CODEN [USA]: IAJPBB ISSN : 2349-7750

INDO AMERICAN JOURNAL OF
PHARMACEUTICAL SCIENCES
SJIF Impact Factor: 7.187
https://doi.org/10.5281/zenodo.24316932

Available online at: http://www.iajps.com Review Article

ARTIFICIAL INTELLIGENCE IN PHARMACEUTICAL FORMULATION DEVELOPMENT

V.Vasu Naik*, Sk. Afrin
Department of Pharmaceutics, Hindu College of Pharmacy, Amaravathi Road, Guntur, Andhrapradesh, India- 522002

Abstract:
Artificial Intelligence (AI) is emerging as a transformative technology in pharmaceutical formulation development by improving the efficiency, accuracy, and speed of drug development processes. Traditional formulation development involves time-consuming experiments, extensive laboratory investigations, and trial-and-error approaches to identify suitable ingredients and optimize formulation parameters. AI-based techniques, including machine learning, deep learning, artificial neural networks, and predictive modelling, offer innovative solutions to overcome these challenges. These technologies can analyze large datasets, predict drug–excipient interactions, optimize formulation compositions, and evaluate critical quality attributes of pharmaceutical products. AI also plays an important role in selecting appropriate excipients, predicting drug solubility, dissolution profiles, stability, and bioavailability, thereby supporting the development of safe, effective, and high-quality drug formulations. Furthermore, the integration of AI with Quality by Design (QbD), process analytical technology (PAT), and continuous manufacturing can enhance process optimization and reduce development costs. AI-driven approaches have potential applications in the development of tablets, capsules, nanoparticles, liposomes, controlled-release systems, and other advanced drug delivery systems. Despite its significant advantages, challenges such as data quality, model validation, regulatory acceptance, and implementation costs need to be addressed. Overall, artificial intelligence represents a promising approach to modernizing pharmaceutical formulation development by enabling faster decision-making, reducing experimental workload, and supporting the development of innovative and personalized drug delivery systems.
Keywords: *Artificial Intelligence, Pharmaceutical Formulation, Machine Learning, Deep Learning, Drug Delivery Systems, Quality by Design, Formulation Optimization.*

Corresponding author:
V. Vasu Naik*,
Professor & HOD, Department of Pharmaceutics,
Hindu College of Pharmacy, Amaravathi Road,
Guntur, Andhrapradesh, India- 522002.
E-mail: *drvasupharma@gmail.com*

QR CODE

Please cite this article in press ***V. Vasu Naik*** *et al.,* ***Artificial Intelligence In Pharmaceutical Formulation Development.,****. Indo Am. J. P. Sci, 2026; 13(11).*

1. INTRODUCTION:

Pharmaceutical formulation development is an essential bridge between an active pharmaceutical ingredient (API) and a finished medicine that is stable, bioavailable, and acceptable to patients. This process has traditionally relied on trial-and-error: excipient compatibility screening followed by one-factor-at-a-time or designed experiments, with prototypes built and tested across successive iterations. Even with rigorous experimental planning, formulation teams typically run dozens, sometimes hundreds, of experiments to define an adequate design space, consuming considerable time, cost, and often scarce early-stage API [16,26].

AI-algorithms trained to recognize patterns and support decision-making-offers a promising alternative. Using machine learning (ML) and deep learning (DL), a model trained on historical experimental data can predict which formulations are likely to succeed, or even propose entirely new candidates for laboratory testing, allowing resources to be directed where they matter most and reducing reliance on guesswork. It is therefore unsurprising that the pharmaceutical industry invested roughly $190 billion in R&D in 2024, with an increasing share directed toward AI-enabled discovery and development [26].

A substantial body of research has already addressed AI in drug discovery and early-stage hit identification; this review focuses instead on its role downstream, in translating discoveries into deliverable products-how formulations are designed, drug products are shaped, and manufacturing is scaled. Although interest in AI for formulation has grown rapidly between 2023 and 2026, the field has not yet received the same degree of critical, systematic review as drug discovery [13,15].

This review pursues three principal aims: first, to characterize "how" these AI/ML models actually solve concrete formulation problems; second, to critically appraise the quality and practical usefulness of the evidence assembled so far; and third, to identify specific, actionable research gaps and outline a roadmap for future work, particularly as regulatory science seeks to keep pace with the technology.

2. Background

2.1 From trial-and-error to data-driven design

Pharmaceutical development began with simple empirical experimentation and progressed gradually toward more sophisticated statistical methods, including factorial design and response surface modeling. Quality by Design (QbD), introduced in the early 2000s, formalized this progression by linking formulation goals-expressed as Quality Target Product Profiles (QTPPs) and Critical Quality Attributes (CQAs)-to process variables such as Critical Material Attributes (CMAs) and Critical Process Parameters (CPPs), all documented within a defined design space.

The limitation is that QbD's statistical backbone—typically multiple regression or response surface modeling—performs adequately with a small number of variables but degrades quickly as formulations involve multiple excipients or multi-step processes. Real-world formulation behavior is nonlinear and often unpredictable: minor changes can produce disproportionate effects, and many interactions remain invisible to simpler models [12,30].

2.2 Terminology and scope

In this review, "machine learning" is used as an umbrella term for any algorithm that learns patterns from data, including random forests, support vector machines, and boosting methods. "Deep learning" refers specifically to large, multi-layer neural networks capable of identifying complex patterns without explicit feature engineering. "Generative AI" denotes models that actively propose new formulation or molecular candidates rather than merely predicting outcomes [4,18]. A "digital twin" is defined as a continuously updated virtual representation of a process or product, linked directly to real-time manufacturing data [24].

