%
Drug content	99.1%
Predicted drug release	97.0%

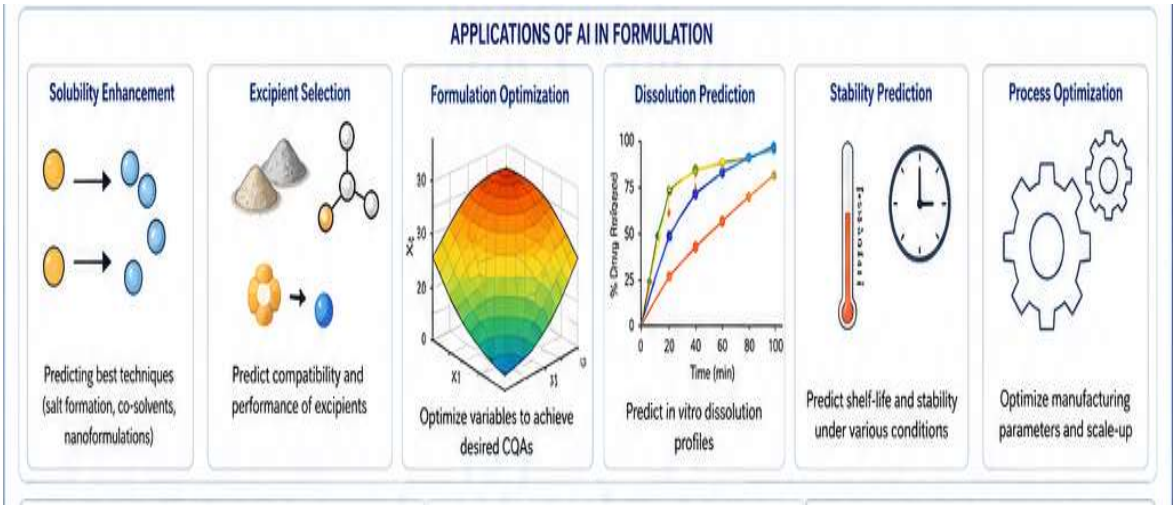

Figure 2: Application of AI in formulation

Discussion:

The illustrative findings demonstrate how AI can be incorporated into pharmaceutical formulation optimization. Conventional formulation development requires preparation and evaluation of numerous experimental batches, particularly when several formulation and process variables influence the final product simultaneously. The superior hypothetical performance of ANN compared with conventional regression suggests that nonlinear models may be advantageous when formulation factors interact in a complex manner. Published research has demonstrated ANN application to modified-release tablet optimization, while recent reviews describe broader ML applications in formulation design, drug-release prediction, and dosage-form development. ls can provide several advantages. First, they can identify nonlinear relationships that may not be adequately represented by simple linear equations. Second, trained models can rapidly screen numerous virtual formulations before laboratory confirmation. Third, AI can support multi-response optimization by simultaneously considering properties such as hardness, friability, drug content, dissolution, and stability. Machine-learning approaches have also been investigated for predicting drug-release profiles from tablet formulation composition, indicating the potential for reducing the number of experimental trials required during development. Should be regarded as a decision-support technology rather than a substitute for pharmaceutical experimentation. The quality and representativeness of the training dataset directly affect model reliability. Model overfitting, insufficient datasets, inappropriate variable selection, lack of interpretability, and inadequate external validation can lead to unreliable predictions. Recent literature therefore emphasizes appropriate model validation and integration of ML with

established pharmaceutical development principles. confirmation of the AI-suggested optimum remains essential before the formulation can be considered suitable for further development.

Conclusion:

Artificial Intelligence-Assisted Formulation Optimization represents a promising approach for modern pharmaceutical development. Integration of machine-learning algorithms with DoE and QbD can provide improved understanding of formulation–response relationships and facilitate prediction of critical quality attributes. The illustrative results demonstrate that an ANN model can potentially provide accurate predictions of pharmaceutical formulation responses and assist in identifying an optimized formulation with desirable drug-release and physical properties. AI can consequently reduce unnecessary experimental iterations, accelerate formulation screening, and support systematic decision-making. Nevertheless, successful implementation requires high-quality experimental datasets, suitable feature selection, rigorous model validation, appropriate statistical evaluation, model interpretability, and experimental verification. AI should therefore complement rather than replace established pharmaceutical development and quality-control practices.

