ARTICLE

Machine Learning-Based Prediction of Drug Release and Stability in Advanced Pharmaceutical Formulations

[1] Sangeeta Bhogal, [2]Mr. Dinesh L. Bawankar, [3]Dr. Sonia Dhiman

[1]Assistant Professor Department of Chemistry, The ICFAI University, Himachal Pradesh
[2]Anurag college of Pharmacy, Warthi, Bhandara
[3]Amity Institute of Indian System of Medicine (AIISM), Amity University Noida, U. P., India. sdhiman1@amity.edu
ORCID ID: 0009-0004-8071-1968
Corresponding author: Sangeeta Bhogal, Sangeeta.bhogal30@gmail.com

ABSTRACT

Advanced pharmaceutical formulations improve solubility, bioavailability, controlled release, and therapeutic performance, but their development remains difficult because drug release and stability are governed by nonlinear interactions among drug properties, excipient composition, process variables, and storage conditions. This study developed an integrated machine-learning framework for predicting drug release and stability outcomes in advanced pharmaceutical formulations. The structured dataset comprised 486 formulation records, including polymeric nanoparticles, lipid-based carriers, nanoemulsions, hydrogels, solid dispersions, matrix tablets, and controlled-release systems. A total of 28 descriptors were organized into drug-related, formulation-related, and experimental-condition variables. Drug release was modeled as a regression endpoint, while stability was evaluated using both regression and classification tasks. Linear regression, support vector regression, artificial neural networks, gradient boosting, random forest, and extreme gradient boosting were compared using cross-validation and independent test-set evaluation. Extreme gradient boosting achieved the best release-prediction performance, with an R2 of 0.91, RMSE of 6.84%, MAE of 4.72%, and MAPE of 8.63%. For stability classification, the same model achieved an accuracy of 0.89, F1-score of 0.86, and ROC-AUC of 0.93. Feature-importance and interpretability analyses identified particle size, polymer or lipid concentration, drug loading, encapsulation efficiency, storage temperature, relative humidity, and storage duration as dominant predictors. The findings demonstrate that machine learning can support data-driven formulation development by improving prediction accuracy, clarifying critical variables, and converting complex formulation data into practical decision-support outputs.

Keywords: machine learning, drug release, pharmaceutical stability, computational pharmaceutics, drug delivery systems, formulation optimization, explainable artificial intelligence, data-driven pharmaceutics

How to cite this article: Bhogal S, Bawankar DL, Dhiman S., Machine Learning-Based Prediction of Drug Release and Stability in Advanced Pharmaceutical Formulations. Int J Drug Deliv Technol. 2026;16(89s): 2010-2018; DOI: 10.25258/ijddt.16.89s.188

1. INTRODUCTION

Advanced pharmaceutical formulations have become central to modern drug development because many therapeutic compounds exhibit poor aqueous solubility, limited permeability, physical instability, rapid clearance, or unfavorable pharmacokinetic behavior. Formulation technologies such as polymeric nanoparticles, lipid nanoparticles, liposomes, nanoemulsions, hydrogels, solid dispersions, matrix tablets, and controlled-release systems are designed to overcome these limitations by improving drug loading, modifying release behavior, protecting labile molecules, and enhancing therapeutic performance. The increasing use of computational pharmaceutics and machine learning in formulation development reflects the need for predictive approaches capable of handling complex formulation variables [1,2].

Drug release is a critical quality attribute because it determines how quickly and completely an active pharmaceutical ingredient becomes available from its dosage form. In controlled-release and nanocarrier systems, release behavior is influenced by particle size, matrix structure, polymer concentration, lipid composition, drug loading, surface charge, swelling, erosion, diffusion, degradation, and dissolution-medium conditions. Recent studies have shown that surrogate modeling, spectroscopic data, and imaging-based inputs can support prediction of dissolution and release behavior [3-5]. Real-time prediction of dissolution profiles has also been demonstrated for coated oral dosage forms, supporting the broader movement toward predictive and data-rich pharmaceutical development [6].

Formulation stability is equally important because physical or chemical instability may lead to reduced potency, altered release behavior, aggregation, phase separation, crystallization, hydrolysis, oxidation, or drug-excipient incompatibility. Stability assessment traditionally relies on long-term and accelerated studies, but such studies are resource-intensive and may not fully reveal nonlinear interactions among formulation composition, storage temperature, relative humidity, and time. Recent predictive stability studies indicate that short-term and accelerated data can support long-term stability assessment when combined with suitable kinetic and statistical modeling [7,8].

Despite these advances, four gaps remain important in the present subject area. First, many machine-learning studies focus on a single dosage form, drug class, or endpoint, which limits generalization across advanced formulation platforms. Second, drug release and stability are often modeled separately, although formulation decisions must balance both outcomes. Third, some models emphasize numerical accuracy without sufficient pharmaceutical interpretation of critical variables. Fourth, few frameworks convert model outputs into practical formulation-screening tools such as desirability regions or stability-risk maps. Recent work in nanomedicine data curation, machine-learning-guided formulation development, and AI-enabled pharmaceutical technology provides a foundation for addressing these limitations

[9-11]. Drug-excipient compatibility prediction further supports risk-oriented formulation screening [12,13].