2.3 Regulatory landscape

Regulatory bodies worldwide, particularly the FDA and EMA, have devoted substantially greater attention to AI in pharmaceuticals since 2023. FDA submissions containing AI or ML components rose from a single instance in 2016 to more than 500 by late 2024, prompting new discussion papers and draft guidance in 2023 and 2025, alongside efforts to develop "New Approach Methodologies" that use AI for more human-relevant testing. The EMA has similarly advanced a multi-year workplan, issuing qualification opinions for its first AI-based methodology in March 2025. For developers and evaluators of AI in pharmaceutical formulation, model performance alone is therefore insufficient; the model must also withstand scrutiny regarding validation, transparency, and real-world applicability

2.4 Why formulation science is a special case for AI application domain

Formulation development presents a particularly demanding application domain for AI. Unlike image recognition, where data are abundant and relatively uniform, formulation data are heterogeneous, combining discrete variables (excipient identity), continuous variables (quantity), process parameters, and complex nonlinear interactions among them. Outputs are

similarly multidimensional: formulators must simultaneously balance mechanical strength, disintegration rate, and release behavior, and improving one attribute frequently compromises another.

Compounding this difficulty, formulation datasets are typically small: a standard formulation program may comprise only 50 to 100 experiments, several orders of magnitude below the datasets common in other machine learning domains, rendering highly complex architectures poorly suited to the task [1,2,4]. What is required instead are methods capable of extracting reliable signal from small, noisy datasets—which is why ensemble tree models and compact neural networks frequently outperform deep learning in this field, and why data efficiency currently matters more than maximizing benchmark accuracy.

3. Mechanisms: How AI "Learns" Formulation–Performance Relationships

Understanding why researchers select a particular algorithm for a given problem requires considering both the scientific question at hand and the structure of the underlying data.

3.1 Supervised learning on table-based data

The majority of published work relies on supervised learning, in which a model is trained on tabular inputs (excipient identity and quantity, process settings such as compression force or mixing time) paired with known outcomes (hardness, dissolution at a given time point). Decision trees, random forests, boosted-tree methods such as XGBoost, support vector machines, and k-nearest neighbors are all established standards within the field

Momeni et al. compared several of these algorithms for predicting tablet disintegration time and hardness from raw-material properties, finding that tree-based models generally performed best, reflecting their robustness on small, noisy datasets [1]. Similarly, Bharathi et al. [2] used an AutoML pipeline to compare these methods, along with a deep neural network, for sustained-release tablets; the random forest again performed best, although fine-tuning narrowed the gap with the deep network.

3.2 Artificial neural networks and deep learning

Artificial neural networks, composed of stacked layers of interconnected nodes, can learn complex patterns given sufficient data and appropriate tuning. In tablet formulation, they are particularly useful for modeling multiple outcomes simultaneously, since formulators are typically concerned with several properties at once rather than a single endpoint.

Deeper networks extend this capability further, discovering latent structure without explicit guidance, but they require larger datasets and careful tuning; otherwise they risk memorizing the training set rather than learning generalizable relationships [2,12].

Convolutional neural networks (CNNs), originally developed for image analysis, are now used to analyze particle micrographs, extracting texture and morphological features relevant to release kinetics or mechanical strength without manual feature engineering. Graph neural networks (GNNs) are particularly promising for this field, as they model molecular or excipient networks directly, capturing not only which components are present but how they interact at the atomic or network level—an ability that traditional tabular models lack [4].

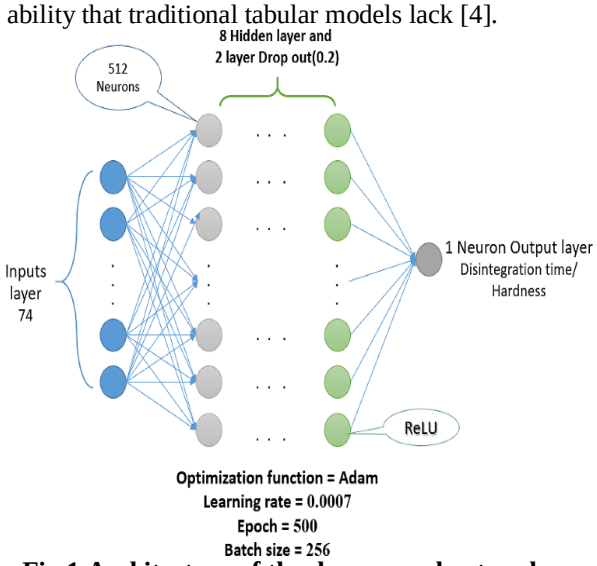

Fig.1.Architecture of the deep neural network

3.3 Bayesian optimization and active/reinforcement learning

Bayesian optimization is fundamentally an efficiency-driven approach. Rather than simply predicting outcomes, it searches for the optimal formulation by asking which experiment would yield the most informative result. By modeling both existing knowledge and remaining uncertainty, it can propose experiments that either refine current leading candidates or reduce knowledge gaps.

In protein formulation, for example, this approach has enabled teams to identify high-stability candidates, such as those meeting a target melting point, using substantially fewer experiments than conventional brute-force designs. Closed-loop design–test–learn cycles accelerate development further, particularly for high-value products where experimental material is limited [20,21].

Reinforcement learning (RL) extends this concept further: an agent learns through trial and error, improving its decision-making by receiving rewards or penalties based on outcomes. RL has been applied primarily to process control, such as

adjusting spray-drying parameters, but has yet to gain comparable traction in core formulation design.

