

CHEM-AI: BRIDGING CHEMISTRY AND AI FOR SMART DRUG DELIVERY

Sivahari R^{1*}, Dr. Sengodan R^{2*}, Akash N S^{3*}

^{1*} Assistant Professor, Department of Chemistry, V.S.B. College of Engineering Technical Campus, kinathukadavu, Coimbatore, Tamilnadu, India.

Email: sivaharikct@gmail.com

^{2*} Assistant Professor, Department of Physics, Kumaraguru College of Technology, Saravanampatti, Chinnavedampatti Post, Coimbatore, Tamilnadu, India.

Email: sengodan78@gmail.com

^{3*} UG Scholar, Department of Artificial Intelligence and Data Science, V.S.B. College of Engineering Technical Campus, kinathukadavu, Coimbatore, Tamilnadu, India.

Email: akashns232006@gmail.com

Abstract

Artificial intelligence (AI) is increasingly useful in pharmaceutical formulation because drug-delivery systems require simultaneous control of composition, process variables, physicochemical properties, stability, release behavior, and therapeutic performance. This manuscript presents an explainable machine-learning framework that connects chemistry-informed molecular descriptors with formulation and drug-delivery variables to support predictive optimization. The framework integrates curated chemical and formulation data, molecular and material representations, task-specific machine-learning models, uncertainty estimation, explainability, multi-objective ranking, and experimental feedback. Particular emphasis is placed on predicting formulation attributes such as particle size, encapsulation efficiency, drug loading, dissolution or release characteristics, stability-related responses, and delivery performance while preserving chemical plausibility and formulation constraints. Three illustrative figures are retained: a compatibility heat map, a bubble map of AI–chemistry application areas, and a molecular-weight histogram. These figures are synthetic methodological illustrations rather than experimental results. Three tables summarize AI strategies for drug-delivery problems, reproducibility requirements, and a multi-objective formulation decision matrix. The proposed framework is intended to help researchers combine chemical knowledge, formulation science, and explainable AI in a reproducible workflow for drug-delivery system design and optimization.

Keywords: Artificial Intelligence; Machine Learning; Pharmaceutical Formulation; Drug Delivery Systems; Explainable AI; Formulation Optimization; Molecular Descriptors; Controlled Release; Nanocarriers; Predictive Modeling

How to cite this article: Sivahari R^{1*} Dr. Sengodan R^{2*} Akash N S^{3*}. CHEM-AI: BRIDGING CHEMISTRY AND AI FOR SMART DRUG DELIVERY. Int J Drug Deliv Technol. 2026;16(83s):593-597. DOI: 10.25258/ijddt.16.83s.73.

Introduction

Drug-delivery development requires the coordinated optimization of drug properties, excipients, composition, processing conditions, dosage-form characteristics, and performance outcomes. Conventional formulation development often relies on sequential experimentation and empirical adjustment, which can become inefficient when many interacting variables influence critical quality attributes. AI-assisted chemistry provides an opportunity to connect molecular-level information with formulation-level responses and to prioritize experiments before extensive laboratory screening.

A useful AI system for drug delivery should not optimize a single response in isolation. A formulation may show high encapsulation efficiency but poor stability, an attractive dissolution profile but inadequate manufacturability, or favorable particle size but unacceptable drug loading. The practical objective is therefore a constrained, multi-objective optimization problem in which chemical compatibility, formulation feasibility, process robustness, and uncertainty are considered together.

The proposed framework, termed ChemAI-Bridge, adapts the earlier AI–chemistry concept to the specific needs of drug-delivery research. It links molecular

structure and chemical descriptors to excipient selection, formulation composition, process conditions, critical quality attributes, and delivery performance. Explainability is treated as a scientific requirement because formulation scientists need to understand which molecular or process variables drive a prediction and whether the proposed formulation is within the model's reliable domain.

The objective of this manuscript is to provide a reproducible, chemistry-informed AI framework for predictive optimization of drug-delivery systems. The framework is methodological and does not claim newly generated experimental efficacy data. Its intended use is to organize formulation datasets, select appropriate machine-learning models, quantify prediction reliability, rank candidate formulations, and guide subsequent experimental validation.