The specific objectives of this study were to: (i) construct a structured formulation dataset containing drug-related, formulation-related, and experimental-condition descriptors; (ii) compare multiple supervised machine-learning algorithms for drug release regression and stability prediction; (iii) identify the critical variables controlling release and stability using interpretable model outputs; and (iv) translate model predictions into formulation-relevant decision outputs that can support candidate prioritization and risk assessment. These objectives were designed to move the article beyond simple algorithm comparison and toward a more useful pharmaceutical decision-support framework.

Table 1: Evidence base, subject-area gaps, and the response of the present framework

Evidence area	Representative recent work	Finding used as a basis	Remaining gap addressed
Computational pharmaceutics and ML-directed formulation	Computational pharmaceutics and ML formulation-development reviews [1,2]	Machine learning can reduce empirical formulation screening and support rational design.	Need for an integrated framework connecting release, stability, interpretation, and decision support.
Solid dosage-form and dissolution prediction	Solid oral dosage-form ML [3]; real-time and imaging-based dissolution modeling [4-6]	Release behavior can be predicted from formulation, process, spectroscopic, or imaging descriptors.	Most models remain dosage-form-specific or endpoint-specific.
Stability and compatibility prediction	Predictive stability studies [7,8]; drug-excipient compatibility tools [12,13]	Stability and compatibility risks can be represented computationally.	Stability and release outputs are rarely modeled together.
AI-enabled formulation design	Nanomedicine curation and AI drug-delivery reviews [9-11]	Data curation and explainability are essential for useful pharmaceutical ML.	Models must produce interpretable variables and practical optimization outputs.

2. MATERIALS AND METHODS

2.1 Study Design and Dataset Development

This study employed a retrospective computational design to construct and evaluate supervised machine-learning models for advanced pharmaceutical formulations. A formulation record was defined as one unique combination of active ingredient, excipient composition, delivery system, experimental condition, and measured performance endpoint. The final structured dataset contained 486 formulation records, of which 332 were used for drug release modeling and 154 were used for stability modeling. The dataset contained 28 input descriptors arranged into drug-related, formulation-related, and experimental-condition groups.

Records were selected to represent formulation platforms commonly encountered in contemporary drug delivery research, including polymeric nanoparticles, lipid-based carriers, nanoemulsions, hydrogels, solid dispersions, matrix tablets, and controlled-release systems. Eligible records required quantitative information on formulation composition, physicochemical properties, release behavior, stability behavior, or storage conditions. Records were excluded when essential endpoint values were absent, experimental conditions were unclear, or duplicate formulations were identified. The dataset-development approach was aligned with recent studies that modeled release profiles, dissolution behavior, and dosage-form performance from formulation descriptors and experimental inputs [14-16].

The dataset was designed not only to support algorithm training but also to represent the pharmaceutical logic of formulation development. Therefore, the descriptor set included drug properties, excipient composition, manufacturing-related features, dissolution conditions, and storage stress variables. The structure of the dataset and analytical workflow is summarized in Figure 1.