4. Recent Advances

One of the most promising trends is the integration of physics-based, first-principle models with data-driven ML. Rather than requiring a deep network to learn every relationship from scratch, the mechanistic model handles well-understood phenomena, such as mass balance or thermodynamics, while ML addresses the residual complexity where behavior is less predictable.

Because they are grounded in established scientific principles from the outset, hybrid models typically require less data than purely data-driven approaches. Digital twins—real-time virtual representations of a process or product—extend this principle to the plant floor, continuously integrating new data and model updates to track, predict, and in some cases control manufacturing, closing the loop between experimentation and production.

4.1 Oral solid dosage forms

Oral solid dosage (OSD) forms continue to dominate pharmaceutical machine learning research, largely because they are produced at scale and generate structured, usable data from historical batch records. Momeni and colleagues addressed one of the field's persistent obstacles, limited and inaccessible datasets, by publishing an open dataset derived from pre-formulation tests for fast-disintegrating tablets. Having trained several machine learning models and found that ensemble learners predicted both disintegration time and hardness well, they made their data publicly available—a practice this review argues should become standard rather than exceptional.

Bharathi and co-workers [2] extended this approach to sustained-release matrix tablets, using an AutoML pipeline to compare six machine learning algorithms against a deep neural network and predicting dissolution time directly from raw-material properties, thereby bypassing the need for new experiments for each formulation. Other researchers have applied optimization strategies such as Grey Wolf Optimization to fine-tune sparse Bayesian learning models, achieving strong predictive accuracy for disintegration time and frequently outperforming competing approaches—underscoring that hyperparameter tuning, not algorithm choice alone, plays a critical role in predictive performance.Deng and colleagues showed that ML models can discern how critical factors such as polymer type, solvent, and processing conditions affect the characteristics of PLGA microspheres, including particle size, encapsulation efficiency, and release kinetics, thus significantly cutting down the experimental workload[12].

Balancing accuracy with interpretability is not merely a technical preference but a necessity, as regulators increasingly demand models capable of explaining their own decisions; this review holds that such approaches should define the agenda for the next generation of AI-driven formulation research.

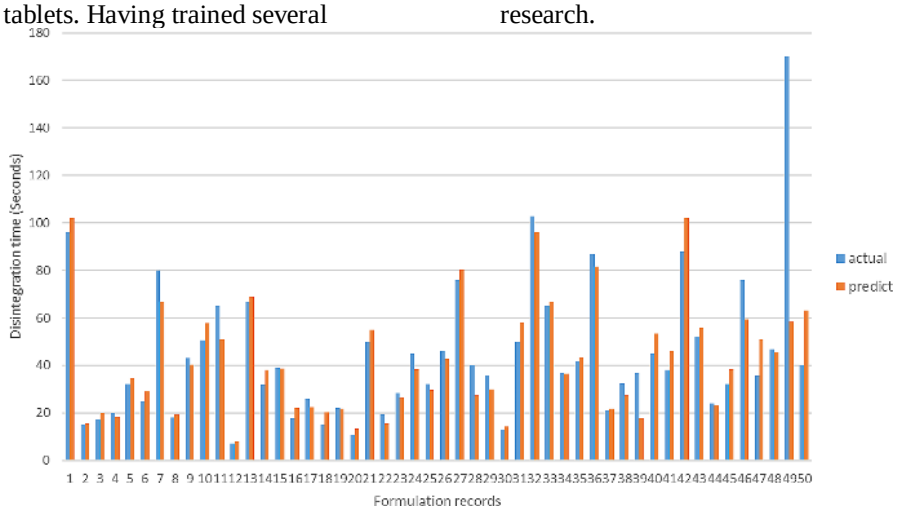

Fig.2.Predicted disintegration time values vs actual values - deep learning method

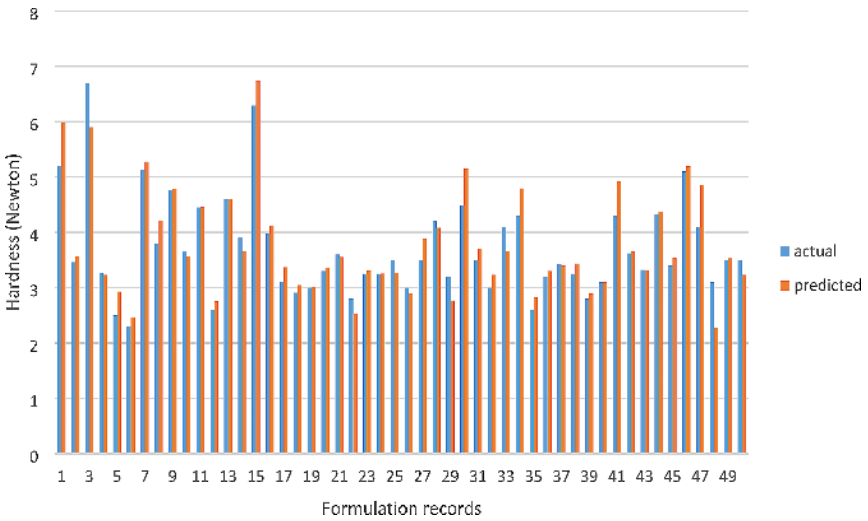

Fig.3.Predicted hardness values vs. actual values - deep learning method

4.2 Nanoparticulate and lipid-based drug delivery systems

Nanoparticulate drug delivery systems, such as polymeric nanoparticles, liposomes, solid lipid nanoparticles, nanostructured lipid carriers, and lipid nanoparticles (LNPs), represent the most dynamic field for AI application in formulation science. This surge is largely attributed to the practical success of mRNA-LNP vaccines. A comprehensive review set for 2025/2026 highlighted the role of machine learning in this domain, revealing that random-forest and support-vector-machine models have enhanced screening efficiency and minimized material waste in over 80 published studies. However, it also noted that these models are often specific to certain nanocarrier types and typically need retraining when applied to a different platform [4].Dorsey et al. [5] reviewed machine learning applications in LNP formulation and process development, noting that most work concentrates on ionizable lipid selection, while helper lipids, PEGylated lipids, cholesterol, and, critically, their combinations have received comparatively little attention—a gap this review revisits in Section 8.