Drug-deliver y proble m class	Chemical/for mulation input	Represe ntative AI strategy	Primary output	Key validati on
Formul ation	Drug + excipient	GNN, gradient	Predict ed	Externa l test

property prediction	descriptors + process variables	boosting, transformers	CQA / uncertainty	set + calibration
Release and dissolution prediction	Molecular descriptors + formulation composition	Ensemble ML / sequence models	Release profile or rate constants	Independent formulation batches
Excipient and carrier selection	Drug structure + excipient/material descriptors	Similarity models, GNN, multi-task ML	Ranked compatible candidates	Chemical compatibility + experiments
Nanocarrier optimization	Drug, carrier, composition, process variables	Multi-objective ML / surrogate models	Particle size, loading, efficiency predictions	Prospective formulation testing
Generative formulation design	Target profile + chemical/formulation constraints	Generative models + optimization	Candidate compositions or materials	Feasibility + laboratory confirmation
Stability and ADME T prioritization	Structure + formulation + assay descriptors	Multi-task ML / calibrated ensembles	Risk and performance profile	External validation + uncertainty

Table 1. A framework-level mapping; entries are methodological examples, not performance claims

I. ChemAI-Bridge Framework for Drug-Delivery Optimization

A. Chemical and formulation data layer

The data layer integrates molecular, excipient, formulation, process, and performance information. Drug molecules can be encoded using SMILES, molecular graphs, fingerprints, physicochemical descriptors, or learned embeddings, while excipients can be represented by chemical descriptors and functional attributes. Public molecular-learning benchmarks demonstrate the value of standardized representations, whereas message-passing models provide a route to learning chemically meaningful graph features. Reaction and process records should retain relevant conditions because synthesis, formulation preparation, and material processing can strongly influence downstream performance. For drug-delivery datasets, measured attributes such as particle size, polydispersity, zeta potential, encapsulation efficiency, drug loading, viscosity, dissolution rate, and release percentage should retain units, assay conditions,

batch information, and provenance. Chemical-structure curation is essential because salts, stereochemistry, tautomerism, duplicates, and inconsistent naming can introduce artificial patterns into models. [1], [2], [3], [4], [5]

B. AI model layer

The model layer is selected according to the formulation question. Tree-based models are useful when the input consists of structured formulation variables, while graph neural networks can capture molecular connectivity and interactions between drug or excipient structures. Transformer-based models can process sequence-like chemical representations and can be extended to formulation-related text or reaction information. Generative models, including molecular graph generation and diffusion approaches, can support exploration of candidate compounds or excipient combinations when chemical constraints are imposed. For formulation optimization, hybrid models can combine molecular descriptors with process variables, enabling predictions of critical quality attributes rather than relying on chemistry alone. The central principle is that model selection should follow the drug-delivery decision being made, not simply the availability of a more complex architecture. [6], [7], [8], [9], [10]

C. Explainability and uncertainty layer

Drug-delivery decisions can involve costly experiments, so predicted values should be accompanied by estimates of reliability. Time-aware validation, calibration, conformal prediction, ensemble uncertainty, and applicability-domain analysis can identify predictions that should be treated cautiously. Explainable AI can further identify influential molecular fragments, excipient descriptors, formulation variables, or process parameters. In a formulation context, explanations should answer practical questions such as whether an increase in polymer concentration is driving a predicted release change, whether a lipid descriptor is associated with particle-size behavior, or whether a molecular feature is responsible for a predicted loading limitation. Chemically meaningful attribution is preferable to generic feature importance when it can be linked to a plausible mechanism. [11], [12], [13], [14], [15]

D. Multi-objective decision and experimental feedback layer

The final layer converts model outputs into an experimental priority list. A generic formulation score can be expressed as $S = w_1Q_1 + w_2Q_2 + w_3Q_3 + w_4Q_4 - w_5R$, where Q_1 – Q_4 may represent target delivery performance, formulation quality, stability, and model confidence, and R represents risk or constraint violations. The weights should be declared before candidate selection or tested through sensitivity analysis. Selected formulations are then prepared and experimentally characterized, and the new observations are returned to the model through an active-learning cycle. This creates a closed loop in which chemical knowledge guides model construction, AI prioritizes experiments, and

experimental evidence updates the next round of formulation decisions. [16], [17], [18], [19], [20]

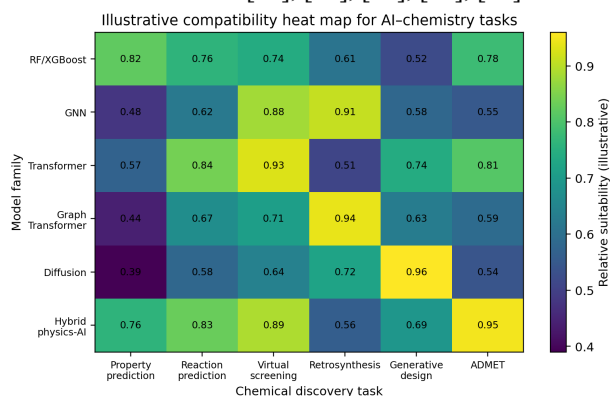


Figure 1. Illustrative compatibility heat map retained from the previous draft. The values are synthetic and demonstrate how AI model families could be compared across drug-delivery-relevant chemical tasks; they do not represent measured performance.

