INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES
[ISSN: 0975-4725; CODEN(USA): IJPS00]
Journal Homepage: https://www.ijpsjournal.com

OPEN ACCESS

Review Paper

Artifical Intelligence And Machine Learning in Solid Dosage Formulation

Pratik Patel*, Preeti Sah, Zahid Husain, Shivanshu Dwivedi, Rajveer Singh Chauhan, Nimita Manocha

Indore Institute of Pharmacy, Rau-Indore, Madhya Pradesh, India.

ARTICLE INFO

Published: 29 Sept 2026
Keywords:
Artificial Intelligence, Machine Learning, Solid Dosage, Process Analytical Technology, Self-Driving Labs, Federated Learning, Explainable AI
DOI:
10.5281/zenodo.23038296

ABSTRACT

Solid dosage forms, tablets, capsules, pellets, and powders, remain the most widely manufactured medicines worldwide, yet their development has long relied on slow, trial-and-error experimentation that contributes to the 10–15 year, multi-billion-dollar cost of bringing a new drug to market. This review examines how artificial intelligence (AI) and machine learning (ML) are reshaping solid dosage formulation across its full lifecycle. The review surveys documented applications spanning pre-formulation screening, formulation design, manufacturing process control, dissolution and stability modeling, analytical method development, and emerging generative AI approaches, supported by published results demonstrating high predictive and detection accuracy. Real-world case studies, including Pfizer's AI-assisted oral formulation work, a Merck convolutional neural network for tablet coating-defect detection, and autonomous formulation platforms such as Intrepid Labs, demonstrate that these methods have moved from research promise toward industrial deployment. The review then examines the data-related, model-related, regulatory, industry, and ethical challenges still limiting broader adoption, including proprietary data scarcity, black-box interpretability, and a regulatory landscape only beginning to formalize AI-specific validation requirements through the FDA's 2025 draft guidance and the FDA-EMA's January 2026 joint guiding principles. Future directions considered include federated learning for cross-industry data sharing, explainable AI for regulatory-ready models, digital twins for end-to-end manufacturing, and AI-guided 3D printing for personalized dosage forms. Rather than replacing the formulation scientist, AI/ML is reshaping that role, and this review concludes that closing the remaining gap depends less on further technical advances than on coordinated progress among academia, industry, and regulators.

***Corresponding Author:** Pratik Patel
Address: *Indore Institute of Pharmacy, Rau-Indore, Madhya Pradesh, India.*
Email ✉: pratik890patel@gmail.com
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

INTRODUCTION

1.1 Overview of Solid Dosage Forms (Tablets, Capsules, Pellets, Powders)

Solid dosage forms, such as tablets, capsules, pellets, and powders, are commonly used in pharmaceutical drug delivery because of their stability, convenience, accurate dosing, and ease of manufacturing. Tablets and capsules are the most common oral solid forms, while pellets offer multiarticulate and modified-release capabilities.

A. Tablet

They are made by compressing drug and other ingredients into a small solid unit. Some tablets release medicine quickly, while others release it slowly over time. A few are made to dissolve in the mouth without water, and some are coated to protect the drug or hide its taste. [1]

B. Capsule

Capsules are medicines packed inside a shell, usually made of gelatin or a plant-based material. Hard capsules usually contain powders or pellets, while soft capsules often contain liquids or oily medicines. [2]

C. Pellets

Pellets are tiny spherical particles of medicine. They are often filled into capsules or compressed into tablets. A major advantage is that pellets can be coated differently, so medicine can be released at different times. This makes them very useful for modified-release products. [3]

D. Powders

Powders are finely divided medicines that may be taken directly, mixed with water, or used externally on the skin. They are simple to prepare and can be useful for children or older patients when flexible dosing is needed. [4]

1.2 Traditional formulation development: limitations and time/cost burden

For decades, pharmaceutical scientists have navigated formulation development largely by intuition and iteration adjusting excipient ratios, tweaking mixing durations, and varying compression forces across batch after batch until an acceptable formulation was obtained. The cost of this approach goes well beyond the laboratory. Raw materials are consumed in failed experiments, skilled technicians spend weeks on repetitive testing, and critical problems such as low bioavailability or poor stability often surface only after months of effort. With industry benchmarks placing the average timeline at 10-15 years and costs exceeding one billion dollars, the inefficiencies embedded in traditional workflows represent a structural problem that the sector can no longer afford. [5],[6],[7]

1.3 The rise of AI/ML sciences

In formulation, ML models suggest the best excipient ratios and processing conditions based on previously collected experimental data, reducing the need for repeated trial batches. In manufacturing, AI-powered sensors monitor processes in real time and catch deviations early. Even supply chains now use predictive algorithms to prevent shortages and manage inventory smartly. [8],[9] The financial case is undeniable. AI-driven improvements are projected to unlock between $350 billion and $410 billion in annual value across the pharmaceutical and biotech sectors, largely by compressing timelines and reducing failure rates. Generative AI, in particular, is being described as a once-in-a-generation opportunity for the industry, with the potential to design entirely new drug candidates from

scratch.[10],[11] Rather than manually screening hundreds of formulations or running repetitive stability checks, researchers now guide and interpret AI-assisted workflows, bringing their scientific judgment to bear on problems the model cannot solve alone. [12]

1.4 Scope and objectives of the review

Solid dosage formulation has always demanded a careful balance getting the appropriate drug, and delivery profile, delivered in the right way. For years, achieving that balance meant running experiment after experiment until something worked. The review traces how AI and machine learning are being used at every stage, starting from the earliest pre-formulation decisions about solubility and excipient compatibility, moving through formulation design and manufacturing, and stretching all the way to stability prediction and quality control.

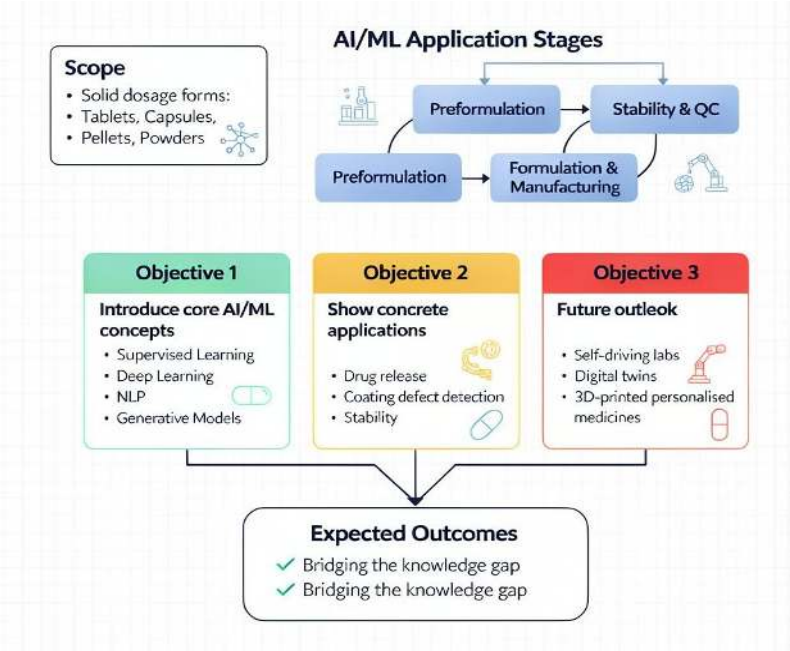

Figure 1 Framework for AI/ML Integration Solid Dosage Forms

2. FUNDAMENTALS OF AI AND ML RELEVANT TO PHARMACEUTICAL FORMULATION

AI and Machine Learning (ML) help in this industries by analysing large amounts of drug and formulation data.
They predict how different ingredients will behave and help select the right combination and quantities.
AI/ML can also predict properties such as drug stability, solubility, release rate, and effectiveness. This makes formulation development faster, more accurate, and less dependent on trial-and-error experiments.

2.1 Key AI/ML concepts

2.1.1 Supervised, unsupervised, and reinforcement learning

Table 2 Supervised Learning Full Overview

Point	Definition / Description
Definition	A learning paradigm where the model is trained on labeled input-output pairs, learning a mapping function to predict outputs for new, unseen inputs. [13]
Working Principle	The algorithm minimizes error between predicted and actual labeled outputs using a loss function (e.g., cross-entropy, MSE) through iterative optimization. [14]
Key Algorithms	Support Vector Machine (SVM), Random Forest, Decision Trees, Gradient Boosting (XGBoost), Artificial Neural Networks (ANN), Deep Neural Networks (DNN)
Task Types	Classification (e.g., active/inactive compound) and Regression (e.g., predicting drug release %)[13]
Pharma Application 1	Predicting drug-excipient compatibility from physicochemical descriptors to avoid formulation instability
Pharma Application 2	ADME/Tox prediction training DNNs on molecular fingerprints to predict absorption, distribution, metabolism, excretion, and toxicity. [15]

Table 3 Unsupervised Learning Full Overview

Point	Definition / Description
Definition	A learning paradigm that works on unlabeled data, allowing the model to discover hidden patterns, structure, or groupings without predefined output categories. [16]
Working Principle	The model minimizes an internal criterion (e.g., intra-cluster variance, reconstruction error) without any external labels it self-organizes data representations. [16]
Key Algorithms	K-Means Clustering, Hierarchical Clustering, Principal Component Analysis (PCA), t-SNE, Autoencoders (AE), Variational Autoencoders (VAE), Self-Organizing Maps (SOM) [17]
Task Types	Clustering (grouping similar formulations), Dimensionality Reduction (compressing feature space), Anomaly Detection (identifying process deviations)[17]
Pharma Application 1	Formulation design space clustering grouping formulations by dissolution behavior or stability using K-Means, supporting QbD (Quality by Design) approaches[17]

Table 4 Reinforcement Learning Full Overview

Point	Definition / Description
Definition	A learning paradigm where an agent learns to make sequential decisions by interacting with an environment, receiving rewards for desirable outcomes and penalties for undesirable ones. [15]
Working Principle	The agent follows a policy (\pi(a
Key Algorithms	Q-Learning, Deep Q-Network (DQN), Proximal Policy Optimization (PPO), SARSA, Actor-Critic Methods, Recurrent Neural Network (RNN)-based policy networks. [18]

Task Types	Sequential decision-making, process optimization, adaptive experimental design, iterative molecular generation. [15]
Pharma Application 1	De novo drug design RL agents generate novel molecules by sequentially adding atoms/fragments, rewarded for drug-likeness (Lipinski's rules), potency, and selectivity

2.1.2 Deep learning and neural networks

Deep learning is an advanced form of neural network modeling that uses multiple hidden layers to learn more complex patterns from large datasets. It is useful for solubility prediction, stability assessment, spectroscopy analysis, and tablet defect detection. These models can reduce trial-and-error work and improve formulation screening, but they still need good data, proper validation, and interpretability before they can be used confidently in regulated development. [19],[20],[21]

Neural networks are used to model complex relationships between formulation variables and product performance. Artificial neural networks (ANNs) have been applied to predict dissolution behavior, optimize drug-excipient ratios, and support quality-by-design-based formulation development by defining process parameter limits. They are useful when the effect of ingredients and processing conditions is non-linear and difficult to describe with simple equations.[22],[23]