Deep learning models have also been applied at the single-particle level, predicting nanoparticle-drug interactions and the efficiency with which triple-negative breast cancer cells take up these particles. Researchers have extended this further, using deep learning to forecast nanoparticle tumor-delivery performance based on both chemical structure and biological characteristics. This reflects a broader shift in which formulation AI is no longer confined to manufacturability but extends into the prediction of biological performance [Section 4.2, ref 8].

4.3 Biologics and protein formulation

Formulating high-concentration protein therapeutics, particularly monoclonal antibodies intended for patient self-administration, presents a distinct set of challenges relative to small-molecule drugs, for which conventional screening methods are poorly suited. Protein aggregation and high viscosity are principal concerns, and these risks often must be predicted from little more than amino acid sequence and minimal biophysical data, sometimes before sufficient protein sample exists for laboratory testing.

Lai et al. [20] adopted a notable approach, combining molecular dynamics simulation data with sequence information from 20 IgG antibodies to train machine learning models predicting aggregation rates. Notably, using just two mechanistically meaningful features—spatial positive charge on the CDRH2 loop and solvent-accessible hydrophobic surface area—the models achieved a strong correlation with experimental measurements ($r = 0.89$), demonstrating that carefully chosen, scientifically grounded features can outperform high-dimensional alternatives.

Sawant et al. [21] used interpretable machine learning to identify which regions of antibody sequence and structure were responsible for excessive formulation viscosity. The mutations proposed by their models were subsequently tested experimentally and did reduce viscosity, providing a rare and unambiguous example of computational predictions directly informing protein engineering

Rai et al. [22] demonstrated that substantial data volumes are not always necessary: their deep learning models, built around biophysically meaningful molecular representations, predicted antibody viscosity even under scarce training data, offering a practical response to the field's characteristic small-dataset problem. Community resources like the Protein-Excipient and Protein-Protein Interactions in Formulation (PIPPI) database have demonstrated their worth by allowing ANN-based models to more reliably forecast protein aggregation and long-term stability in realistic formulation settings compared to studies conducted in isolated laboratories.

4.4 Three-dimensional (3D) printed and personalized dosage forms

Three-dimensional printing allows for the creation of dosage forms that are tailored to individual patients, with specific adjustments in shape, dosage, and drug-release profiles. However, determining the ideal printing parameters, such as nozzle temperature, infill density, and print speed, and understanding the role of geometry in drug release, present complex, nonlinear challenges that are well-suited for machine learning solutions

Sundarkumar et al. [19] developed a machine learning framework that unifies formulation and print-parameter optimization within a single model, addressing both simultaneously rather than as separate sequential steps, thereby improving process efficiency and integration.

Mazur et al. [17] trained artificial neural networks to predict the geometric shape a dosage form should take to achieve a target dose and release profile in 3D-printed oral medicines. Notably, they reported a modest test-set accuracy of 44.4%; this candor is valuable, as it clarifies the current limitations of geometry-to-release prediction in contrast to studies that report unverified high-accuracy results, and it identifies a clear direction for future research. More recently, conditional GANs have been applied to design multi-layered capsules with controlled, multi-phasic release patterns, an approach that seeks to balance formulation novelty against real-world physical and chemical plausibility [18].

4.5 Continuous manufacturing, digital twins, and Quality by Design

Digital twins integrated with process analytical technology have reportedly achieved API content consistency of up to 99.95% in continuous manufacturing, and patient-specific digital twins have predicted appropriate dosing within approximately 7% of actual clinical outcomes. These figures, however, derive from isolated case reports rather than large, multi-site studies, and should be interpreted as best-case demonstrations rather than representative performance [24].

A 2025 review mapped how digital twins integrate across the pharmaceutical pipeline, from tools such as AlphaFold3 accelerating molecular target discovery through to continuous manufacturing and personalized dosing. The review identified data integration, real-world model robustness, and regulatory complexity as the principal barriers to wider adoption, noting that proof-of-concept successes exist but have not yet overcome these obstacles at scale [24].

Generative AI is increasingly being incorporated into digital twin platforms, enabling researchers to simulate virtual patients and manufacturing runs. In silico trial design is consequently becoming relevant to tasks such as evaluating formulation changes and assessing bioequivalence risk [25].

In terms of industry adoption, digital twins and virtual manufacturing represented the fastest-growing category of AI deployment in pharmaceuticals as of 2024. Case studies report that AI-based visual inspection detects approximately 70% more defects while reducing false rejects by 60%, and that AI-guided deviation management shortens investigation times by 50–60% at adopting sites.

Cloud-based platforms are also improving the accessibility of these tools: the KNIME-based smart-formulation web platform for compounded drugs, for example, allows bench scientists to access predictive algorithms and integrated descriptor databases in real time without requiring a dedicated data scientist on every project [11].