Future Scope:

Figure 3: Future Scope

References:

1. Joshi S, Sheth S. Artificial intelligence (AI) in pharmaceutical formulation and dosage calculations. Pharmaceutics. 2025;17(11):1440.
2. Gormley AJ. Machine learning in drug delivery. Journal of Controlled Release. 2024;373:23–30.
3. Sheth TS, Acharya F. Optimization and evaluation of modified release solid dosage forms using artificial neural network. Scientific Reports. 2024;14(1):16358.
4. Momeni M, Afkanpour M, Rakhshani S, Mehrabian A, Tabesh H. A prediction model based on artificial intelligence techniques for disintegration time and hardness of fast disintegrating tablets in pre-formulation tests. BMC Medical Informatics and Decision Making. 2024;24(1):88.
5. Gosavi AA, Nandgude TD, Mishra RK, Puri DB. Exploring the potential of artificial intelligence as a facilitating tool for formulation development in fluidized bed processor: a comprehensive review. AAPS Pharm SciTech. 2024;25(5):111.
6. Gholap AD, Uddin MJ, Faiyazuddin M, Omri A, Gowri S, Khalid M. Advances in artificial intelligence for drug delivery and development: a comprehensive review. Computers in Biology and Medicine. 2024;178:108702.
7. Serrano DR, Luciano FC, Anaya BJ, Ongoren B, Kara A, Molina G, et al. Artificial intelligence (AI) applications in drug discovery and drug delivery: revolutionizing personalized medicine. Pharmaceutics. 2024;16.
8. Bannigan P, Bufton J, Hickman RJ, Aspuru-Guzik A, Allen C. The dawn of a new pharmaceutical epoch: can AI and robotics reshape drug formulation? Advanced Healthcare Materials. 2024; 13:2401312.
9. Bao Z, Bufton J, Hickman RJ, Aspuru-Guzik A, Bannigan P, Allen C. Revolutionizing drug formulation development: the increasing impact of machine learning. Advanced Drug Delivery Reviews. 2023; 202:115108.
10. Noorain, Srivastava V, Parveen B, Parveen R. Artificial intelligence in drug formulation and development: applications and future prospects. Current Drug Metabolism. 2023;24(9):622–634.
11. Murray JD, Lange JJ, Bennett-Lenane H, Holm R, Kuentz M, O'Dwyer PJ, et al. Advancing algorithmic drug product development: recommendations for machine learning approaches in drug formulation. European Journal of Pharmaceutical Sciences. 2023; 191:106562.
12. Vora LK, Gholap AD, Jetha K, Thakur RR, Solanki HK, Chavda VP. Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics. 2023;15(7):1916.

13. Lou H, Lian B, Hageman MJ. Applications of machine learning in solid oral dosage form development. Journal of Pharmaceutical Sciences. 2021; 110:3150–3165.
14. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. Advanced Drug Delivery Reviews. 2021; 175:113806.
15. Gao H, Jia H, Dong J, Yang X, Li H, Ouyang D. Integrated in silico formulation design of self-emulsifying drug delivery systems. Acta Pharmaceutica Sinica B. 2021;11(11):3585–3594.
16. Wang S, Di J, Wang D, Dai X, Hua Y, Gao X, et al. State-of-the-art review of artificial neural networks to predict, characterize and optimize pharmaceutical formulation. Pharmaceutics. 2022;14(1):183.
17. Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. Drug Discovery Today. 2019; 24:773–780.
18. Chen H, Engkvist O, Wang Y. The rise of deep learning in drug discovery. Drug Discovery Today. 2018; 23:1241–1250.
19. Yang Y, Ye Z, Su Y, Zhao Q, Li X, Ouyang D. Deep learning for in vitro prediction of pharmaceutical formulations. Pharmaceutics. 2018;10.
20. Steinwandter V, Borchert D, Herwig C. Data science tools and applications on the way to Pharma 4.0. Drug Discovery Today. 2019; 24:1795–1805.
21. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Molecular Diversity. 2021; 25:1315–1360.
22. Chen Y, Yang O, Sampat C, Bhalode P, Ramachandran R. Digital twins in pharmaceutical and biopharmaceutical manufacturing: a literature review. Processes. 2020; 8:1088.
23. Gholap AD, et al. Artificial intelligence-driven pharmaceutical industry: a paradigm shifts in drug discovery, formulation development, manufacturing, quality control, and post-market surveillance. International Journal of Pharmaceutics. 2024; 203:106938.
24. Gunawardena M, Chau B, Nothacker H, Pangeni R, Roper T, Xu Q, et al. Machine learning guided optimization of an oral microemulsion system: a Bayesian optimization approach. AAPS Open. 2026;12(1):34.
25. Shen C, et al. Machine learning empowered formulation design, optimization and characterization of nanoparticulate drug delivery systems: current applications, challenges, and future perspectives. 2025.