2.1.3 Natural language processing (NLP)

Natural language processing (NLP) is used to extract useful information from unstructured sources such as research articles, patents, regulatory labels, and technical reports. It helps identify excipients, process parameters, formulation strategies, and drug-related relationships from large volumes of text, which makes literature mining and knowledge gathering much faster and more efficient. [24],[25]

NLP is also useful for supporting formulation decisions and development planning. This is valuable for tasks such as drug-excipient information extraction, stability data review, and identifying formulation trends across published studies. [26]

2.1.4 Genetic algorithms and evolutionary computing

Genetic algorithms and evolutionary computing are used to optimize difficult problems such as formulation composition, lead selection, and process parameters. They are based on repeated cycles of selection, crossover, and mutation, which help them search a large solution space and move toward better designs over time.[27]

They are especially useful in drug and formulation development because they can handle multiple objectives at once, such as improving potency while reducing toxicity, or improving tablet hardness while maintaining fast disintegration. This makes them valuable for de novo drug design, optimization of formulation variables, and other non-linear problems where traditional methods are less effective. [28]

2.2 Data Types Used in Formulation Science

2.2.1 Physicochemical properties (solubility, particle size, polymorphism)

Physicochemical properties are the measurable characteristics of a drug substance that strongly influence how it behaves during formulation and after administration. In preformulation studies, the most important physicochemical properties include solubility, particle size, and polymorphism, because these factors directly

affect dissolution, stability, bioavailability, manufacturability, and the overall performance of the dosage form. [29]

Solubility refers to the amount of drug that can dissolve in a given solvent and is one of the most critical factors controlling absorption and bioavailability. Particle size influences surface area, wetting, dissolution rate, and content uniformity, so reducing particle size often improves the performance of poorly soluble drugs. Polymorphism means that the same drug substance can exist in more than one crystalline form, and each form may show different melting point, solubility, stability, and dissolution behavior. [30]

2.2.2. Excipient databases

Excipient databases are organized collections of information about inactive ingredients. In formulation development, they help scientists choose suitable excipients by providing data on prior use, acceptable routes of administration, dosage forms, and safety-related information. [31]

The most important example is the FDA Inactive Ingredient Database (IID), which lists inactive ingredients present in FDA-approved drug products and can be used to support excipient selection during development. Other useful resources include Pharma Excipients. [32]

2.2.3 Process parameters and in-process controls

Process parameters are the controllable conditions used during pharmaceutical manufacturing, such as temperature, pressure, mixing speed, compression force, granulation time, and coating conditions. In solid dosage formulation, these parameters strongly influence critical quality attributes like hardness, disintegration, dissolution, content uniformity, and stability.[33]

In-process controls (IPCs) are the tests and checks carried out during manufacturing to ensure that the process remains within the desired limits.[34]

2.2.4 Clinical and in vivo data

Clinical and in vivo data are one of the data types used in formulation science. It means data collected from animal studies and human clinical studies that show how a formulation behaves in the body, especially for safety, bioavailability, pharmacokinetics, and product performance. [35],[36]

Clinical and in vivo data are importantbecause they provide direct evidence of how a drug product performs in biological systems. Such information is valuable for comparing formulations, supporting dose selection, and understanding.

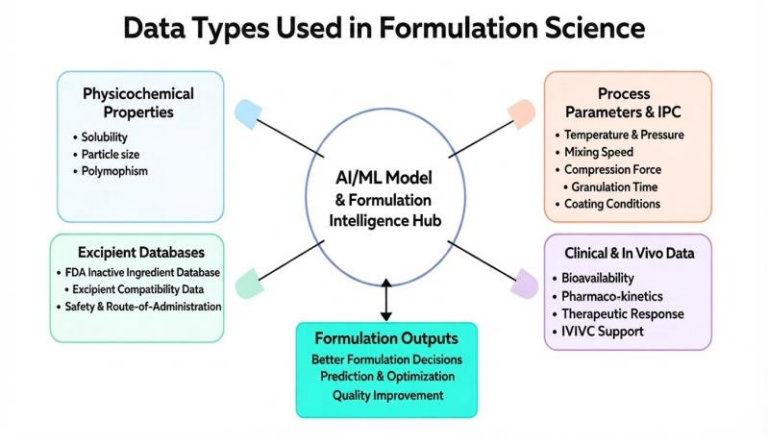

Figure 2 From Physicochemical Properties to Therapeutic Response: A Machine Learning Framework for Formulation Development

3. APPLICATIONS OF AI/ML IN SOLID DOSAGE FORMULATION

AI/ML can help in solid dosage formulation by predicting the right combination and quantity of ingredients such as drug, binder, and disintegrant. They can optimize tablet properties like hardness, dissolution, flowability, and stability. AI can also predict manufacturing problems such as poor compression or tablet defects before production.
This reduces trial-and-error, saves time and cost, and improves the quality of tablets and capsules.

3.1 Pre-formulation Studies

Pre-formulation studies are done before developing the final dosage form to understand the physical, chemical, and biopharmaceutical properties of the drug. They include tests such as solubility, pH, melting point, particle size, hygroscopicity, dissolution, powder flow, and polymorphism.

3.1.1 Drug-excipient compatibility prediction

Drug–excipient compatibility prediction is a key part of pre-formulation because it helps identify possible interactions before a formulation is finalized. If a drug and excipient are incompatible, the product may show reduced stability, poor manufacturability, or changes in performance. Traditionally, formulators rely on analytical methods such as DSC, FTIR, NMR, chromatography, and accelerated stability studies to evaluate these risks.[37],[38]
In recent years, artificial intelligence and machine learning have offered a faster and more data-driven way to study compatibility. For example, DE-INTERACT used PubChem fingerprints and a curated collection of more than 3,500 drug–excipient instances to build a predictive model, showing that molecular data can be used to estimate interaction risk. In another study, PharmDE was developed as a rule-based expert system using 532 data items collected from 228 articles and 60 interaction rules, which shows that knowledge-based methods can also support compatibility assessment.[39],[40]

3.1.2 Polymorphism and salt form screening

Polymorphism and salt form screening are essential steps in preformulation because they help identify the most suitable solid form of an API before development begins. A single drug substance may exist in multiple crystalline forms, and each form can differ in stability, solubility, hygroscopicity, and manufacturability, which makes early screening important for product performance and process selection. [41],[42]
Salt screening is used to improve properties such as dissolution behavior, chemical stability, and crystallinity, while polymorph screening helps detect alternate solid-state forms that may appear during development or storage. Even after a promising salt is selected, polymorphism can still occur, so both evaluations are usually needed to reduce development risk and support robust formulation design. [42],[43].

3.1.3 Solubility and permeability prediction (ADMET modeling)

Solubility and permeability prediction are core parts of ADMET modeling because they help estimate how well a compound dissolves and crosses biological membranes before extensive laboratory testing. [44] Modern ADMET platforms provide models for aqueous solubility, biorelevant solubility, pKa, logP, logD, and several permeability endpoints such as Caco-2 permeability, human effective permeability, MDCK permeability, skin permeability, and blood-brain barrier penetration. For example, ADMET Predictor reports models for water

solubility, simulated GI fluid solubility, and permeability-related endpoints. [45],[46] One study found that ADMET Predictor gave good predictions for LogD and was useful for categorizing permeability and metabolic stability, while another online tool, ADMET-PrInt, combined prediction with explainability so users could see which molecular features affected solubility and membrane permeability.[47]

3.1.4 API physicochemical characterization

API physicochemical characterization is the systematic evaluation of an active pharmaceutical ingredient to define the properties that influence formulation, processing, stability, and performance. It usually includes identification of solid-state form, particle size, solubility, stability, flow behavior, and related molecular or biological attributes.[48],[49]
This characterization is important because APIs can exist in different crystalline or amorphous states, and each form may show different melting point, solubility, dissolution rate, hygroscopicity, and stability. [50]

3.2 Formulation Design and Optimization

Formulation design and optimization means selecting the right ingredients, their quantities, and processing conditions for a drug product. AI/ML can predict properties like hardness, dissolution, stability, and drug release, reducing trial-and-error and helping find the best formulation faster.

3.2.1 Excipient selection using ML classifiers

Excipient selection using ML classifiers is a data-driven approach that predicts which excipients are most likely to work well with a given drug product before extensive laboratory screening.[51]
For small-molecule such as DE-INTERACT use molecular fingerprints and artificial neural networks to classify drug excipient compatibility as likely acceptable or risky. A newer stacking model combined Mol2vec and 2D molecular descriptors and reported strong performance, with accuracy 0.98, precision 0.87, recall 0.88, AUC 0.93, and MCC 0.86, and it detected 10 of 12 incompatible validation cases. [51]

3.2.2 Design of Experiments (DoE) augmented with AI

DoE provides the structured framework for studying factor effects, while AI helps analyze patterns in the data, predict outcomes, and suggest the most informative next experiments.[52],[53]
This approach is useful for optimizing excipient concentration, granulation conditions, compression force, coating parameters, and dissolution behavior. AI-assisted DoE can reduce the number of experiments needed, save time and materials, and make the optimization process more efficient. [53],[54]

3.2.3 Quality by Design (QbD) integration

QbD helps researchers link formulation variables and process parameters with critical quality attributes such as assay, dissolution, hardness, and stability. This reduces trial-and-error work, improves consistency between batches, and makes scale-up and regulatory submission more robust. [55],[56]
It is especially useful in solid dosage development because it supports better understanding of excipient behavior, process limits, and acceptable operating ranges.[57]

3.2.4 Formulation of BCS Class II & IV drugs (solubility enhancement)

BCS Class II and IV drugs are often difficult to develop because they show poor aqueous solubility, and Class IV compounds also have low permeability. To improve their oral absorption and

bioavailability, formulators use approaches such as particle size reduction, salt formation, pH adjustment, solid dispersions, cyclodextrin complexes, self-emulsifying systems, and lipid-based formulations.[58],[59]

These solubility-enhancement methods are commonly supported by characterization techniques such as PXRD, DSC, and SEM to confirm successful formulation changes. Overall, these approaches help make poorly soluble drugs more suitable for effective dosage form development.[60]

3.2.5 Fixed-dose combination product development

Fixed-dose combination (FDC) product development involves combining two or more active drugs into a single dosage form to improve convenience and treatment adherence.[61]

The main challenges in FDC formulation are dose balance, compatibility between the drugs and excipients, stability, and achieving proper dissolution from the final product. Because each active ingredient may behave differently, the formulation often needs careful optimization to make sure both drugs work effectively together.[62]

3.3 Manufacturing Process Optimization

Manufacturing process optimization means improving the process conditions and parameters used to produce a drug product. AI/ML can help optimize factors like mixing time, compression force, drying temperature, and speed to improve product quality and reduce manufacturing errors.