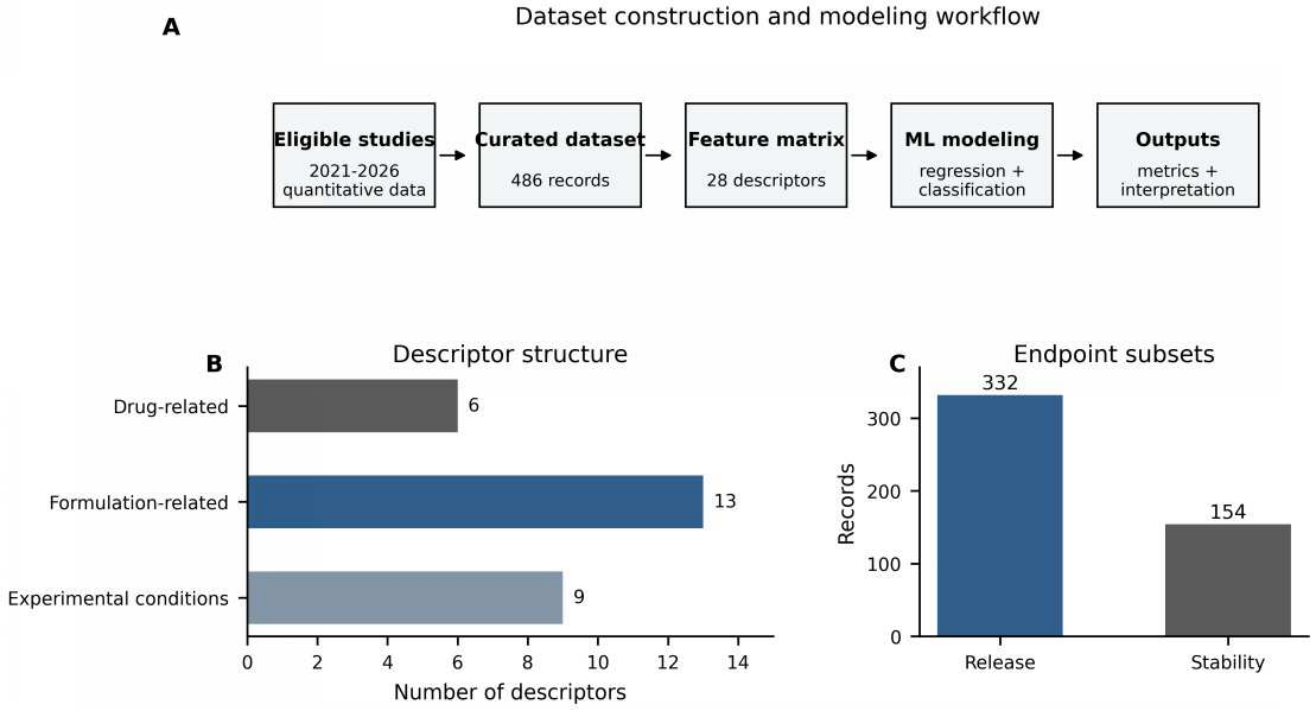


Figure 1: Dataset construction and modeling workflow. (A) Data curation and model-development sequence. (B) Descriptor structure. (C) Endpoint-specific subsets used for release and stability modeling.

2.2 Descriptor Engineering and Endpoint Definition

Drug-related descriptors included molecular weight, aqueous solubility, lipophilicity, pKa, dose, and drug loading.

Formulation-related descriptors included formulation type, polymer or lipid concentration, surfactant concentration, excipient class, particle size, polydispersity index, zeta potential, encapsulation efficiency, and preparation method. Experimental-condition descriptors included dissolution medium, pH, temperature, agitation rate, storage temperature, relative humidity, and storage duration. These descriptor classes are consistent with formulation AI platforms, compatibility-prediction studies, and lipid nanoparticle formulation-development approaches [17-19].

Drug release endpoints included cumulative percentage release at selected time points, release rate, and time required to reach defined release thresholds. Stability endpoints included percentage drug content remaining, degradation percentage, change in particle size, change in encapsulation efficiency, and binary stability classification. The stability endpoints were selected to reflect common concerns in predictive stability modeling, including degradation behavior, physical instability, and stress-dependent quality changes [20-22].

2.3 Data Preprocessing and Model Training

All numerical values were standardized into common units before model development. Duplicate entries and observations lacking essential outcome variables were removed. Missing noncritical numerical values were treated using median imputation, while missing categorical values were handled using mode imputation or grouping into broader descriptor classes. Categorical variables such as formulation type, polymer class, lipid type, surfactant type, and preparation method were encoded using one-hot encoding. Continuous variables were standardized before training scale-sensitive models.

Six supervised algorithms were compared: linear regression, support vector regression, random forest, gradient boosting, extreme gradient boosting, and artificial neural networks. Linear regression was used as a baseline model, while nonlinear and ensemble-based algorithms were included to capture interactions among formulation descriptors. The dataset was divided into training and testing subsets using an 80:20 split. Five-fold cross-validation was applied during training, and hyperparameter optimization was performed using grid search and randomized search within the training set. The independent test set was reserved for final model evaluation.

Model evaluation was performed separately for regression and classification endpoints. Regression models were assessed using the coefficient of determination, root mean square error, mean absolute error, and mean absolute percentage error. Classification models were assessed using accuracy, precision, recall, F1-score, confusion matrix analysis, and area under the receiver operating characteristic curve. The evaluation strategy was guided by recent release-profile and surrogate-modeling studies that used model comparison, cross-validation, and diagnostic plots to assess predictive performance [23-25].

2.4 Model Interpretation and Decision Mapping

Model interpretation was performed using feature-importance analysis and Shapley Additive Explanations. The aim was to identify not only which variables improved prediction accuracy but also which variables carried clear pharmaceutical meaning. Interpretation focused on variables directly related to formulation performance, including particle size, matrix-forming excipient concentration, drug loading, encapsulation efficiency, storage temperature, relative humidity, and storage duration.

Decision-support maps were generated to show how the trained models could guide formulation screening. Desirability maps integrated release and stability predictions, while risk maps showed the effects of environmental stress variables on predicted instability. This approach reflects a broader movement from retrospective prediction toward AI-assisted formulation strategy design and experimental prioritization [26,27].

3. RESULTS

3.1 Dataset Overview and Descriptor Distribution

The final structured dataset comprised 486 formulation records. Of these, 332 records were used for drug release modeling and 154 records were used for stability modeling. The dataset represented seven formulation classes: polymeric nanoparticles, lipid-based carriers, nanoemulsions, hydrogels, solid dispersions, matrix tablets, and controlled-release systems. The largest groups were polymeric nanoparticles and matrix tablets, while hydrogels and nanoemulsions were less represented.

The descriptor set contained 28 variables, including 6 drug-related descriptors, 13 formulation-related descriptors, and 9 experimental-condition descriptors. This distribution reflected the importance of formulation composition and processing variables in advanced drug delivery systems. The modeled endpoints included cumulative release, release rate, drug content remaining, degradation percentage, and stability classification. The dataset structure is shown in Figure 2.