II. Methodology and Reproducibility Protocol

A study using ChemAI-Bridge should define the drug, excipient, formulation, and process datasets before model development. Each dataset should document inclusion and exclusion criteria, chemical standardization, missing-data handling, batch effects, measurement units, and assay conditions. Data should be partitioned so that closely related formulations or molecular analogues do not leak across training and test sets. Scaffold-based or time-aware partitions can be useful when the goal is prospective prediction. The complete workflow should record model architecture, hyperparameters, random seeds, preprocessing operations, evaluation metrics, and the criteria used to select formulations for laboratory testing. [21], [22], [23], [24], [25]

Component	Minimum reporting requirement	Preferred practice
Dataset provenance	Drug, excipient, formulation and assay sources	Versioned archive + persistent identifier
Chemical standardization	Salt, tautomer, stereochemistry and duplicate policy	Documented automated pipeline + audit sample
Formulation variables	Composition, process conditions, units and batch identifiers	Machine-readable data dictionary + provenance
Data splitting	Train/validation/test strategy	Scaffold, formulation-family or temporal split with leakage audit
Model specification	Architecture and major hyperparameters	Code, environment file and fixed

		seed
Evaluation	Metrics and confidence intervals	External formulation dataset + calibration analysis
Interpretability	Method used and scope	Chemically and formulation-relevant attribution / counterfactuals
Uncertainty	Whether confidence is estimated	Conformal or calibrated ensemble + applicability domain
Candidate selection	Ranking rule and thresholds	Predefined multi-objective decision rule
Experimental loop	How new formulation data enter the model	Active-learning protocol with versioned updates

Table 2. Recommended reproducibility checklist for an AI-assisted drug-delivery study

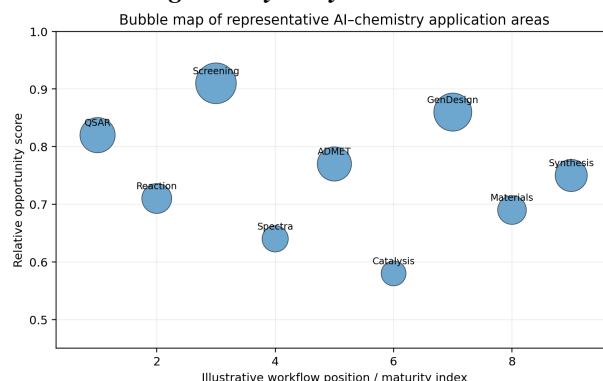


Figure 2. Illustrative bubble map retained from the previous draft. The visual is synthetic and is used only to show how major AI-chemistry application areas can be positioned within a drug-delivery decision workflow; bubble size and position do not represent validated performance or publication volume.

III. Results and Interpretation

A. Predictive targets for drug-delivery optimization

A chemistry-informed drug-delivery model should evaluate multiple critical quality attributes rather than a single endpoint. Useful targets include particle size, polydispersity, zeta potential, encapsulation efficiency, drug loading, viscosity, dissolution behavior, release kinetics, stability indicators, and—where appropriate—permeation or targeting-related outcomes. Molecular descriptors can provide mechanistic context, while formulation and process variables describe the controllable experimental space. The practical value of the model is therefore determined by whether it can support reliable formulation decisions within the intended chemical and process domain. ^{26,27,28}

B. Multi-objective formulation prioritization

Formulation selection is inherently multi-objective. A candidate with excellent encapsulation may have poor release control, while a formulation with rapid release may compromise stability or manufacturability. The decision layer should therefore rank candidates using explicit objectives and penalties. The retained heat map and bubble map can be interpreted as workflow-level illustrations: they show how different AI model families and application areas may be matched to different drug-delivery tasks, rather than reporting experimental scores. In a real study, the same framework can be populated with measured formulation data and validated prediction intervals.