3.3.1 Wet granulation and dry granulation process control

In wet granulation, control often centers on liquid addition rate, mixing time, impeller speed, kneading, moisture level, and granule consistency, while in dry granulation the main variables are roll pressure, roll gap, roll speed, feed rate, ribbon density, and milling conditions.[63]

For wet granulation, process monitoring may use power consumption, moisture probes, temperature probes, or near-infrared sensors to track the end point and scale-up behavior. For dry granulation, control strategies often rely on in-line measurement of ribbon density, mass flow, and granule size distribution, especially in roller compaction systems.[64]

3.3.2 Tablet compression prediction of hardness, friability, disintegration

Tablet compression influences hardness, friability, and disintegration because compression force changes how tightly particles bond and how much porosity remains in the tablet.higher compression force increases hardness and lowers friability, but it can also slow down disintegration if the tablet becomes too dense.[65],[66]

Predictive studies have shown that tablet hardness can be estimated from NIR-based chemometric models, with one report giving an R^2 of 0.925 for predicted versus actual hardness. Friability has also been modeled from mechanical properties, and one study showed that it can be predicted accurately enough to support formulation optimization. [67],[68]

3.3.3 Coating process optimization (film coating, enteric coating)

In film coating, the most important variables usually include spray rate, atomization pressure, inlet air temperature, bed temperature, pan speed, and droplet size, because these directly affect coating efficiency and film quality.[69],[70]

For enteric coating, the process must also ensure that the polymer layer is strong enough to resist gastric fluid but still dissolves in intestinal pH.

This means controlling coating thickness, spray rate, drying conditions, and plasticizer level so the tablets remain intact in acid but release properly later.[71]

3.3.4 Continuous manufacturing real-time process monitoring

Continuous manufacturing uses real-time process monitoring to keep production running within the desired quality range instead of checking only at the end. It relies on sensors, data analytics, and process analytical technology (PAT) to track critical process parameters and critical quality attributes as the process runs.[72],[73]

3.3.5 Powder blending and mixing uniformity prediction

Powder blending and mixing uniformity prediction is used to estimate whether an API and excipients will distribute evenly throughout a blend before compression or encapsulation. This is important because poor mixing can cause content non-uniformity, segregation, and variability in tablet strength or dose.[74]

Prediction methods often combine material attributes such as particle size, density, flow behavior, and moisture sensitivity with process variables like mixing speed, time, and blender design. Recent machine-learning work showed that blend uniformity can be predicted from a large industrial dataset, and one neural-network model achieved an R^2 of 0.9919, showing strong ability to capture nonlinear blending behaviour.[75],[76]

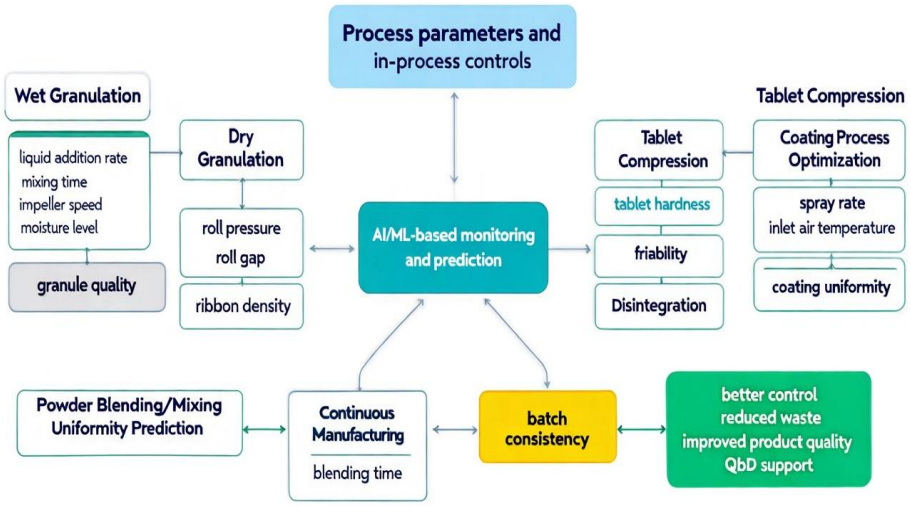

Figure 3 Machine learning Enhanced Process Optimization and Control Strategy for Pharmaceutical for Solid Dosage Manufacturing

3.4 Dissolution and Drug Release Modeling

Dissolution and drug release modeling studies how quickly and how much a drug is released from a dosage form.

AI/ML can predict drug release rates and dissolution profiles, helping researchers design formulations with the desired release pattern.

3.4.1 In vitro dissolution profile prediction

In vitro dissolution profile prediction is used to estimate how a tablet or capsule will release drug over time before running repeated laboratory tests. It is important because dissolution strongly affects bioavailability, product performance, and the ability to compare formulations during development.[77],[78]

Prediction methods include empirical models, artificial neural networks, spectroscopy-based approaches, and multivariate modeling from coating or formulation data. Recent studies show that coating quality attributes and spectroscopic signals such as Raman, NIR, OCT, and terahertz data can be used to predict dissolution behavior with useful accuracy.[79],[80]

3.4.2 In vitro–in vivo correlation (IVIVC) modelling

An in vitro–in vivo correlation (IVIVC) model is a predictive mathematical relationship between a dosage form. It is used to predict human performance from dissolution data and can reduce the need for repeated bioequivalence studies.[81]
The most common form is Level A IVIVC, which gives a point-to-point link between dissolution and absorption, while Level B and Level C give weaker statistical relationships.[82]
I

3.4.3 Modified/controlled release formulation optimization

Modified and controlled-release formulation optimization focuses on designing a dosage form that releases the drug at the desired rate and for the desired duration. The aim is to keep plasma levels within the therapeutic window, improve patient compliance, and reduce dosing frequency.[83],[84]
Optimization usually involves changing polymer type, polymer concentration, compression force, coating level, and matrix structure to control diffusion, swelling, erosion, or membrane permeability. QbD and DoE are commonly used to identify the most important variables and build a robust design space for sustained or extended release.[85]

3.5 Analytical Method Development

Analytical method development involves creating methods to identify, measure, and test the quality of a drug.
AI/ML can help optimize analytical conditions and improve the accuracy, speed, and reliability of drug testing.

3.5.1 Near-infrared (NIR) spectroscopy + ML for blend uniformity

Near-infrared (NIR) spectroscopy combined with machine learning is used to monitor blend uniformity in real time during powder blending. This approach helps determine whether the API is evenly distributed in the blend without stopping the process for repeated sampling and laboratory testing.[86]
NIR provides spectral data linked to blend composition, while ML or multivariate models such as PLS, PCA, random forest, or neural networks convert that data into predictions of uniformity or blend endpoint. In one study, a PLS calibration model tracked API content during blending across 67 production-scale batches and supported the determination of the blend uniformity endpoint. A newer ML study reported very strong prediction performance for blend uniformity, with an R^2 of 0.9919 using a feed-forward neural network.[87]

3.5.2 Raman spectroscopy for tablet content analysis

Raman spectroscopy is a rapid, nondestructive technique for tablet content analysis and content uniformity testing, especially when transmission Raman is used to sample the bulk of the tablet. It has been validated for quantifying APIs in whole tablets and can also be applied to low-dose and multi-API products, making it useful for quality control in solid dosage manufacturing.[88]

Raman methods are often paired with chemometric models such as PLS to convert spectral data into accurate content predictions. Studies have shown good agreement with reference methods like HPLC, and one transmission Raman application reported an R^2 of 0.997 with low prediction error for tablet assay.[89]

3.5.3 Computer vision for tablet defect detection

Computer vision is increasingly used for tablet defect detection because it can inspect tablets quickly and automatically during manufacturing. Deep learning systems can identify defects such as chipping, breaking, color non-uniformity, speckling, sticking, picking, and surface abrasion from tablet images.[90]

These methods usually rely on cameras or digital microscopes, followed by image processing and neural-network-based classification or object detection. Recent studies showed that computer vision can detect defects with high precision, and one approach combined Faster R-CNN with deterministic image analysis to quantify sticking, picking, chipping, and abrasion without false negatives. Another study used CNN-based tablet inspection to identify defective tablets before blister packing, helping reduce wastage and improve quality control.[91]

3.5.4 Process Analytical Technology (PAT) integration

Process Analytical Technology (PAT) integration is a science-based approach to monitoring and controlling pharmaceutical manufacturing in real time. It uses in-line, on-line, or at-line analytical tools to track critical process parameters and connect them with critical quality attributes so that product quality is built into the process rather than checked only at the end.[92]

The main benefit of PAT integration is improved control, reduced waste, shorter cycle times, and support for real-time release testing. It is especially valuable in modern pharmaceutical development because it strengthens QbD, scale-up, and regulatory confidence.[93]

3.6 Stability Prediction

Stability prediction helps determine how long a drug product remains safe, effective, and stable during storage.

AI/ML can predict the effects of temperature, humidity, and time on drug stability, reducing the need for lengthy stability studies.

3.6.1 Accelerated stability study data modelling

Accelerated stability study data modeling uses data collected under elevated temperature and humidity conditions to predict how a drug product will behave during long-term storage. This approach helps estimate shelf life, identify degradation trends earlier, and reduce the time needed for conventional stability studies.[94]

A common approach is to fit degradation data with kinetic models such as zero-order or first-order kinetics and then use Arrhenius-based relationships to extrapolate stability at recommended storage conditions. More advanced methods, such as accelerated predictive stability and ASAP-style modeling, use temperature and relative humidity as predictors to calculate degradation rate constants and forecast when a product will reach specification limits.[95]

3.6.2. Prediction of degradation pathways

Prediction of degradation pathways is used to anticipate how a drug or chemical may break down under stress, storage, or environmental conditions.This helps identify likely degradants early and supports forced-degradation studies,

impurity profiling, and stability-indicating method development.[96]

Most pharmaceutical pathway prediction systems use expert-rule knowledge bases combined with structural alerts to suggest probable transformations such as oxidation, hydrolysis, dealkylation, dimerization, and photodegradation. Zeneth is a well-known system for predicting forced degradation pathways, and benchmarking studies showed that its predictive coverage improved as the knowledge base expanded, with observed degradant prediction rising from 31% to 54% over time.[97]

3.6.3 Shelf-life estimation using AI-driven models

Shelf-life estimation using AI-driven models uses data from storage studies to predict how long a product will remain within specification. These models can combine temperature, humidity, time, and degradation measurements to forecast product stability more quickly than traditional trial-based methods.[98]

Common AI approaches include artificial neural networks, regression-based models, and accelerated predictive stability frameworks such as ASAP. These methods can capture nonlinear degradation behaviour and help estimate when a product will reach its failure limit or shelf-life endpoint.[99]

3.7 Generative AI in Formulation Discovery

Generative AI can help create and explore new drug formulations by suggesting suitable drugs and excipients.

It can generate formulation options and predict their properties and performance, making formulation discovery faster and reducing trial-and-error experiments.