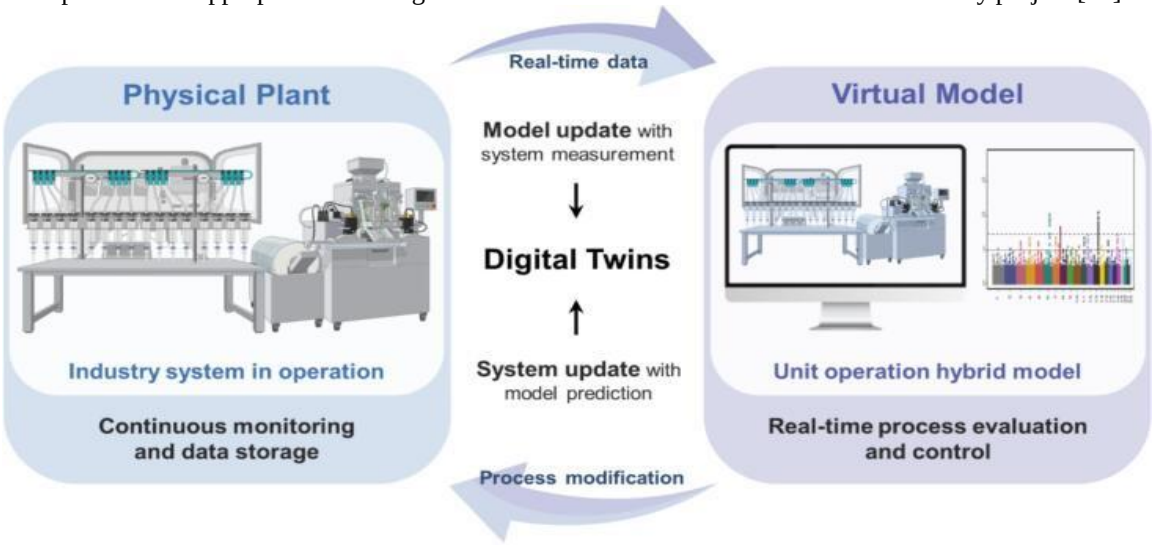

Figure.4 Digital Twin framework in pharmaceutical manufacturing, depicting the bidirectional loop between the physical plant and the virtual model. Real-time data from the physical system update the virtual model, while optimized control commands from the model support process evaluation, modification, and continuous improvement

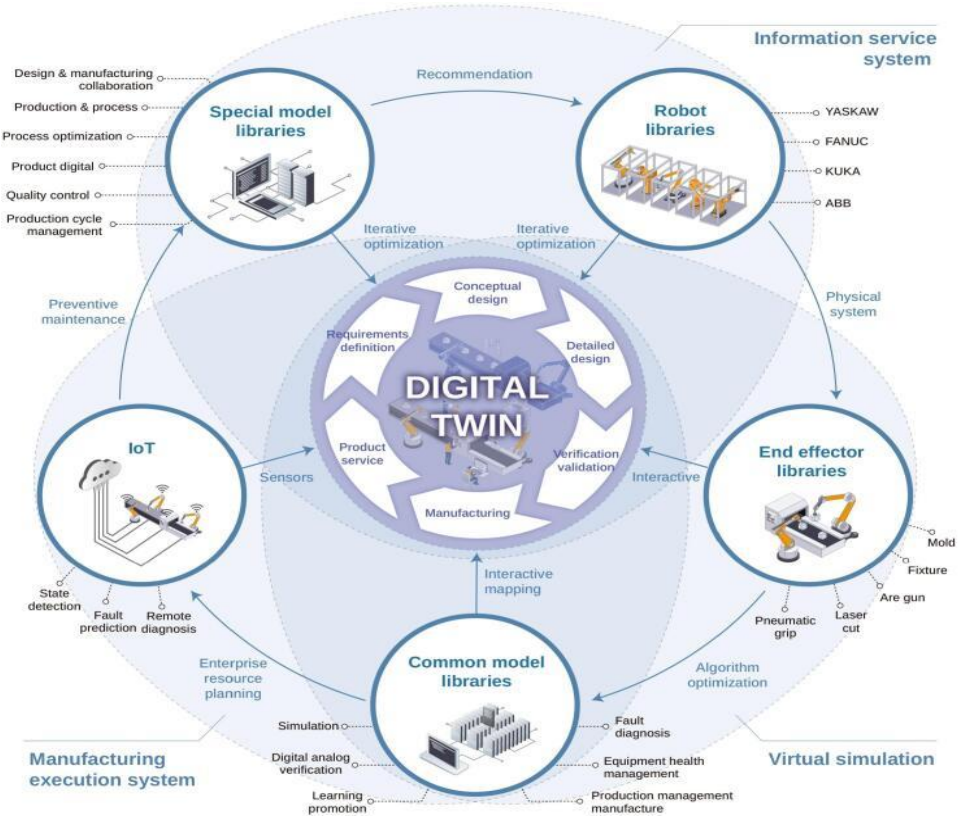

Fig. 5. Interactive mapping of conceptual design, details, verification/validation, manufacturing, product service, requirements, and optimization of a Digital Twin model (Robot libraries → End effector libraries → Common model libraries → IoT → Special model libraries) based on information service system, virtual simulation, and manufacturing execution system

5. Current Challenges

5.1 Data scarcity, quality, and standardization

Where public datasets exist, such as the fast-disintegrating tablet dataset underlying Momeni et al.'s work [1] or the PIPPI protein-formulation database, they represent isolated efforts rather than a coordinated, field-wide infrastructure, and inter-study comparability is undermined by differences in excipient nomenclature, measurement protocols, and reported units. This review's critical assessment is that this data problem, rather than algorithmic sophistication, is the binding constraint on further progress in formulation AI.