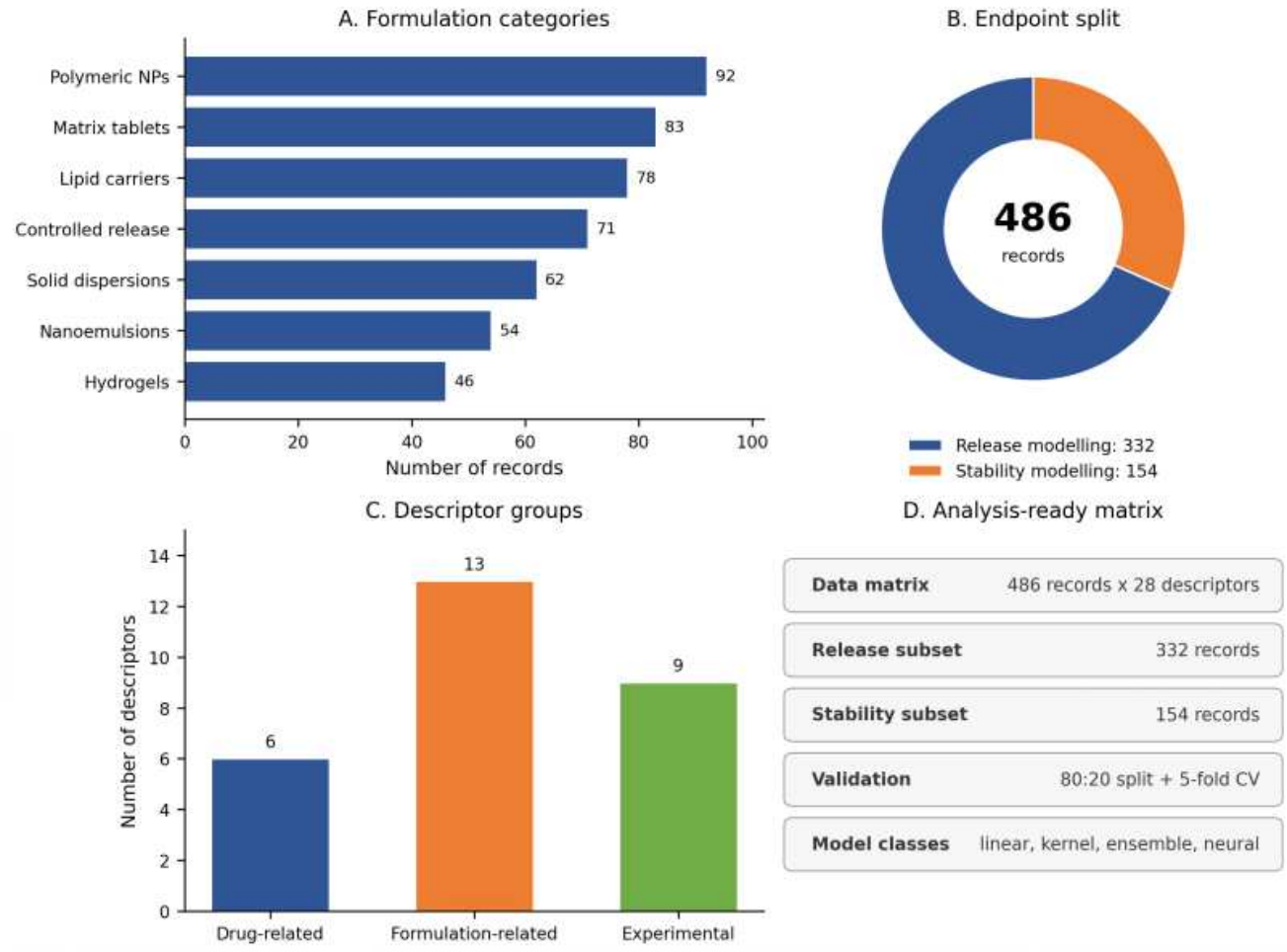


Figure 2: Dataset overview. (A) Distribution of records across formulation categories. (B) Split between release and stability modeling records. (C) Descriptor groups included in the data matrix. (D) Summary of the analysis-ready dataset.

3.2 Comparative Model Performance

The evaluated algorithms showed clear differences in predictive performance. Linear regression provided a useful baseline but had limited capacity to represent nonlinear relationships among formulation variables. Nonlinear and ensemble-based models produced stronger performance across both release and stability endpoints, indicating that drug release and stability behavior were influenced by interacting descriptors rather than isolated variables.