3.7.1 Generative adversarial networks (GANs) for novel formulation design

Generative adversarial networks (GANs) can be used for novel formulation design because they learn patterns from existing data and then generate new candidate structures or composition ideas. This makes them useful for exploring new molecules, formulations, or design spaces more quickly than trial-and-error experimentation.[100]

A GAN works with two networks: a generator that creates new outputs and a discriminator that checks how realistic those outputs are. By competing with each other, the model can produce candidates that resemble valid pharmaceutical designs while still introducing new combinations that may not have been tested experimentally.[101]

3.7.2 Large language models (LLMs) for literature mining and formulation suggestion

Development, large language models (LLMs) are being explored as tools for literature mining, formulation suggestion, and decision support across the drug development pipeline. Recent reviews note that LLMs can help extract patterns from biomedical literature, assist with protocol generation, and support formulation-related tasks such as predicting excipient interactions, solubility, stability, and process risk.[102]

LLMs are most useful when they are guided by domain-specific data and expert oversight. A study showed that fine-tuned LLMs could recommend excipients for fused deposition modeling formulations and predict filament properties from a dataset of over 1,400 formulations, but it also showed that model choice and parameter tuning strongly affect performance. Another line of work suggests that LLMs can act as “virtual instruments” for drug formulation by encoding formulation metadata into structured prompts,

which may help with early-stage composition design. [103],[104]

3.7.3 Autonomous formulation platforms (self-driving labs)

Autonomous formulation platforms, or self-driving labs, combine robotics, automation, and AI to run a closed loop of experiment design, execution, analysis, and optimization with minimal human intervention. or materials formulation, this can speed up screening, reduce trial-and-error, and help identify better excipient combinations or process conditions more quickly.[105]

These platforms are especially promising for complex development tasks where many variables interact, but they still require high-quality data, careful validation, and human oversight to ensure reliable and regulator-ready decisions.[106]

4. AI/ML TOOLS AND PLATFORMS USED IN SOLID DOSAGE DEVELOPMENT

4.1 Commercial software (MODDE, JMP, Simca)

These three are commercial statistical and data analytics software packages widely used, formulation optimization, and process development:

Integration: SIMCA 13 can directly communicate with MODDE 9 by exporting scores, enabling a "quality by design" solution.

Table 5 Commercial statistical and data analytics software

Software	Primary Purpose	Developer	Key Applications in Pharma
MODDE	Design of Experiments (DoE)	Sartorius (formerly Sartorius Stedim Data Analytics)	Experimental design & optimization for product/process development,qualitycontrol, R&D. [107]
JMP	Statistical discovery & predictive analytics	JMP (subsidiary of SAS Institute)	Data preparation, statistical analysis, graphing, predictive modeling, machine learning for scientists/engineers. [108]
SIMCA	Multivariate data analysis (MVA)	Sartorius (formerly Umetrics)	Multivariate calibration, predictive modeling, spectroscopic data analysis, batch process monitoring, chemometrics reddit. [109]

4.2 Open-source frameworks (scikit-learn, TensorFlow, PyTorch)

Scikit-learn is a classical machine learning library ideal for small/medium datasets with traditional algorithms like SVM and random forests, requiring no GPU. TensorFlow (Google) handles deep learning with scalable production deployment for mobile/cloud, excellent for CNN-based image analysis. PyTorch (Facebook's FAIR) prioritizes research flexibility for novel neural networks and state-of-the-art deep learning models.[110],[111]

In solid dosage development, these frameworks predict tablet disintegration time (scikit-learn NODE achieved R^2=0.9805), detect tablet defects using CNNs/YOLOv5 (99.2% accuracy), predict dissolution profiles, and optimize 3D-printed formulation parameters. Scikit-learn is best for formulation optimization with small datasets; PyTorch/TensorFlow for image-based quality control; PyTorch for research prototyping.[112]

4.3 Pharmaceutical-specific platforms (Dataiku, Signals Notebook, TIBCO)

Dataiku is an enterprise AI and data science platform useddevelopment for tasks such as pharmacovigilance, drug repurposing, clinical data analysis, and manufacturing support. It is suitable for the full development pipeline because it helps teams prepare data, build predictive models, and manage workflows in a regulated environment.[113]

Signals Notebook is a cloud-based electronic lab notebook (ELN) used mainly for early research and formulation work. It helps scientists record experiments, organize chemical and biological data, and maintain a structured digital record of laboratory work, which is useful in preformulation and solid dosage development.[114]

TIBCO Spotfire is a visualization and analytics platform used in pharma for dashboards, clinical data analysis, and process monitoring. It is especially helpful in manufacturing and quality control because it turns complex datasets into interactive charts and decision-support visuals.[115]

4.4 Integrated laboratory automation and robotics

The major value of laboratory automation lies in its ability to increase throughput, reduce operator-dependent variability, and support more reliable generation of high-quality datasets for AI and machine learning models.This is especially relevant development, where robust and well-structured experimental data are essential for linking material attributes, process parameters, and final product performance.[116],[117]

Integrated robotics can support high-throughput screening, automated Design of Experiments execution, formulation preparation, and real-time monitoring of critical quality attributes. These platforms are increasingly connected with orchestration software and laboratory infrastructure, enabling more seamless movement toward autonomous or self-driving laboratories in which experiment planning, execution, and data feedback occur in a closed loop. [118]

4.5 Digital twins in manufacturing

Digital twins in manufacturing are virtual models of physical equipment, processes, or entire production systems that are continuously updated using real-time data from the shop floor. They allow manufacturers to simulate process changes, predict performance, and identify possible issues before they affect actual production. This makes them useful for improving efficiency, reducing downtime, and supporting better decision-making in modern industrial operations.[119]

Digital twins can be especially valuable for process monitoring, quality control, and optimization of production workflows. They help reduce variability, improve consistency, and support predictive maintenance in complex manufacturing environments. [120]

5. CHALLENGES AND LIMITATIONS

5.1 Data-Related Challenges

5.1.1 Limited and proprietary datasets

Many of the predictive capabilities listed in previous sections depend on data that companies don't have a strong incentive to release. Whereas, pharmaceutical formulation know-how is usually protected as a trade secret; the information and data on the amount of excipients, processing conditions, compatibility information and so on that took years, and a significant investment of resources, to develop is carefully guarded, and if published, would give a working formula to competitors.[121]

A 2025 industry survey shows the scale of the resulting problem: roughly half of life-sciences

respondents named "AI-ready data" as a leading hurdle to AI adoption, and the large majority said their organization's data was not structured or ready for AI use at all. [121]

A few late 2025, several big pharmaceuticals (AbbVie, Johnson & Johnson, Astex Pharmaceuticals, Bristol Myers Squibb and Takeda) entered into a federated learning program, specifically designed to allow each company to finetune an open-source, shared structural biology model using their own proprietary data without compromising it, which remains on their own servers.[122]

5.1.2 Data heterogeneity and inconsistency

Pharmaceutical data is routinely fragmented across clinical trials, laboratory results, and manufacturing systems, often using inconsistent naming conventions and incompatible platform formats that complicate even basic integration before any modeling can begin. [15]

Tablet hardness testing illustrates the scale of this problem well. Depending on which instrument and convention a laboratory uses, the same physical measurement can be reported in Newtons, kiloponds, pounds-force, or Strong-Cobb units, measured on meaningfully different equipment that has historically included Monsanto, Stokes, Pfizer, Strong-Cobb, and Schleuniger hardness testers. [123],[124]

5.1.3 Lack of standardized data formats in pharma

Standardization efforts such as CDISC and MedDRA have made genuine progress in harmonizing terminology for clinical trial data, but no comparably adopted framework exists specifically for solid dosage formulation and manufacturing data, leaving most of the work of reconciling formats to manual, labor-intensive curation before a dataset becomes usable for modeling. This absence of a shared data standard is arguably a quieter but equally limiting barrier than data scarcity itself even abundant data offers little benefit to a machine learning pipeline if every source structures it differently.

5.1.4 Imbalanced datasets (rare failure events)

Manufacturing quality-control research more broadly has found that even in datasets specifically curated to study defects, defective items typically represent only around 10 to 15 percent of total output, and the proportion is generally far lower still in tightly controlled, validated commercial pharmaceutical lines. [125]

Benchmark studies of imbalanced quality-control data show the same pattern at a smaller scale defect datasets such as the NEU Surface Defect Dataset contain only a few hundred images per defect category against a much larger pool of normal samples, which is why accuracy alone is considered a misleading metric in this setting, with recall on the minority class used instead to evaluate whether a model can actually catch rare events. [126]

5.2 Model-Related Challenges

5.2.1 Black-box nature of deep learning (interpretability gap)

A second modelling problem is whether or not someone can explain why a model concluded a specific point. Many high-performance deep learning architectures operate as a black box, delivering predictions without explaining why they make those predictions; this opacity is particularly important for a prediction about a tablet batch's quality one of the most important and safety-critical features of a pharmaceutical product.[127]

The European Medicines Agency has expressed a clear preference for interpretable models in drug

development, and where black-box models are used because of genuinely superior performance, its framework requires accompanying explainability metrics and thorough documentation of model architecture and behavior. [128]

The deeper concern is not simply technical inconvenience: decisions made using black-box outputs can directly affect patient safety, and a decision that cannot be traced back to a clear rationale is difficult to defend during a regulatory audit, regardless of how statistically accurate it has proven in practice. [128]

5.2.2 Domain shift between preclinical and clinical data

A model's reliability also depends on whether the conditions it is deployed in actually resemble the conditions it was trained on. Process parameters developed and validated at laboratory scale frequently cannot be directly applied once a product moves to large-scale commercial manufacturing, because the equipment, operating methods, and production environment at lab scale differ meaningfully from those of a full production facility. [129]

This shift in statistical patterns the model has learned is not a matter of calibration for an AI/ML model; it's a fundamental change. A model which was trained on a small-scale, tightly controlled lab-bench data set is being asked to make predictions for a commercial line operating at different batch sizes, mixing dynamics and heat-transfer properties. That's a reason why pharmaceutical scale up has traditionally required structured pilot scale studies and protocols for technology transfer, prior to the process being considered validated at commercial scale.

5.3 Regulatory and Validation Challenges

5.3.1 Lack of clear FDA/EMA guidelines for AI-driven formulation

The FDA requires for an AI model used in formulation development was that no single, finalized rulebook existed. That gap has been narrowing quickly. The FDA issued its first draft guidance specifically addressing AI in drug and biological product development in January 2025, introducing a structured, risk-based credibility assessment framework for evaluating AI models used to support regulatory submissions. [130]

This guidance still remains draft until 2025 and FDA and EMA jointly published "Guiding Principles of Good AI Practice in Drug Development" in January 2026, which extends the entire product lifecycle, including manufacturing.[131]

5.3.2 Model validation and qualification requirements

Traditional analytical methods are validated once and used and trusted to do so for as long as the method and instrument do not change. AI/ML models complicate this picture because they are not all "locked" in this same way: a model can be "locked", giving the same results for the same input each time; or "adaptive", continually updating to new data as it becomes available, and these two types of models pose fundamentally different questions for validation. [132]

There is no single answer as to what regulators will do about the adaptive case. There is an emerging trend in the FDA approach toward allowing some model evolution, via a Predetermined Change Control Plan, where the range of acceptable model changes are prespecified before deployment takes place, and the EMA has focused more on fixed algorithm configurations with fewer post-deployment changes. [133]

5.3.3 Documentation and audit trail for AI decisions

AI model modifies a process during the batch, GMP would expect to see these changes documented in the same manner as all other manufacturing documentation "what" changed, "when" it changed, and "why" it was changed. It is not new a secure, time-stamped audit trail of who and/or what changed and why is required by longstanding data integrity requirements including the FDA's 21 CFR Part 11 and EU's GMP Annex 11. [134]

The recently proposed Annex 22 from the EU focuses on the specific needs for the use of AI and machine learning in the pharmaceutical manufacturing process, while FDA issued its risk-based Computer Software Assurance guidance in September 2025, to modernize software validation in production and quality system activities. [135]

5.4 Ethical Considerations

5.4.1 Bias in training data affecting formulation decisions

Bias in AI/ML formulation models is seldom about a model being purposefully discriminatory typically, it just reflects the earlier identified constraints of data within this section here. Pharmacogenomics has a clear, measurable basis: models trained predominantly on data from one population or region can consistently underperform when applied to others.[136]