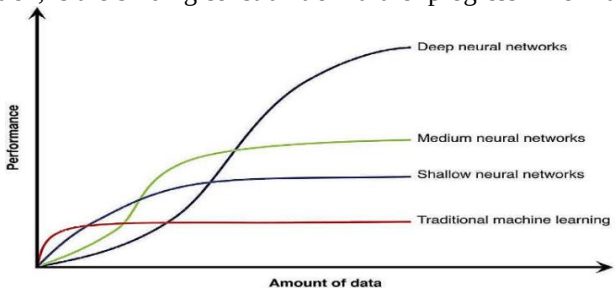

Fig.6.Comparison of different methods with data volume increases

5.2 Model interpretability and the "black-box" problem

Regulatory and clinical decision-making requires that a model's predictions be explicable in terms a reviewer or quality assessor can evaluate, yet many of the higher-performing architectures reviewed here, deep neural networks, GNNs, and generative models in particular, provide limited native interpretability. Encouragingly, several recent studies have deliberately incorporated interpretability tools, such as SHAP value analysis in long-acting injectable release prediction, or have selected simpler, mechanistically grounded features specifically to preserve interpretability in antibody viscosity and aggregation models [20,21,22]. However, such practices remain the exception rather than the norm across the broader corpus, and regulators have explicitly identified the "black box" problem, the difficulty of verifying how a model arrived at a given prediction, as a central barrier to regulatory acceptance of AI-derived data supporting product quality claims [13,49,50].

6. Clinical Applications

Most solid evidence so far comes from early-stage formulation work, but some applications are edging closer to clinical reality. In personalized medicine,

AI-driven 3D printing now makes it possible to tailor tablets for individual patients—you can tweak dose, shape, or drug release based on age, kidney or liver function, even genetics. That matters, especially for children and older adults, where standard doses often fall short. There's also work on building digital twins for patients. By merging PBPK simulations with AI, researchers can fine-tune how drugs are released, and these models, in early trials, have come within about 7% of matching real clinical dosing outcomes.

Sawant and colleagues showed how this improves convenience and eases the load on hospitals. AI tools sped up the rollout of mRNA vaccines, and now they're adapting the same systems for other RNA and gene-editing therapies, where keeping formulations stable and ensuring successful delivery into cells really decide whether the treatment works[19,24,28].

7. Comparison Tables

Table 1 summarizes the main AI and machine learning tools used in formulation development, matching each technique to specific problems, data needs, and key studies. Table 2 lines up traditional (DoE/QbD) and AI-powered approaches side by side, comparing them on factors relevant to how close they are to being used in real-world settings. The idea isn't to declare a winner—what works best depends a lot on the context and the dataset—but to help formulation scientists and reviewers pick the right method for each challenge.

Table 1. Principal AI/ML techniques applied in pharmaceutical formulation development

Technique	Typical Formulation Application	Mechanistic Basis	Typical Dataset Size	Representative Study
Random forest / gradient boosting	Disintegration time, hardness, dissolution prediction from tabular formulation data	Ensemble of decision trees; recursive variance-reducing splits	50–300 formulations	Momeni et al. [1]; Bharathi et al. [2]
Support vector machine	Formulation optimization / classification of pass–fail CQAs	Maximum-margin hyperplane separation in kernel-transformed feature space	30–200 formulations	Wang et al. [12]
Artificial neural network (ANN)	Multi-output CQA prediction; geometry–release mapping in 3D printing	Layered weighted summation with nonlinear activation	40–250 formulations	Mazur et al. [17]
Deep neural network	Dissolution-time prediction from raw-material databases	Multiple hidden layers enabling hierarchical feature learning	100s–1000s of records	Bharathi et al. [2]
Convolutional neural network (CNN)	Particle/granule/printlet image-morphology to performance correlation	Convolutional filters extracting spatial image features	Image datasets, 100s–1000s images	Reviewed in Shen et al. [4]
Graph neural network (GNN)	Excipient/polymer/lipid interaction modeling; LNP component design	Message passing over molecular/formulation graphs	Variable; often supplemented by virtual libraries	AGILE platform, in Shen et al. [4]; Bae et al. [6]
Bayesian optimization	Efficient search for optimal antibody/vaccine/adjuvant formulations	Gaussian-process surrogate model with acquisition-function-guided sampling	10–40 experiments (iterative)	Cited in Shen et al. [4]
Generative adversarial network (GAN)	Generation of novel 3D-printed capsule/formulation designs	Adversarial generator–discriminator training	Requires larger reference libraries	Elbadawi et al. [18]

Technique	Typical Formulation Application	Mechanistic Basis	Typical Dataset Size	Representative Study
Variational autoencoder / diffusion model	Novel excipient-combination proposal for nanoparticle design	Latent-space sampling / iterative denoising generation	Variable; benefits from pre-training	Shen et al. [4]
Digital twin (hybrid mechanistic–ML)	Real-time release testing; continuous manufacturing control	Mechanistic unit-operation model calibrated/corrected by ML using PAT data streams	Continuous real-time process data	Maharjan et al. [24]

8. Future Perspectives

Addressing the gaps above suggests several concrete directions for the field over the coming five years. First, pre-competitive data-sharing consortia, potentially modeled on existing public–private partnerships in drug discovery, could pool de-identified formulation datasets across companies and academic laboratories, dramatically expanding the training data available for model development while preserving proprietary formulation details through federated learning architectures that train shared models without centralizing raw data [4,26].

Second, hybrid mechanistic–ML models are likely to supplant purely data-driven deep learning as the dominant paradigm for regulatory-facing applications, since embedding physical constraints (mass balance, thermodynamics, established pharmacokinetic structure) both reduces data requirements and improves interpretability relative to unconstrained deep networks, a direction already visible in digital-twin and PBPK-coupled applications [24,28].