C. Chemistry-informed reaction and material intelligence

Drug-delivery development also depends on chemical and material transformations, including synthesis of carriers, polymer modification, nanoparticle preparation, cross-linking, lipid or polymer assembly, and other formulation-enabling steps. Reaction and molecular models can help organize these chemical relationships, identify relevant structural patterns, and prioritize feasible material or carrier candidates. For this reason, the framework treats reaction intelligence as supporting formulation development rather than as an independent drug-discovery endpoint.^{29,30}

D. Molecular-property and formulation-profile analysis

Distributional analysis is an important quality-control step before model training. The retained histogram can be used to illustrate inspection of molecular-weight distributions, but in an IJDDT-focused study the same analysis should be extended to formulation variables and critical quality attributes. Researchers should examine skewness, outliers, batch effects, missingness, and the overlap between training and prospective formulation spaces. Such profiling helps identify whether the model is being asked to extrapolate beyond the chemistry and formulation domain represented in the training data.

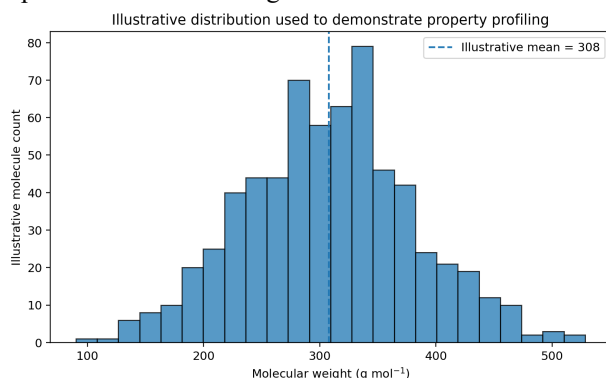


Figure 3. Illustrative molecular-weight distribution retained from the previous draft. The distribution is synthetic and is intended to demonstrate property profiling for chemistry-informed drug-delivery datasets.

Criterion	Example	Directio	Illustrativ
-----------	---------	----------	-------------

	measurement	n	e weight
Drug-delivery performance	Predicted release / delivery target	Higher is better	0.25
Encapsulation efficiency	Predicted or measured loading efficiency	Higher is better	0.15
Stability	Predicted stability-related response	Higher is better	0.20
Formulation feasibility	Process and manufacturability score	Higher is better	0.15
Model confidence	Calibrated prediction confidence	Higher is better	0.15
Risk penalty	Instability / incompatibility flags	Lower is better	0.10

Table 3. Illustrative decision matrix. The weights are examples for framework demonstration and should be replaced by study-specific, experimentally justified values.

IV. Discussion

The main value of combining AI-assisted chemistry with drug-delivery science is the ability to connect molecular information with formulation decisions in a transparent workflow. Instead of treating formulation optimization as a sequence of isolated experiments, the framework supports integrated prediction of quality attributes, identification of influential variables, uncertainty-aware candidate ranking, and iterative experimental learning. This approach is particularly useful when formulation spaces contain many interacting drug, excipient, and process variables.

V. Limitations and Future Work

This manuscript is a methodological framework and does not report a newly generated experimental formulation dataset. The three figures therefore remain synthetic illustrations and should not be cited as measured performance. The framework also depends on the quality, consistency, and chemical diversity of the underlying formulation data. Future validation should use independent drug-delivery datasets and prospective experiments covering multiple dosage forms and delivery technologies.

Future work should benchmark the framework on oral, transdermal, ocular, injectable, pulmonary, nanoparticle, lipid-based, polymeric, and controlled-release systems. Model comparisons should include calibration and applicability-domain analysis, while prospective experiments should test whether uncertainty-aware ranking reduces laboratory workload and improves formulation success rates. Integration with automated formulation platforms and active-learning systems could further shorten the design–test–learn cycle.

Conclusion

Artificial Intelligence-Assisted Chemistry for Predictive Optimization of Drug Delivery Systems provides a chemistry-informed framework for integrating molecular descriptors, formulation variables, process conditions, machine-learning prediction, explainability, uncertainty estimation, and experimental feedback. The approach is designed to support practical drug-delivery decisions while preserving chemical plausibility and reproducibility. By combining AI with formulation science rather than treating AI as a stand-alone prediction tool, the framework offers a structured basis for future experimental validation and data-driven development of optimized drug-delivery systems.