And CYP2D6 alone metabolizes roughly 25 percent of all pharmaceuticals, while the prevalence rates for "poor metabolizer" and "ultra-rapid metabolizer" phenotypes vary widely by ancestry--the former class represents less than one percent in East Asian populations but five to 10 percent in European populations, whereas ultra-rapid metabolizers are up to 29 percent prevalent in certain Ethiopian and Saudi Arabian groups but just one to two percent among Northern Europeans.[136]

5.4.2 Patient safety implications of AI-driven decisions

When a formulation or quality decision is influenced by an AI/ML model, the implications of making a mistake are concrete. Between 2022 and 2024, the number of healthcare AI malpractice-related claims increased by 14%. A common theme in this litigation is the danger of using an AI tool outside of the population or conditions it was truly validated on, such as using a model that was primarily trained on one demographic group to make decisions that affect another.[137]

5.4.3 Transparency and accountability

As AI/ML becomes more deeply entrenched in decisions regarding formulation and production, the accountability question emerges as who should be held responsible for a product failure or patient harm that results from an AI-influenced decision? A 2026 analysis of AI's impact on pharmaceutical decision-making reveals that liability risks from an AI decision that can be traced back to a flawed AI system are no longer theoretical and that AI models are now involved in consequential decisions across the spectrum of target selection to process optimization a pace and scale that current corporate governance frameworks have not been designed to manage.[138]

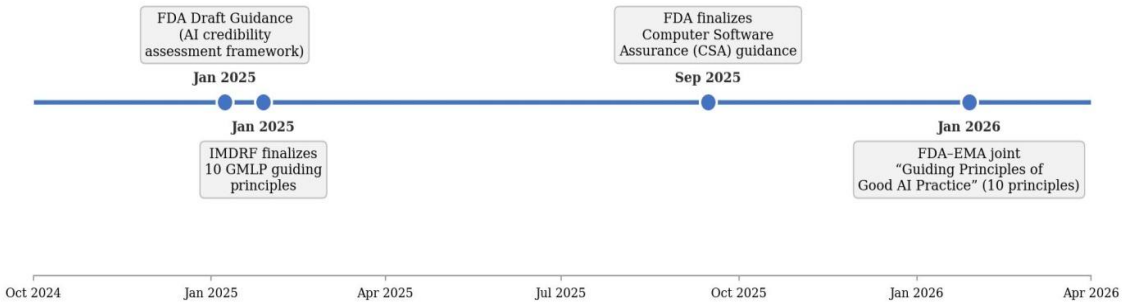

Figure 4 Timeline of Major FDA and EMA AI-in-Drug-Development Regulatory Milestones, January 2025 – January 2026

6. CASE STUDIES AND REAL-WORLD EXAMPLES

6.1 AI-driven tablet formulation optimization (e.g., Pfizer, AstraZeneca initiatives)

Pfizer's development of Paxlovid, its oral antiviral for COVID-19, is one of the most visible examples of AI-assisted formulation and delivery optimization at a major pharmaceutical company. Pfizer combined supercomputing resources with advanced computational modeling to identify a formulation that could be delivered orally rather than intravenously, a change that allowed patients to take the treatment at home rather than in a hospital setting. [139]

The reason I wanted to feature this example is less because it is tablet compression optimization in the strictest sense that we defined in Section 3.3, but rather how an AI-augmented computational method fed directly into a formulation decision that had a tangible result: turning a treatment that was once administered in a hospital into something a patient can pick up at their local pharmacy to take at home.[139]

6.2 ML for dissolution prediction in immediate-release tablets

A 2025 study published in The AAPS Journal offers a directly relevant, well-documented example. Researchers produced 377 direct-compression tablet formulations in-house and measured their drug release profiles at 11 time points across 480 minutes under dynamic dissolution conditions, then trained six different machine learning techniques to predict the resulting release curves from formulation composition alone. [140]

Random forest and extreme gradient boosting performed best, achieving five-fold cross-validation R^2 values of 0.635 and 0.601 respectively, with root-mean-square errors around 13 to 14 percent. [140]

6.3 Deep learning in coating defect detection

In 2025, researchers at Merck & Co. applied a convolutional neural network to detect coating defects on film-coated tablets, addressing a specific practical limitation of earlier approaches. A previous detection method had achieved high accuracy, but required tablets to be manually placed in a 3D-printed tray for consistent imaging, a setup not feasible on an actual production line. The Merck team's CNN-based approach could instead inspect tablets directly on the line without this fixturing requirement, simplifying integration into real manufacturing. [141]

Such an application has already been successfully implemented by a pharmaceutical firm that needed

to inspect hundreds of thousands of tablets a day for a printed logo and found this deep learning vision system was able to distinguish true defects from false alarms with greater accuracy than the previous rule-based approach, resulting in lower batch waste due to rejections. [142]

6.4 NLP for excipient knowledge extraction from literature

A 2025 study built specifically to extract pharmaceutical manufacturing information from patents offers a concrete, measurable example of NLP applied to excipient knowledge extraction. Researchers collected 208,596 pharmaceutically relevant patents and built a two-stage system: an unsupervised model to identify which sections of a patent actually contained manufacturing information, followed by a named entity recognition model trained to extract specific entities, including excipients, from those sections. [143]

The section-identification model achieved a Cohen's kappa of 91.1 percent agreement with manual review, and the named entity recognition model reached an overall F1-score of 84.2 percent, with excipient-specific extraction performing somewhat lower at 81.1 percent. [143]

6.5 Self-driving laboratories for automated formulation screening

Intrepid Labs is a spin-off from the self-driving-lab academic research, with a machine-learning powered robotic platform that can explore formulation design space of up to one billion possible candidate formulations. The company's CEO said the key feature of the platform is that it doesn't need a big curated initial set of data, it can start exploring a formulation space without any data whatsoever and learn from the results of their experiments as they go.[144]

To illustrate NLP's application to excipient knowledge extraction, a study was developed specifically for extracting pharmaceutical manufacturing information from patents was considered in 2025. They collected 208,596 pharmaceutically relevant patents and created a two-stage process: first an unsupervised system to check if a patent has any manufacturing information in it and second a named entity recognition system trained to identify manufacturing-related entities like excipients in the manufacturing part of the patents. [145]

Table 7 Representative Case Studies of AI/ML Application in Solid Dosage Formulation, 2024–2026

Source/Organization	Application	AI/ML Method	Key Reported Result
Pfizer	Oral antiviral formulation/delivery optimization (Paxlovid)	Computational modeling + supercomputing	Enabled oral vs. intravenous delivery
AAPS Journal (2025)	Dissolution profile prediction, 377 IR/CR tablet formulations	Random forest, XGBoost	$R^2 = 0.635$ (RF, 5-fold CV)
Merck & Co. (2025)	Film-coated tablet coating-defect detection	Convolutional neural network	Inspection without manual tray-fixturing
Industrial deployment (Cognex)	Printed-logo defect detection on tablets	Deep learning vision (edge AI)	Reduced batch waste from false rejections
Alvarado-Maldonado et al. (2025)	NLP extraction of manufacturing data from 208,596 patents	BiLSTM-CRF named entity recognition	F1 = 84.2% overall; 81.1% for excipients

Intrepid Labs	Self-driving formulation screening platform	ML-guided robotics	Explored design spaces up to ~1 billion candidates
RSC Digital Discovery (2025)	Self-driving injectable formulation (curcumin)	Semi-autonomous closed-loop robotic formulator	7 lead formulations, solubility >10 mg/mL

7. FUTURE PERSPECTIVES

7.1 multi-omics and patient-specific data for personalized solid dosage forms

The approach of personalizing formulation decisions is expanding to incorporate information from the multi-omics, that is, all the information from an individual patient's genome, transcriptome, proteome, and metabolome, together and not just the genetic data. Models, for instance, DeepDRA, which demonstrated a precision-recall area under the curve of 0.99 in predicting drug response with transcriptomic and genomic data and deep learning, were cited. [146] For solid dosage formulation, this would suggest a future scenario that would become an opportunity rather than a risk: When a patient's true metabolic and genomic profile is available and not an average of the population, in principle, a model could be used to fine-tune the formulation or the release rate recommendation for that individual's actual CYP enzyme profile, as opposed to the more general classification of population categories.

7.2 Federated learning for cross-industry data sharing without IP compromise

A relatively early, but true example of cross-industry collaboration at scale is the federated learning project described earlier in this review (Sections 5.1.1 and 5.4.4) that involved the creation of shared infrastructure by AbbVie, Johnson & Johnson, Astex Pharmaceuticals, Bristol Myers Squibb and Takeda to optimise a structural biology model without sharing proprietary data.

7.3 Explainable AI (XAI) for regulatory-ready models

This is a clear preference in regulatory thinking as discussed in Sections 5.2.2 and 5.3.4, and in cases where black-box models are applied, EMA has stated the expectation of a document that explains the model.

However, the current limitations in achieving interpretability (introduced earlier in this review) do not make the full gap close with interpretability tools like SHAP and attention-based visualization. It's not going to be one breakthrough technique that will dominate the futurethere's going to be a combination of regulatory expectation and capability, and that disparity right now is likely to be narrowed but not eliminated as explainability techniques evolve, become more effective, and as other frameworks like the FDA's Predetermined Change Control Plan (Section 5.3.2) become more entrenched.

7.4 Digital twins for end-to-end pharmaceutical manufacturing

Although digital twins have been used mainly as a manufacturing tool in this review (Section 4.5), they are now being proposed as a tool that could be applied to the entire drug development lifecycle, from early discovery to commercial manufacturing. A digital twin in continuous manufacturing that incorporates process analytical technology (PAT) has been shown to increase active ingredient consistency to 99.95 percent, and patient-specific digital twins in early personalised-medicine applications have been shown to predict optimal doses within 7 percent of real clinical

results, according to a review in September 2025.[147]

A specific example of this phenomenon in solid dosage manufacturing is a study published in March 2025 which developed a digital real-time-release-testing (RT-RTT) strategy based on the ConsiGma-25 continuous tablet manufacturing line that combined and validated a set of residence-time-distribution (RTD) models and material-tracking algorithms through a study to track single tablets from the manufacturing line to the process data that produces them, thereby tracing individual tablets throughout the manufacturing process. [148]

7.5 Convergence of AI with 3D printing for personalized dosage forms

Unlike the tablet manufacturing industry, three-dimensional printing has a unique value proposition: it could provide a single dosage form developed to the specific dose, release or drug combination of a single patient. A 2025 study from the University of Mississippi's Pharmaceutical Engineering and 3D Printing Lab specifically targeted the practical issue of scaling beyond single-unit production, optimizing print parameters for both batch and continuous manufacture of 3D-printed dosage forms using AI/ML. [149]

The most obvious short-term use is the polypill: a single 3D-printed dosage form with multiple compartments, each with a different release profile, which helps to minimize the pill burden for patients with multiple chronic diseases.[150]

Table 8 Future Directions for AI/ML in Solid Dosage Formulation: Current Maturity and Primary Barriers

Direction	Current State (2025–2026)	Primary Barrier
Multi-omics personalization	Research-stage; e.g., DeepDRA AUC = 0.99	Population diversity in training data
Federated learning	Early industry pilots (5 major companies)	Requires new technical/legal infrastructure
Explainable AI (XAI)	Tools (SHAP, attention maps) narrow but do not close the gap	Inherent black-box/performance trade-off
Digital twins	CM digital twins reaching 99.95% API consistency	Not yet widely deployed at tablet level
AI + 3D printing	Print-parameter optimization demonstrated	Scale-up beyond single-unit production
Autonomous formulation labs	Operating, but human-supervised	Validation/documentation framework
Regulatory frameworks	FDA-EMA joint guiding principles (Jan 2026)	High-level principles, not yet operational rules

CONCLUSION

AI and machine learning are becoming useful parts of modern pharmaceutical formulation. They can support researchers in areas such as preformulation, excipient selection, manufacturing, dissolution studies, and stability testing. By learning from existing experimental data, these technologies can help reduce repeated laboratory work, identify useful formulation conditions, and speed up the development process. At the same time, AI/ML should not be seen as a complete replacement for laboratory experiments or scientific judgment. Their performance depends strongly on the quality and quantity of available data. In pharmaceutical research, much of the

useful data is limited, confidential, or stored in different formats, which makes model development difficult. Some models can also produce results that are hard to explain or validate, creating additional concerns for quality control and regulatory approval.