Third, explainable AI is likely to shift from an optional methodological enhancement to a near-mandatory feature of any formulation-AI application intended for regulatory submission, particularly as the FDA's risk-based credibility framework and the EMA's tiered explainability expectations mature into more prescriptive, formulation-specific guidance [13,15].

Fourth, generative AI isn't just about screening formulations anymore—it's reaching for something much bigger: actually designing new formulations from scratch, especially for nanoparticulate and 3D-printed systems. But if you look at the field's own assessment—take Mazur et al.'s reported 44.4% accuracy on geometry predictions—it's clear that generative models need a lot more rigorous validation before we can trust them to propose clinical-grade formulations without heavy human involvement.

Fifth, digital twins combined with real-time release testing stand out as the most immediate and concrete opportunity for regulatory and commercial progress. The pharmaceutical industry is pouring money and resources into digital twins-as of 2024, they're considered the fastest-growing AI segment in pharma manufacturing. Regulators already prioritize continuous manufacturing and process analytical technology (PAT)[24,25], so this is a natural fit. But unlocking real value won't come from isolated, single-company pilots[13].Regulators are already testing these tools internally, for example, the FDA's 'Elsa' assistant. These systems promise to support formulation scientists not just through prediction or generation, but by helping synthesize literature, suggest experiments, and draft regulatory documentation. [13,51].

9. CONCLUSION:

Conclusion In just five years, artificial intelligence has moved from an experimental curiosity to a practical tool in pharmaceutical formulation science. There's solid, critically examined evidence of AI's impact—whether in oral solid dosage forms, nanoparticulate and lipid carriers, biologics, 3D-printed personalized medicines, or continuous manufacturing. The strongest results show up in areas with well-bounded problems and good data, or where the science itself constrains the models—antibody viscosity reduction validated in the lab, ensemble models predicting tablet properties from open datasets. By contrast, evidence thins out just where AI's promise is greatest: for entirely new chemical entities with no prior history, for solutions that work across different technology platforms, and for digital twins that regulators have fully endorsed and validated at commercial scale.

10.REFERENCES:

1. Momeni M, Afkanpour M, Rakhshani S, Mehrabian A, Tabesh H. A prediction model based on artificial intelligence techniques for disintegration time and hardness of fast disintegrating tablets in pre-formulation tests. BMC Med Inform DecisMak. 2024;24:87.https://pubmed.ncbi.nlm.nih.gov/38539201/
2. Bharathi M, Kamaraj R, Murugaanandam S, Navyaja K, Kumar TS. A data-driven approach to predict the in vitro dissolution time of

sustained-release tablets using raw material databases and machine learning algorithms. Pharmacia. 2024;71:1–7.https://pharmacia.pensoft.net/article/122772/
3. Deng J, Ye Z, Zheng W, Chen J, Gao H, Wu Z, et al. Machine learning in accelerating microsphere formulation development. Drug DelivTransl Res. 2023;13:966–982.link.springer.com/article/10.1007/s13346-022-01253-z
4. Shen C, Zhang M, Lu M, Chang E, Gao Z, Ban W, Liu Q, Zuo Z, Jiang C. Machine learning empowered formulation design, optimization and characterization of nanoparticulate drug delivery systems: current applications, challenges, and future perspectives. Acta Pharm Sin B. 2026 [Epub ahead of print]. doi:10.1016/j.apsb.2025.12.011.https://doi.org/10.1016/j.apsb.2025.12.011
5. Dorsey PJ, Lau CL, Chang TC, Doerschuk PC, D'Addio SM. Review of machine learning for lipid nanoparticle formulation and process development. J Pharm Sci. 2024;113:3413–3433.https://pubmed.ncbi.nlm.nih.gov/39341497/
6. Bae S, et al. Artificial intelligence-driven nanoarchitectonics for smart targeted drug delivery. Adv Mater. 2025;37:e2510239.https://pubmed.ncbi.nlm.nih.gov/40772340/
7. Sun Y, Qin S, Li Y, Hasan N, Li YV, Liu J. Machine learning integrated with in vitro experiments for study of drug release from PLGA nanoparticles. Sci Rep. 2025;15:4218.https://pubmed.ncbi.nlm.nih.gov/39905099/
8. Ali R, Balamurali M, Varamini P. Deep learning-based artificial intelligence to investigate targeted nanoparticles' uptake in TNBC cells. Int J Mol Sci. 2022;23:16070.https://pubmed.ncbi.nlm.nih.gov/36555718/
9. Suriyaamporn P, Pamornpathomkul B, Patrojanasophon P, Ngawhirunpat T, Rojanarata T, Opanasopit P. The artificial intelligence-powered new era in pharmaceutical research and development: a review. AAPS PharmSciTech. 2024;25:188.https://www.researchgate.net/publication/383155390
10. Serrano DR, Luciano FC, Anaya BJ, Ongoren B, Kara A, Molina G, et al. Artificial intelligence (AI) applications in drug discovery and drug delivery: revolutionizing personalized medicine. Pharmaceutics. 2024;16:1328. https://doi.org/10.3390/pharmaceutics16101328
11. Grigoryan A, Helfrich S, Lequeux V, Lapras B, Marchand C, Merienne C, et al. Smart formulation: AI-driven web platform for optimization and stability prediction of compounded pharmaceuticals using KNIME. Pharmaceuticals. 2025;18:1240.https://pubmed.ncbi.nlm.nih.gov/40872628/
12. Wang S, Yang J, Chen H, et al. A strategy for the effective optimization of pharmaceutical formulations based on parameter-optimized support vector machine model. AAPS PharmSciTech. 2022;23:210.https://pubmed.ncbi.nlm.nih.gov/35102463/
13. Niazi SK. Regulatory perspectives for AI/ML implementation in pharmaceutical GMP environments. Pharmaceuticals. 2025;18:901.https://pmc.ncbi.nlm.nih.gov/articles/PMC12195787/
14. Mirakhori F, Niazi SK. Harnessing the AI/ML in drug and biological products discovery and development: the regulatory perspective. Pharmaceuticals. 2025;18:47.https://pmc.ncbi.nlm.nih.gov/articles/PMC11769376/
15. Niazi SK. A critical review of the FDA's draft guidance on artificial intelligence in drug and biological product regulation. J Chem. 2026;2026:5202999.https://doi.org/10.1155/joch/5202999
16. Vora LK, Gholap AD, Jetha K, Thakur RRS, Solanki HK, Chavda VP. Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics. 2023;15:1916.https://pmc.ncbi.nlm.nih.gov/articles/PMC10385763/
17. Mazur H, Erbrich L, Quodbach J. Investigations into the use of machine learning to predict drug dosage form design to obtain desired release profiles for 3D printed oral medicines. Pharm Dev Technol. 2023;28:219–231.https://doi.org/10.1080/10837450.2023.2173778
18. Elbadawi M, Gaisford S, Basit AW. AI-driven design of customized 3D-printed multi-layer capsules with controlled drug release profiles for personalized medicine. Int J Pharm. 2024;658:124204.https://pubmed.ncbi.nlm.nih.gov/38615804/
19. Sundarkumar V, Nagy ZK, Reklaitis GV. Machine learning enabled integrated formulation and process design framework for a pharmaceutical 3D printing platform. AIChE J. 2023;69:e17990.https://pmc.ncbi.nlm.nih.gov/articles/PMC10785158/
20. Lai PK, Gallegos A, Mody N, Sathish HA, Trout BL. Machine learning prediction of antibody aggregation and viscosity for high concentration formulation development of protein therapeutics. mAbs.