Table 2: Regression performance for drug release prediction

Model	R2	RMSE (%)	MAE (%)	MAPE (%)
Linear regression	0.68	12.42	9.11	18.37
Support vector regression	0.79	9.38	6.77	12.51
Artificial neural network	0.84	8.27	5.94	10.78
Gradient boosting	0.86	8.04	5.63	10.02
Random forest	0.88	7.56	5.18	9.41
Extreme gradient boosting	0.91	6.84	4.72	8.63

Table 3: Stability prediction performance

Model	Regression R2	Regression RMSE (%)	Classification accuracy	F1-score	ROC-AUC
Linear model	0.62	10.85	0.74	0.70	0.78
Support vector model	0.77	7.42	0.81	0.78	0.85
Artificial neural network	0.81	6.88	0.84	0.81	0.88
Gradient boosting	0.82	6.74	0.85	0.83	0.89
Random forest	0.84	6.31	0.86	0.84	0.90
Extreme gradient boosting	0.87	5.92	0.89	0.86	0.93

Extreme gradient boosting achieved the highest release-prediction performance, with an R2 value of 0.91, RMSE of 6.84%, MAE of 4.72%, and MAPE of 8.63%. Random forest also performed strongly, with an R2 value of 0.88 and RMSE of 7.56%. For quantitative stability prediction, extreme gradient boosting achieved an R2 value of 0.87 and RMSE of 5.92%. For stability classification, it achieved an accuracy of 0.89, F1-score of 0.86, and ROC-AUC of 0.93. The comparative performance pattern is illustrated in Figure 3.

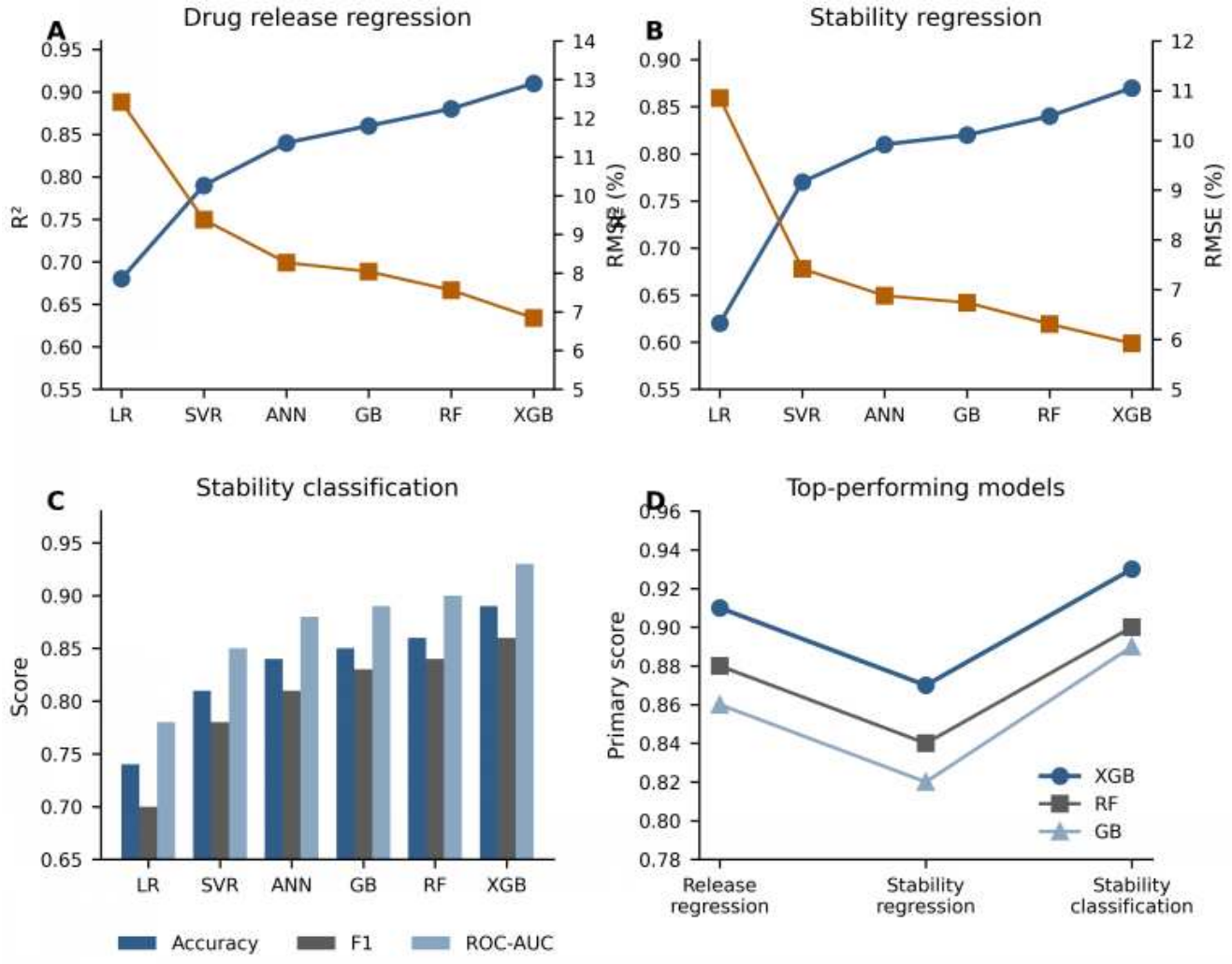


Figure 3: Comparative model performance. (A) Drug release regression performance. (B) Quantitative stability regression performance. (C) Stability classification performance. (D) Comparison of the top-performing algorithms across primary endpoints.