Future progress will require closer cooperation between formulation scientists, data scientists, pharmaceutical companies, and regulatory authorities. Better datasets, proper model validation, transparent methods, and practical regulatory guidelines will be important. Overall, AI/ML has strong potential to make formulation development faster and more efficient, but its success will depend on how responsibly and effectively these tools are integrated with established pharmaceutical practices.

REFERENCES

1. Dwivedi P, Kumari A, Devi A, Maqbool SZ. Tablets as a dosage form: a comprehensive review of design, manufacture, evaluation, and innovations. Int J Novel Res Dev. 2025;10(7):a110-a129.
2. Naharros-Molinero A, Caballo-González MÁ, de la Mata FJ, García-Gallego S. Shell formulation in soft gelatin capsules: design and characterization. Adv Healthc Mater. 2024;13(19):2302250. doi:10.1002/adhm.202302250.
3. Khandelwal H, Chaturvedi A, Gupta M. Formulation and evaluation of mouth dissolving tablets: a review. Res J Pharm Technol. 2019;12(12):6060-6068. doi:10.5958/0974-360X.2019.01050.3.
4. N.S. University. Powders [Internet]. Chhattisgarh: N.S. University; 2025 [cited 2026 Jun 20].
5. Joshi S, Singh A, Kumar R, Patel V, Sharma P. Artificial intelligence (AI) formulation and dosage form design: transforming the future of drug development. AAPS Open. 2025;11:5. doi:10.1186/s41120-024-00109-7.
6. Kalonia DS, Vora LK, Thakur RRS. Artificial intelligence and machine learning guided optimization in drug delivery. Trends Biotechnol. 2026. doi:10.1016/j.tibtech.2026.01.001.
7. Upperton Pharma Solutions. The transition from formulation development to Phase 1 clinical trials [Internet]. Nottingham (UK): Upperton Pharma Solutions; 2024 [cited 2026 Jun 20].
8. Deloitte. AI in pharma and life sciences [Internet]. New York (NY): Deloitte; [cited 2026 Jun 20].
9. SCW.AI. AI in pharma: use cases, success stories, and challenges in 2026 [Internet]. Atlanta (GA): SCW.AI; 2025 Jan 23 [cited 2026 Jun 20].
10. Coherent Solutions. Artificial intelligence in pharmaceuticals and biotechnology: current trends and innovations [Internet]. Minneapolis (MN): Coherent Solutions; 2026 Apr 24 [cited 2026 Jun 20].
11. McKinsey & Company. Generative AI in the pharmaceutical industry: moving from hype to reality [Internet]. New York (NY): McKinsey & Company; 2024 Jan 9 [cited 2026 Jun 20].
12. Gambhir D, Tiwari K, Saini G, Dhakad H, Yadav K. Pharmaceutical science in the era of artificial intelligence. Zenodo [Internet]. 2026 Feb 27 [cited 2026 Jun 20]; Version 1.
13. Panda SK, Dash SK, Samal SK, et al. Artificial intelligence product development: current applications and future perspectives. Smart Health. 2024;34:100501. doi:10.1016/j.smhl.2024.100501.
14. Obaido G, Djoumbou-Feunang Y, Parvin R, et al. Supervised machine learning in drug discovery and development: algorithms, applications, challenges, and prospects.

WIREs Comput Mol Sci. 2024;14(6):e1744. doi:10.1002/wcms.1744.
15. Clinical Programming Team. Machine learning in the pharmaceutical industry [Internet]. Hitchin (UK): Quanticate; 2025 Sep 25 [cited 2026 Jun 20].
16. Polański J, Kaczmarek K, Tkacz A, Dobosz A. Unsupervised learning in drug design from self-organization to deep learning. Molecules. 2022;27(5):1523. doi:10.3390/molecules27051523.
17. Bruce JL. An overview of unsupervised machine learning in drug discovery, pharmacogenomics, and pharmacovigilance. Biomed Res Clin Rev. 2025;11(2):1-6. doi:10.31579/2692-9406/220.
18. GeeksforGeeks. Supervised vs Unsupervised vs Reinforcement Learning [Internet]. GeeksforGeeks; 2026 Mar 24 [cited 2026 Jun 20].
19. Pushkarova Y. Artificial neural networks . In: Chemical engineering: innovations in materials and processes [Internet]. Tallinn: ROUTE Academic Publishing; 2025 [cited 2026 Jun 20].
20. Shunmugaperumal T. Artificial intelligence and deep/machine learning or artificial neural network. Pharma Focus Europe [Internet]. 2022 [cited 2026 Jun 20].
21. Jha NK, Sharma A, Bhatia S, et al. Artificial intelligence product development: current applications, challenges, and future opportunities. Eur J Pharm Sci. 2025;205:107043. doi:10.1016/j.ejps.2025.107043.
22. Elbadawi M, Gaisford S, Basit AW. Advanced machine-learning techniques in drug discovery. Drug Discov Today. 2021;26(3):769-777. doi:10.1016/j.drudis.2020.11.009.
23. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. Adv Drug Deliv Rev. 2021;175:113806. doi:10.1016/j.addr.2021.05.027.
24. Withers CA, Rufai AM, Venkatesan A, Tirunagari S, Lobentanzer S, Harrison M, et al. Natural language processing in drug discovery: bridging the gap between text and therapeutics with artificial intelligence. Expert Opin Drug Discov. 2025;20(6):765-783. doi:10.1080/17460441.2025.2490835.
25. Bhatnagar R, Farhangfar A, Capan M, Sikaroodi M, Vangala S, Tatonetti NP, et al. How can natural language processing help model-informed drug development? Clin Transl Sci. 2022;15(7):1589-1601. doi:10.1111/cts.13297.
26. Fast Data Science. AI in clinical trials: transforming drug development with artificial intelligence [Internet]. London (UK): Fast Data Science; 2024 Aug 9 [cited 2026 Jun 20].
27. Hexaware Technologies. Evolutionary computation [Internet]. Navi Mumbai (India): Hexaware Technologies; [cited 2026 Jun 20].
28. Rana PS, Deep K. Ensemble of genetic algorithm and artificial neural network for pharmaceutical process optimization. Appl Soft Comput. 2015;27:456-469. doi:10.1016/j.asoc.2014.11.055.
29. Arora P, Saini V. Physico-chemical properties of solid drugs: a review. Asian J Pharm Technol. 2019;9(1):61-66. doi:10.5958/2231-5713.2019.00010.4.
30. Chistyakov D, Chistyakov P. The polymorphism of drugs: new approaches to the synthesis of nanostructured polymorphs and cocrystals. Pharmaceutics. 2020;12(1):34. doi:10.3390/pharmaceutics12010034.
31. U.S. Food and Drug Administration. Inactive ingredients in approved drug products search: frequently asked questions [Internet]. Silver Spring (MD): U.S. Food and Drug

Administration; 2022 Jan 26 [cited 2026 Jun 20].
32. Pharma Excipients. Pharma Excipients: all about excipients & inactive ingredients [Internet]. Steinhausen (Switzerland): Pharma Excipients; [cited 2026 Jun 20].
33. U.S. Food and Drug Administration. Inactive Ingredients Database download [Internet]. Silver Spring (MD): U.S. Food and Drug Administration; 2026 Apr 13 [cited 2026 Jun 20].
34. Pharmuni. In-process controls (IPC) in drug manufacturing [Internet]. Pharmuni; 2026 [cited 2026 Jun 20].
35. Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. Drug Discov Today. 2019;24(3):773-780. doi:10.1016/j.drudis.2018.11.014.
36. U.S. Food and Drug Administration. Overview of in vivo bioavailability (BA) and bioequivalence (BE) studies supporting NDAs and ANDAs [Internet]. Silver Spring (MD): U.S. Food and Drug Administration; 2022 Sep 14 [cited 2026 Jun 20].
37. Bannigan P, Bao Z, Hickman RJ, Aldeghi M, Häse F, Aspuru-Guzik A, et al. Machine learning models to accelerate the design of polymeric long-acting injectables. Nat Commun. 2023;14(1):35. doi:10.1038/s41467-022-35343-w.
38. Dave VS, Haware RV, Sangave NA, Sayles M, Popielarczyk M. Drug-excipient compatibility studies in formulation development: current trends and techniques. AAPS Formulation Design and Development (FDD) Section Newsletter. 2015 Jan:9-15.
39. Wang N, Sun H, Dong J, Ouyang D. PharmDE: a new expert system for drug-excipient compatibility evaluation. Int J Pharm. 2021;607:120962. doi:10.1016/j.ijpharm.2021.120962.
40. Patel S, Patel M, Kulkarni M, Patel MS. DE-INTERACT: A machine-learning-based predictive tool for the drug-excipient interaction study during product development—Validation through paracetamol and vanillin as a case study. Int J Pharm. 2023;637:122839. doi:10.1016/j.ijpharm.2023.122839.
41. Jubilant Biosys. Solid state capabilities [Internet]. Noida (India): Jubilant Biosys Ltd.; [cited 2026 Jun 20].
42. Yu LX, Amidon G, Khan MA, Hoag SW, Polli J, Raju GK, et al. Understanding pharmaceutical quality by design. Org Process Res Dev. 2014;18(4):556-564. doi:10.1021/op300241f.
43. Bevill M, Seadeek C, Setiawan N, Comella S, Mibeck B, Boerrigter S. A comprehensive approach for solid form selection in preclinical development and beyond [Internet]. Lafayette (IN): Triclinic Labs; 2023 Nov [cited 2026 Jun 20].
44. Simulations Plus. Physicochemical & biopharmaceutical (PCB) module: ADMET Predictor® [Internet]. Lancaster (CA): Simulations Plus Inc.; [cited 2026 Jun 20].
45. ADMETlab. ADMETlab 2.0: systematic ADMET evaluation platform [Internet]. Changsha (China): Computational Biology & Drug Design Group, Central South University; [cited 2026 Jun 20].
46. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 2021;49(W1):W5-W14. doi:10.1093/nar/gkab255.
47. Venkataraman M, Vijayakumar S, Alshammari TM, Alruwaili NK, Alhazmi HA. Leveraging machine learning models in evaluating ADMET properties for drug discovery and development. ADMET DMPK.