2022;14:2026208.https://pmc.ncbi.nlm.nih.gov/articles/PMC8794240/
21. Sawant MS, Streu CN, Wu L, Underhill P, Pelegri-O'Day E, Maglalang E, Winters D, Tessier PM. Reduction of monoclonal antibody viscosity using interpretable machine learning. mAbs. 2024;16:2303781.https://pmc.ncbi.nlm.nih.gov/articles/PMC10939158/
22. Rai BK, Apgar JR, Bennett EM. Low-data interpretable deep learning prediction of antibody viscosity using a biophysically meaningful representation. Sci Rep. 2023;13:2917.https://pmc.ncbi.nlm.nih.gov/articles/PMC9941094/
23. Kingsbury JS, Saini A, Auclair SM, Fu L, Lantz MM, Halloran KT, et al. A single molecular descriptor to predict solution behavior of therapeutic antibodies. Sci Adv. 2020;6:eabb0372.https://pmc.ncbi.nlm.nih.gov/articles/PMC7457339/
24. 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. 2025;10:100409.https://pmc.ncbi.nlm.nih.gov/articles/PMC12516570/
25. Bordukova M, Makarov N, Rodriguez-Esteban R, Schmich F, Menden MP. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. Expert Opin Drug Discov. 2024;19:33–42.https://pubmed.ncbi.nlm.nih.gov/37887266/
26. Pharma Excipients. Current status on the convergence of artificial intelligence and formulation development in industry: a review [Internet]. 2025 Dec [cited 2026 Jul 7]. Available from: https://www.pharmaexcipients.com/news/convergence-ai-formulation-industry/.
27. U.S. Food and Drug Administration. Artificial intelligence for drug development [Internet]. Silver Spring (MD): FDA; 2025 [cited 2026 Jul 7]. Available from: https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/artificial-intelligence-drug-development.
28. European Pharmaceutical Review. Applying AI to enhance drug formulation and development [Internet]. 2026 [cited 2026 Jul 7]. Available from: https://www.europeanpharmaceuticalreview.com/article/271637/artificial-intelligence-ai-drug-formulation-development/.
29. Future Journal of Pharmaceutical Sciences. Influence of artificial intelligence in modern pharmaceutical formulation and drug development. Future J Pharm Sci. 2024;10:33.https://pubmed.ncbi.nlm.nih.gov/41092928/
30. Akseli I, Xie J, Schultz L, Ladyzhynsky N, Bramante T, He X, et al. A practical framework toward prediction of breaking force and disintegration of tablet formulations using machine learning tools. J Pharm Sci. 2017;106:234–247.https://pubmed.ncbi.nlm.nih.gov/28340955/
31. Aksu B, Paradkar A, De Matas M, Özer Ö, Güneri T, York P. Quality by design approach: application of artificial intelligence techniques of tablets manufactured by direct compression. AAPS PharmSciTech. 2012;13:1138–1146.https://pubmed.ncbi.nlm.nih.gov/22956056/
32. Fink E, Celikovic S, Rehrl J, Sacher S, Urich JAA, Khinast J. Predicting dissolution kinetics of pharmaceutical tablets using data-driven models. Int J Pharm. 2023;641:123056.https://pubmed.ncbi.nlm.nih.gov/40930319/
33.Food and Drug Law Institute. Regulating the use of AI in drug development: legal challenges and compliance strategies [Internet]. 2025 [cited 2026 Jul 7]. Available from: https://www.fdli.org/2025/07/regulating-the-use-of-ai-in-drug-development-legal-challenges-and-compliance-strategies/https://www.fdli.org/2025/07/regulating-the-use-of-ai-in-drug-development-legal-challenges-and-compliance-strategies/