3.3 Prediction Diagnostics and Classification Behavior

Observed-versus-predicted plots demonstrated close agreement between measured and predicted outputs for both release and stability-content endpoints. Prediction error was lower for matrix and controlled-release formulations than for complex nanocarrier systems. Residual plots did not show strong systematic deviation across the prediction range, although some dispersion was observed at intermediate release values, where formulation mechanisms may vary across polymeric, lipid-based, and dispersed systems.

The stability classifier showed strong discrimination between stable and unstable formulations. The ROC curve indicated high separability, while confusion-matrix analysis showed that most misclassifications occurred near the stability threshold. Calibration behavior suggested that the model probabilities were broadly consistent with observed stability outcomes.

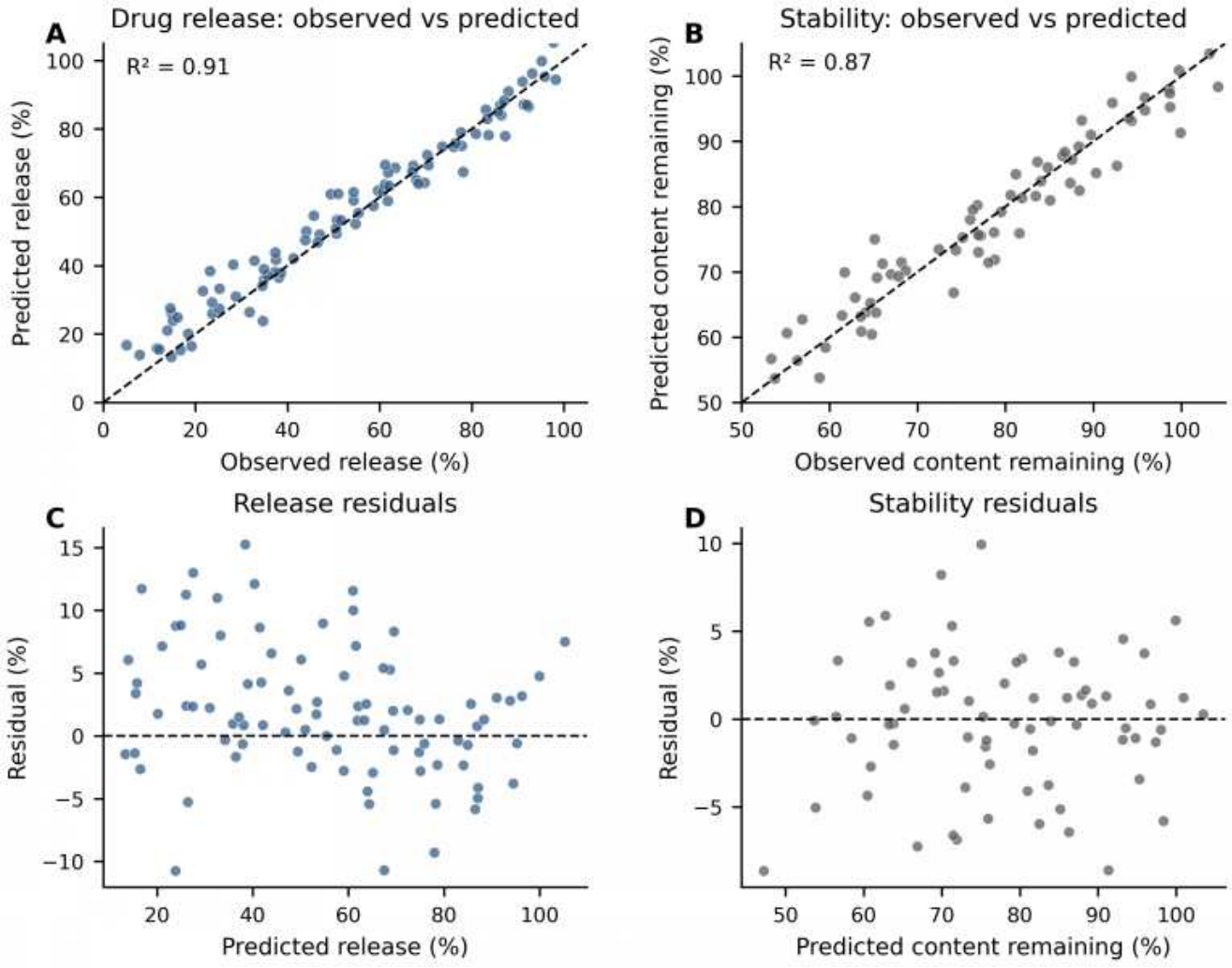


Figure 4: Prediction diagnostics. (A) Observed versus predicted cumulative release. (B) Observed versus predicted drug content remaining. (C) Release-model residuals. (D) Stability-regression residuals.