2025;13(3):2772. doi:10.5599/admet.2772. PMID:40585410. PMCID: PMC12205928.
48. NETZSCH Analyzing& Testing. API characterization [Internet]. Selb (Germany): NETZSCH-Gerätebau GmbH; [cited 2026 Jun 20].
49. Patel P. Pharmaceutical solid state studies [Internet]. Bioprocess Online; 2013 [cited 2026 Jun 20].
50. Karpinski PH. Polymorphism of active pharmaceutical ingredients. Chem Eng Technol. 2006;29(2):233-237. doi:10.1002/ceat.200500397.
51. Hang NT, Thanh HT, Hai NT, Linh NTH, Dung PT, Van Quy N, et al. Machine learning approaches to predict drug-excipient compatibility using molecular descriptors and molecular fingerprints. ComputBiol Med. 2024;176:108489. doi:10.1016/j.compbiomed.2024.108489.
52. Minitab. AI: the future of design of experiments [Internet]. State College (PA): Minitab LLC; 2024 [cited 2026 Jun 20].
53. Xu X, Li D, Bi J, Moeckel M. AutoML based workflow for design of experiments (DOE) selection and benchmarking data acquisition strategies with simulation models. Sci Rep. 2024;14:32170. doi:10.1038/s41598-024-83581-3.
54. Arboretti R, Ceccato R, Pegoraro M, Salmaso L. Machine learning and design of experiments with an application in process modeling and optimization for pharmaceutical and industrial manufacturing. Qual Eng. 2022;34(3):403-421. doi:10.1080/08982112.2021.1959336.
55. Yu LX, Amidon G, Khan MA, Hoag SW, Polli J, Raju GK, et al. Understanding pharmaceutical quality by design. AAPS J. 2014;16(4):771-783. doi:10.1208/s12248-014-9598-3. PMCID: PMC4070262. PMID: 24854893.
56. European Medicines Agency. Quality by design [Internet]. Amsterdam (Netherlands): European Medicines Agency; [cited 2026 Jun 20].
57. Synthace. Quality by Design (QbD): a guide for pharmas [Internet]. London (UK): Synthace Ltd.; 2023 May 22 [cited 2026 Jun 20].
58. Kumar S, Gupta SK, Sharma PK. Drug carrier systems for solubility enhancement of BCS class II drugs: a critical review. Crit Rev Ther Drug Carrier Syst. 2013;30(3):217-256. doi:10.1615/CritRevTherDrugCarrierSyst.2013007232.
59. Chowdary KPR, Kumar BV, Rao YS, et al. Enhancing the solubility of BCS class II and IV drugs by SEDDS approach: a structured review. J Pharm Clin Res. 2024;7(2):1-14.
60. Premakumar SV, Vishwanath BA, Swapna D. Enhancement of solubility and dissolution rate of BCS class II drugs. Int J Pharm Sci Rev Res. 2024;84(7):153-161. doi:10.47583/ijpsrr.2024.v84i07.023.
61. Hens B, Corsetti M, Bermejo M, Löbenberg R, González PM, Mitra A, et al. "Development of fixed dose combination products" workshop report: considerations of gastrointestinal physiology and overall development strategy. AAPS J. 2019;21(4):75. doi:10.1208/s12248-019-0346-6.
62. Central Drugs Standard Control Organization. Fixed dose combinations (FDC) [Internet]. New Delhi: Directorate General of Health Services, Ministry of Health and Family Welfare, Government of India; [cited 2026 Jun 20].
63. Faure A, York P, Rowe RC. Process control and scale-up of pharmaceutical wet granulation processes: a review. Eur J Pharm Biopharm. 2001;52(3):269-277. doi:10.1016/S0939-6411(01)00184-9.

64. Jang EH, Mansouri S, Linkes M, Shah NH, Venkatasubramanian V. Model-based scale-up methodologies for pharmaceutical granulation processes: a review. Pharmaceutics. 2020;12(6):517. doi:10.3390/pharmaceutics12060517.
65. Faure A, York P, Rowe RC. Process control and scale-up of pharmaceutical wet granulation processes: a review. Eur J Pharm Biopharm. 2001;52(3):269-277. doi:10.1016/S0939-6411(01)00184-9.
66. Tnokovska K, Toshev K, TrifunovskaVulovska B, Antovska P, Ugarkovic S. Optimization of tablet compressing process using experimental design approach. Maced Pharm Bull. 2020;66(Suppl 1):133-134. doi:10.33320/maced.pharm.bull.2020.66.03.066.
67. Paul S, Sun CC. Dependence of friability on tablet mechanical properties and a predictive approach for binary mixtures. Pharm Res. 2017;34(12):2901-2909. doi:10.1007/s11095-017-2273-5.
68. Moin A, Alghamdi A, Sindi IA, Alharbi KS, Alhazmi HA. Development of machine learning models for estimation of tablet disintegration time using formulation composition and physical properties. AAPS Open. 2025;11:46. doi:10.1186/s41120-025-00146-8.
69. Hlinak AJ, Kuriyan K, Morris KR, Reklaitis GV, Basu PK. Understanding critical material properties for solid dosage form design. Int J Pharm. 2012;423(2):351-363. doi:10.1016/j.ijpharm.2012.01.017.
70. Asian Journal of Chemistry. About the journal [Internet]. New Delhi: Asian Journal of Chemistry; [cited 2026 Jun 20].
71. Chaudhary A, Sharma R, Patel M, et al. Polymers used in enteric coating: an overview. Int J Pharm Res Appl. 2025;10(2).
72. Sartorius. Real-time process monitoring software [Internet]. Göttingen (Germany): Sartorius AG; [cited 2026 Jun 20].
73. METTLER TOLEDO. Process analytical technology (PAT) for pharmaceutical continuous manufacturing [Internet]. Columbus (OH): METTLER TOLEDO; [cited 2026 Jun 20].
74. Abatzoglou N. Process analytical technology for blending. In: Levin M, editor. Pharmaceutical Blending and Mixing. Hoboken (NJ): John Wiley & Sons; 2015. p. 401-430. doi:10.1002/9781118682692.ch15.
75. Tekie HK, Beshah T, Gernaey KV, Gebrehiwot NH, Gidey K, Gebremeskel HG, et al. Machine learning model for the prediction of blend uniformity powder blending process [Internet]. Kongens Lyngby: Technical University of Denmark; 2026 [cited 2026 Jun 20].
76. Tekie HK, Beshah T, Gernaey KV, Gebrehiwot NH, Gidey K, Gebremeskel HG, et al. Machine learning model for the prediction of blend uniformity powder blending process [Internet]. Kongens Lyngby: Technical University of Denmark; 2026 [cited 2026 Jun 20].
77. Su Q, Hermant P, Casati F, Halkude BS, Wu W, Ramnath A, et al. Model predictive *in vitro* dissolution testing continuous manufacturing: an equivalence study. AIChE J. 2023;69(9):e18124. doi:10.1002/aic.18124.
78. Galata DL, Kleinebudde P, Wiedey R, Dressman JB. Fast, spectroscopy-based prediction of *in vitro* dissolution for oral extended-release tablets using process analytical technology and machine learning. Int J Pharm. 2019;572:118732. doi:10.1016/j.ijpharm.2019.118732.
79. Sacher S, Fink E, Alva C, Afonso Urich JA, Doğan A, Herndler V, et al. Real-time prediction of dissolution profiles of coated

oral dosage forms. Int J Pharm. 2025;672:125336. doi:10.1016/j.ijpharm.2024.125336. https://www.sciencedirect.com/science/article/pii/S0378517324010755

80. Nasar FF, Abu Bakar NF, Mahmud J. Prediction of *in vitro* dissolution profile for oral dispersible film (ODF) using artificial neural network. In: 2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET). Piscataway (NJ): IEEE; 2024. doi:10.1109/IICAIET63584.2024.10872643.
81. Jacob S, Nair AB. An updated overview with simple and practical approach for developing *in vitro–in vivo* correlation. Drug Dev Res. 2018;79(3):97-110. doi:10.1002/ddr.21435.
82. Gillespie WR. Convolution-based approaches for *in vivo–in vitro* correlation modeling. Adv Drug Deliv Rev. 1997;23(1-3):99-109. doi:10.1016/S0169-409X(96)00405-8.
83. Nepolean R, Chandramohan K, Rajkumar K, Anbarasu D, Sivasankari S. A comprehensive review of modified release dosage forms and its formulation approaches, in-vitro characterizations, and pharmaceutical applications. PriMera Sci Surg Res Pract. 2026;7(3):70-80. doi:10.56831/PSSRP-07-252.
84. Shah V, Khambhla E, Nivsarkar M, Trivedi R, Patel RK. An integrative QbD approach for the development and optimization of controlled release compressed coated formulation of water-soluble drugs. AAPS PharmSciTech. 2022;23(5):120. doi:10.1208/s12249-022-02225-9.
85. Singh B, Chakkal SK, Ahuja N. Formulation and optimization of controlled release mucoadhesive tablets of atenolol using response surface methodology. AAPS PharmSciTech. 2006;7(1):E1-E10. doi:10.1208/pt070103.
86. Roggo Y, Martinez L, Peinado A, Matero S. Near infrared spectroscopy for blend uniformity monitoring: an innovative qualitative application based on the coefficient of determination. J Near Infrared Spectrosc. 2022;30(6):322-336. doi:10.1177/09670335221130430.
87. Popo M, Wu H, Khan MA. Blend uniformity analysis using stream sampling and near infrared spectroscopy. Int J Pharm. 2003;262(1-2):35-43. doi:10.1016/S0378-5173(03)00341-0.
88. Griffen JA, Owen H, Matousek P. Comprehensive quantification of tablets with multiple active pharmaceutical ingredients using transmission Raman spectroscopy. J Pharm Biomed Anal. 2015;115:277–282. doi:10.1016/j.jpba.2015.07.019.
89. Li J, Gee S, Wigman L, Davis J, Zhang W. Determination of API content in tablets by transmission Raman spectroscopy. *American Pharmaceutical Review* [Internet]. 2019 Mar 28 [cited 2026 Jun 20].
90. Diószegi A, Ficzere M, Mészáros LA, Péterfi O, Farkas A, Galata DL, et al. Automated tablet defect detection and the prediction of disintegration time and crushing strength with deep learning based on tablet surface images. Int J Pharm. 2024;667:124896. doi:10.1016/j.ijpharm.2024.124896.
91. Karthikeyan Pillai AK, Ellis A, Rossato HH, Fallon E. Real-time tablet inspection using computer vision for blister packing machines. In: Alja'Am JM, Alja'Am JM, Elseoud SA, Karam O, editors. *Proceedings of the 2023 5th International Conference on Computer and Applications (ICCA)*; 2023 Nov 28–30; Cairo, Egypt. Piscataway (NJ): IEEE; 2023. doi:10.1109/ICCA59364.2023.10401685.
92. U.S. Food and Drug Administration. *PATA Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality*

Assurance [Internet]. Silver Spring (MD): FDA; 2004 [updated 2018 Aug 24; cited 2026 Jun 20].