References

- Shilpa; Kashyap, G.; Sunoj, R. B. Recent Applications of Machine Learning in Molecular Property and Chemical Reaction Outcome Predictions. *J. Phys. Chem. A* 2023, 127, 8253–8271. DOI: 10.1021/acs.jpca.3c04779.
- Wu, Z. et al. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chem. Sci.* 2018, 9, 513–530.
- Schwaller, P. et al. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent. Sci.* 2019, 5, 1572–1583.
- Lowe, D. M. Extraction of Chemical Structures from the Literature Using a Neural Network. PhD Thesis, University of Cambridge, 2012.
- Gilmer, J. et al. Neural Message Passing for Quantum Chemistry. *Proc. ICML* 2017, 1263–1272.
- Coley, C. W. et al. A Graph-Convolutional Neural Network Model for the Prediction of Chemical Reactivity. *Chem. Sci.* 2019, 10, 370–377.
- Sorkun, M. C. et al. A Quantitative Structure–Property Relationship Database for Machine Learning. *J. Cheminform.* 2019, 11, 43.
- Tetko, I. V. et al. Critical Assessment of QSAR Models of Melting Point. *J. Chem. Inf. Model.* 2020, 60, 158–174.
- Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify: On the Importance of Chemical Structure Curation in QSAR Modeling and Virtual Screening. *J. Chem. Inf. Model.* 2010, 50, 1189–1204.
- Breiman, L. Random Forests. *Mach. Learn.* 2001, 45, 5–32.
- Yang, K. et al. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* 2019, 59, 3370–3388.
- Das, M.; Ghosh, A.; Sunoj, R. B. Advances in Machine Learning with Chemical Language Models in Molecular Property and Reaction Outcome Predictions. *J. Comput. Chem.* 2024, 45, 1160–1176. DOI: 10.1002/jcc.27315.
- Schwaller, P. et al. Mapping the Space of Chemical Reactions Using Attention-Based Neural Networks. *Nat. Mach. Intell.* 2021, 3, 144–152.
- Gómez-Bombarelli, R. et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* 2018, 4, 268–276.
- Jin, W.; Barzilay, R.; Jaakkola, T. Junction Tree Variational Autoencoder for Molecular Graph Generation. *ICML* 2018, 2323–2332.
- Hoogeboom, E. et al. Equivariant Diffusion for Molecule Generation in 3D. *ICML* 2022, 2793–2803.
- Venkatasubramanian, V. The Promise of Artificial Intelligence in Chemical Engineering: Is It Ready for Prime Time? *AIChE J.* 2019, 65, 466–478.
- Tu, Z.; Stuyver, T.; Coley, C. W. Predictive Chemistry: Machine Learning for Reaction Deployment, Reaction Development, and Reaction Discovery. *Chem. Sci.* 2023, 14, 226–244. DOI: 10.1039/D2SC05089G.
- Sheridan, R. P. Time-Split Cross-Validation as a Method for Estimating the Goodness of Prospective Prediction. *J. Chem. Inf. Model.* 2013, 53, 783–790.
- Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K. Q. On Calibration of Modern Neural Networks. *ICML* 2017, 1321–1330.
- Angelopoulos, A. N.; Bates, S. Conformal Prediction: A Gentle Introduction. *Found. Trends Mach. Learn.* 2023, 16, 494–591.
- Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles. *NeurIPS* 2017, 30.
- Yang, Y.; Li, X.; Song, R.; Zeng, X. Applicability Domain Analysis for Chemical Machine Learning. *J. Cheminform.* 2022, 14, 70.
- Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug Discovery with Explainable Artificial Intelligence. *Nat. Mach. Intell.* 2020, 2, 573–584.
- Schwaller, P. et al. Data-Driven Chemical Reaction Prediction with Attention-Based Neural Networks. *Chem. Sci.* 2019, 10, 8989–8996.
- Huang, K. et al. Therapeutics Data Commons: Machine Learning Datasets and Tasks for Therapeutics. *NeurIPS Datasets and Benchmarks* 2021.
- Brown, N.; Fiscato, M.; Segler, M. H. S.; Vaucher, A. C. R. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* 2019, 59, 1096–1108.
- Winter, R. et al. Molecular Optimization by Evolutionary Search with Uncertainty-Aware Objectives. *J. Chem. Inf. Model.* 2020, 60, 420–432.
- Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Next-Generation Experiment Design: AI-Driven Adaptive Chemistry. *Trends Chem.* 2019, 1, 282–291.
- Burger, B. et al. A Mobile Robotic Chemist. *Nature* 2020, 583, 237–241.