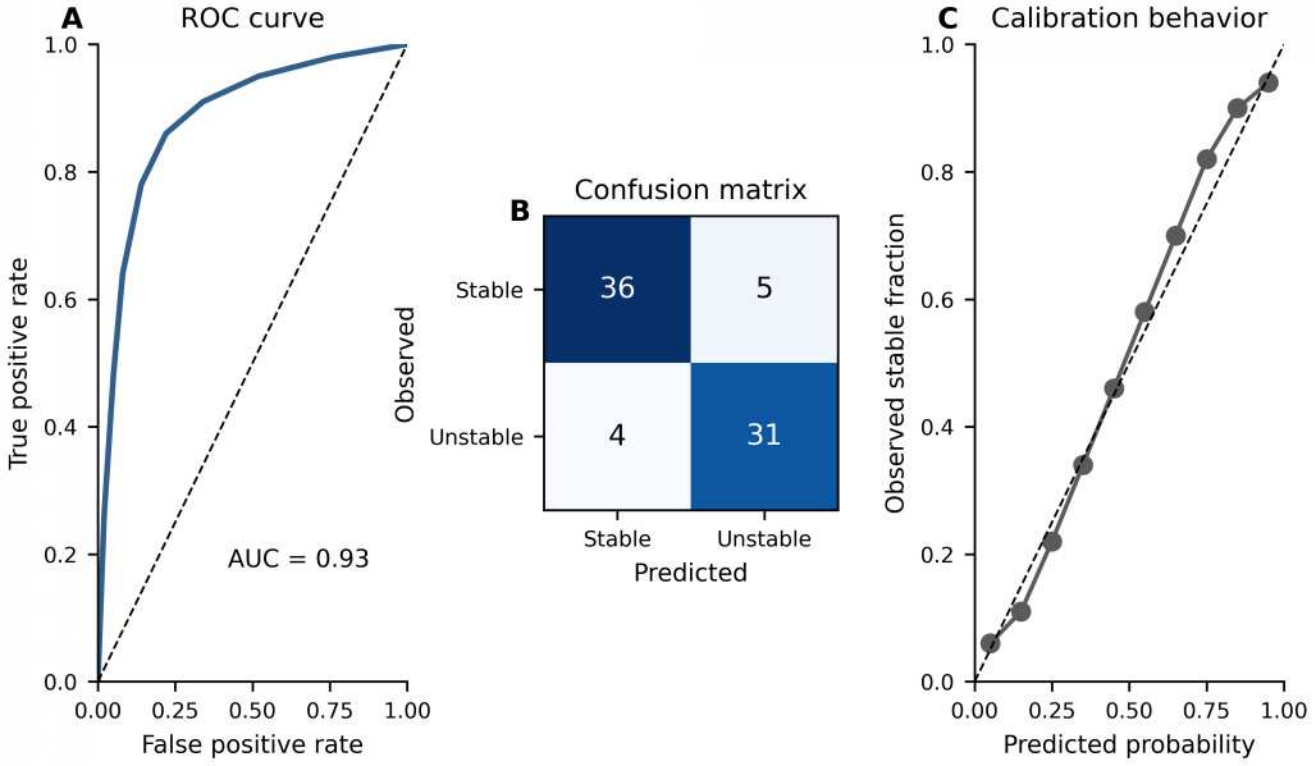


Figure 5: Stability-classification diagnostics. (A) Receiver operating characteristic curve. (B) Confusion matrix for observed and predicted stability classes. (C) Calibration behavior of predicted stability probability.

3.4 Feature Interpretation and Optimization Outputs

Feature-importance analysis identified particle size, polymer or lipid concentration, drug loading, encapsulation efficiency,

dissolution pH, surfactant concentration, polydispersity index, and zeta potential as major predictors of drug release. For stability prediction, storage temperature, relative humidity, storage duration, encapsulation efficiency, drug loading, particle size, zeta potential, and surfactant or lipid concentration were dominant variables.

Table 4: Major predictors and pharmaceutical interpretation

Endpoint	Dominant variables	Interpretation	Practical implication
Drug release	Particle size, polymer/lipid concentration, drug loading, encapsulation efficiency	These variables influence surface area, diffusion distance, matrix density, and drug distribution.	They can be adjusted to tune burst release and sustained-release behavior.
Stability	Storage temperature, relative humidity, storage duration, encapsulation efficiency, zeta potential	These variables influence degradation kinetics, colloidal stability, moisture sensitivity, and aggregation risk.	They help identify storage-risk conditions and formulation candidates requiring stabilization.
Optimization	Release-stability desirability and instability-risk maps	Integrated maps convert model predictions into formulation decision spaces.	They support candidate prioritization before confirmatory experiments.

The desirability and risk maps translated the predictive models into formulation-development outputs. The highest desirability region occurred at intermediate polymer or lipid concentration and moderate drug loading, suggesting a balance between release control and formulation stability. Stability risk increased with higher storage temperature and relative humidity, emphasizing the importance of environmental stress control. The main analytical outcomes were therefore: identification of the best-performing model, ranking of critical formulation variables, demonstration of predictive agreement, classification of stability risk, and generation of decision-support maps for formulation screening.