93. Jha A, Kumar A, Singh R, Patel S. Artificial intelligence manufacturing: applications, challenges and future perspectives. J Res Tech Dev Des. 2024;7(5):1044–1052.
94. Lennard A, Zimmermann B, Clenet D, Molony M, Tami C, Oliva Aviles C, Moran A, Pue-Gilchrist P, Flores E. Stability modeling methodologies to enable earlier patient access. J Pharm Sci. 2024;113(12):3406-3412. doi:10.1016/j.xphs.2024.09.018.
95. Srokosz S. Making accelerated predictive stability studies even faster with software. ACD/Labs Blog [Internet]. 2025 Feb 27 [cited 2026 Jun 20].
96. Parenty ADC, Zaben N, Williams AC, Hardy IJ, Maheswaran R. An expert system to predict the forced degradation of organic molecules. Mol Pharm. 2013;10(8):2962–2974. doi:10.1021/mp400083h.
97. Wyttenbach N, Janas D, Siam M, Laffan A, Ilott AJ, Roberts CJ, et al. Predicting solid-state stability using computational and experimental approaches. Mol Pharm. 2014;11(12):4288–4299. doi:10.1021/mp5004467.
98. Srokosz S. Making accelerated predictive stability studies even faster with software. ACD/Labs Blog [Internet]. 2025 Feb 27 [cited 2026 Jun 20].
99. Hou M, Ji Z, Jia Z, Shi C, Yang X, Ying X, Zhao Z. Artificial neural network-based shelf life prediction in the food storage process: A review. Crit Rev Food Sci Nutr. 2024;64(35):12009–12024. doi:10.1080/10408398.2023.2245899.
100. Li Z, Nash WT, O'Brien SP, Qiu Y, Gupta RK, Birbilis N, et al. *cardiGAN: A generative adversarial network model for design and discovery of multi principal element alloys* [Preprint]. arXiv; 2022. doi:10.48550/arXiv.2202.00966.
101. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems 27 (NeurIPS 2014)*. Red Hook (NY): Curran Associates, Inc.; 2014. p. 2672–2680.
102. Zhao Y, Wang H, Li X, Chen J, Zhang Y, Liu Z, et al. Artificial intelligence-driven pharmaceutical development: Current applications and future perspectives. Smart Mater Med. 2025;6:100194. doi:10.1016/j.smaim.2025.100194.
103. Okubena A, Mohammed YA, Elbadawi M. FormuLLA: A large language model approach to generating novel 3D printable formulations [Preprint]. arXiv; 2026. doi:10.48550/arXiv.2601.02071.
104. *Advances in Neural Information Processing Systems (NeurIPS 2025)*. San Diego (CA): Neural Information Processing Systems Foundation; 2025.
105. Tom G, Granda JM, Aspuru-Guzik A, Burke MD, Coley CW, Cronin L, et al. Self-driving laboratories for chemistry and materials science. Chem Rev. 2024;124(16):9633–9732. doi:10.1021/acs.chemrev.4c00055.
106. U.S. Food and Drug Administration. Artificial Intelligence and Machine Learning (AI/ML) in Drug Development: Guidance for Industry. Silver Spring (MD): FDA; 2025. FDA Guidance
107. Sartorius. MODDE® – Design of experiments (DOE) software [Internet]. Göttingen (Germany): Sartorius AG; [cited 2026 Jun 20].
108. JMP Statistical Discovery LLC. JMP statistical discovery software [Internet]. Cary (NC): JMP Statistical Discovery LLC; [cited 2026 Jun 20].

109. Umetrics Automation. Umetrics releases SIMCA® 13 software. Chemical Processing [Internet]. 2012 Mar 6 [cited 2026 Jun 20].
110. Thinkitive Technologies Pvt. Ltd. 8 popular machine learning frameworks to use in Python [Internet]. Pune: Thinkitive Technologies Pvt. Ltd.; 2023 Sep 21 [cited 2026 Jun 20].
111. Kanagavell M. Exploring TensorFlow, PyTorch, scikit-learn, and Keras with Python: real-life examples for machine learning and deep learning tasks [Internet]. Medium. 2023 Oct 15 [cited 2026 Jun 20].
112. Pharma Excipients. Manufacture of a triple fixed-dose combination tablet [Internet]. Pharma Excipients; 2023 [cited 2026 Jun 20].
113. Dataiku. Driving AI innovation in life sciences with AI and Dataiku [Internet]. New York (NY): Dataiku; [cited 2026 Jun 20].
114. ADDITIVE Soft- und Hardware für Technik und Wissenschaft GmbH. Signals Notebook: integrated electronic laboratory notebook (ELN) for modern laboratories [Internet]. Friedrichsdorf (DE): ADDITIVE Soft- und Hardware für Technik und Wissenschaft GmbH; [cited 2026 Jun 20].
115. Revvity Signals Software Inc. Spotfire® as essential tool at major global pharmaceutical company [Internet]. Clinical Leader. [cited 2026 Jun 20].
116. Holland I, Davies JA. Automation in the life science research laboratory. Front BioengBiotechnol. 2020;8:571777. doi:10.3389/fbioe.2020.571777.
117. Van Noorden R. Lab automation. Chem World [Internet]. 2016 Dec 20 [cited 2026 Jun 20].
118. Geoui T. Lab automation in pharma: turn complexity into progress [Internet]. Zühlke Group. 2025 Jun 27 [cited 2026 Jun 20].
119. Ansys. Digital twins in manufacturing: transforming production with virtual models [Internet]. Canonsburg (PA): Ansys; [cited 2026 Jun 20].
120. Camarella S, Conway MP, Goering K, Huntington M. Digital twins: the next frontier of factory optimization [Internet]. McKinsey & Company. 2024 Jan 10 [cited 2026 Jun 20].
121. IntuitionLabs. Pharma AI pilots: fixing data foundations for scale [Internet]. IntuitionLabs; 2026 [cited 2026 Jun 20].
122. Apheris GmbH. AISB Network expands federated OpenFold3 initiative with three new pharma contributors [Internet]. Berlin (Germany): Apheris GmbH; 2025 Oct 1 [cited 2026 Jun 20].
123. Torontech. Tablet hardness: how to measure, calculate & control it [Internet]. Torontech; 2025 Dec 12 [cited 2026 Jun 20].
124. Soft chew pharmaceutical formulations. US Patent No. 11,633,361 [Internet]. United States Patent and Trademark Office; [cited 2026 Jun 20].
125. Comparison of deep transfer learning against contrastive learning in industrial quality applications for heavily unbalanced data scenarios when data augmentation is limited [Internet]. PMC; 2025 [cited 2026 Jun 20].
126. A comprehensive survey on imbalanced data learning [Internet]. arXiv; 2025 Sep 12 [cited 2026 Jun 20].
127. Explainable AI methods for drug discovery: a survey of interpretability, metrics and mechanistic insight. ScienceDirect [Internet]. 2026 Mar 9 [cited 2026 Jun 20].
128. Lenarczyk G, Minssen T, Price N, Rai A. The future of AI regulation in drug development: a comparative analysis. J Law Biosci. 2025;12(2):lsaf028. doi:10.1093/jlb/lsaf028.
129. Aligned Machinery. From lab to commercial production: pharmaceutical process scale-up and technology transfer [Internet]. Aligned Machinery; 2025 Dec 24 [cited 2026 Jun 20].

130. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. Artificial intelligence for drug development [Internet]. Silver Spring (MD): FDA; 2026 [cited 2026 Jun 21].
131. FDA and EMA align on ten principles to guide artificial intelligence use in drug development. Appl Clin Trials [Internet]. 2026 Jan 28 [cited 2026 Jun 21].
132. IntuitionLabs. AI post-market surveillance: locked vs. continuous learning [Internet]. IntuitionLabs; 2026 May 10 [cited 2026 Jun 21].
133. Risk-based validation of software, automation and artificial intelligence in pharmaceuticals. BiosciBiotechnol Res Asia [Internet]. 2025 Dec 30 [cited 2026 Jun 21].
134. IntuitionLabs. Audit trails: 21 CFR Part 11 & Annex 11 compliance [Internet]. IntuitionLabs; 2026 Feb 10 [cited 2026 Jun 21].
135. European Commission. EudraLex Volume 4: Good Manufacturing Practice Guidelines—New Annex 22: Artificial Intelligence. Brussels: European Commission; 2025. Official EU Annex 22 consultation page
136. Diversity in clinical trials: why it matters and how to achieve it [Internet]. Kapsule Technologies; 2026 Feb 13 [cited 2026 Jun 21].
137. Algorithmic accountability: liability frameworks for AI-driven clinical decisions [Internet]. Censinet Inc.; 2026 Apr 9 [cited 2026 Jun 21].
138. AI and drug development decisions: a framework for accountability [Internet]. Santa Clara (CA): Markkula Center for Applied Ethics, Santa Clara University; 2026 [cited 2026 Jun 21].
139. Klover.ai. Pfizer's AI strategy: analysis of dominance in pharma [Internet]. Klover.ai; 2025 Jul 22 [cited 2026 Jun 21].
140. Machine learning predicts drug release profiles and kinetic parameters based on tablets' formulations. AAPS J [Internet]. 2025 Jul 28 [cited 2026 Jun 21].
141. Pathak KA, Kafle P, Vikram A. Deep learning-based defect detection in film-coated tablets using a convolutional neural network. Int J Pharm [Internet]. 2025 [cited 2026 Jun 21].
142. IntuitionLabs. Computer vision quality control: enhancing drug manufacturing [Internet]. IntuitionLabs; 2026 May 8 [cited 2026 Jun 21].
143. Alvarado-Maldonado D, Johnston B, Brown CJ. Natural language processing (NLP) tools for pharmaceutical manufacturing information extraction from patents [Internet]. arXiv; 2025 [cited 2026 Jun 21].
144. AI/ML advancements in drug formulation: the potential of self-driving labs [Internet]. Oxford Global; 2024 Oct 16 [cited 2026 Jun 21].
145. Efficient discovery of new medicine formulations using a semi-self-driven robotic formulator. Digit Discov [Internet]. 2025 Jul 17 [cited 2026 Jun 21].
146. Artificial intelligence and multi-omics in pharmacogenomics: a new era of precision medicine. Mayo Clin Proc Digit Health [Internet]. 2025 Jun 26 [cited 2026 Jun 21].
147. Maharjan R, Kim NA, Kim KH, Jeong SH. Transformative roles of digital twins from drug discovery to continuous manufacturing: pharmaceutical and biopharmaceutical perspectives. Int J Pharm X [Internet]. 2025 Sep 27 [cited 2026 Jun 21].
148. A modern strategy for digital real-time release testing in continuous tablet manufacturing. Eur J Pharm Biopharm [Internet]. 2025 Mar 24 [cited 2026 Jun 21].
149. Optimization of print parameters for batch and continuous manufacturing of three-

dimensional (3D) printed dosage forms using artificial intelligence and machine learning. Drug Deliv Transl Res [Internet]. 2025 Nov 4 [cited 2026 Jun 21].

150. 3D-printed polypills for personalized medicine and precision oral drug delivery practice: a review [Internet]. PMC; 2026 [cited 2026 Jun 21].

HOW TO CITE: Pratik Patel, Preeti Sah, Zahid Husain, Shivanshu Dwivedi, Rajveer Singh Chauhan, Nimita Manocha, Artifical Intelliegence And Machine Learning in Solid Dosage Formulation, Int. J. of Pharm. Sci., 2026, Vol 4, Issue 9, 3920-3951, https://doi.org/10.5281/zenodo.23038296