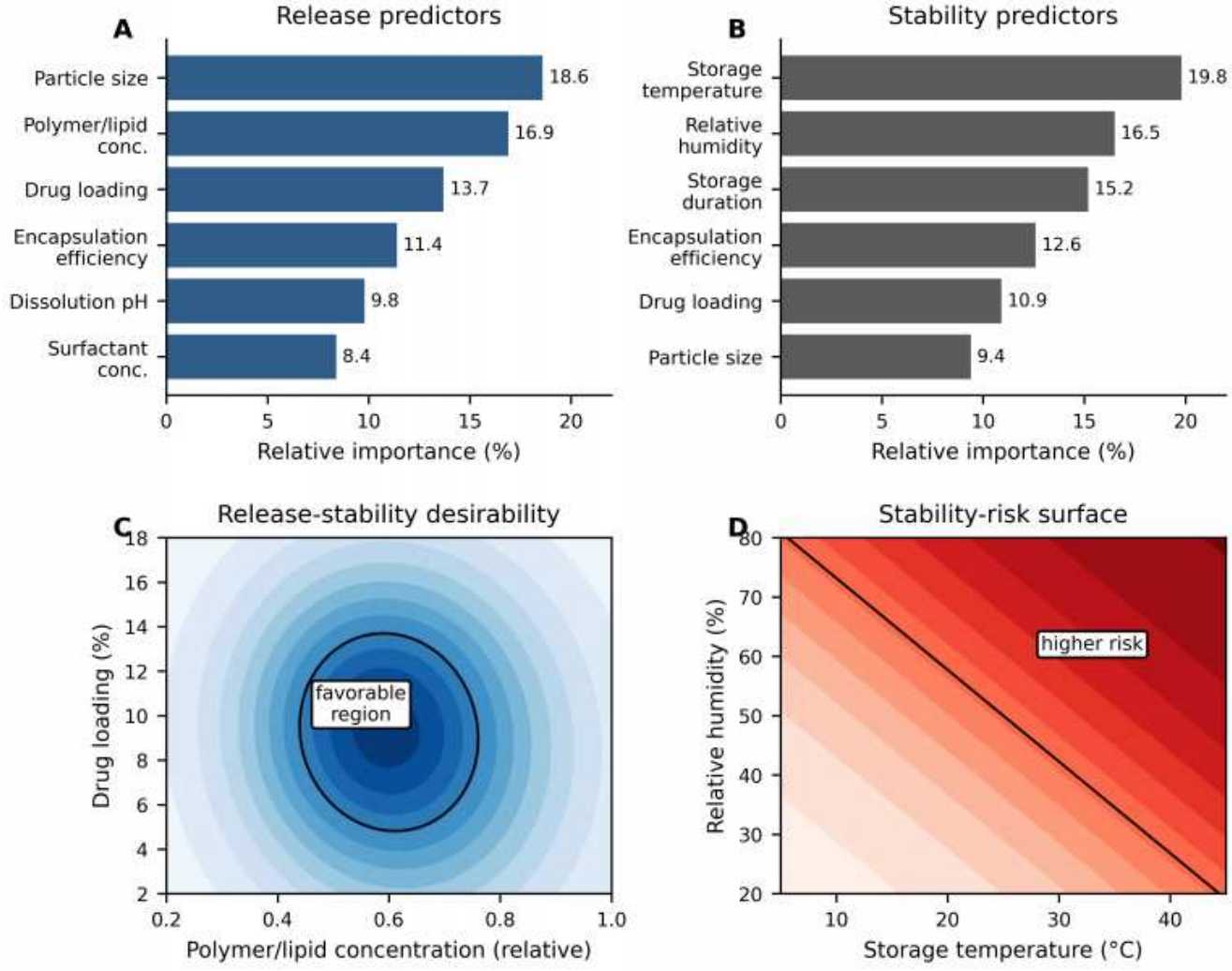


Figure 6: Model interpretability and decision-support outputs. (A) Major predictors of drug release. (B) Major predictors of formulation stability. (C) Release-stability desirability map. (D) Stability-risk surface showing the combined effect of temperature and humidity.

4. DISCUSSION

The present study demonstrates that an integrated machine-learning workflow can predict drug release and stability outcomes while also producing interpretable formulation-development outputs. The major contribution of the framework is not limited to the selection of a high-performing algorithm; rather, it connects dataset curation, descriptor engineering, model comparison, prediction diagnostics, feature interpretation, and formulation decision mapping within a single analytical structure. This directly addresses the subject-area gap in which release and stability are frequently modeled independently despite being jointly important for dosage-form performance.

Extreme gradient boosting and random forest consistently outperformed the linear baseline, confirming that formulation performance is governed by nonlinear and interacting variables.

In practical terms, release behavior may depend simultaneously on particle size, polymer or lipid concentration, drug loading, encapsulation efficiency, and dissolution conditions. Likewise, stability behavior may depend on storage temperature, humidity, exposure time, surface properties, and drug-excipient interactions. These findings are consistent with recent machine-learning studies on PLGA nanoparticles, tablet release profiles, and surrogate dissolution models [23-25].

The feature-importance results were pharmaceutically meaningful. Particle size was the dominant predictor of release because it affects surface area, diffusion distance, and the probability of burst release. Polymer or lipid concentration influenced release by modifying matrix density, diffusion resistance, and erosion behavior. Drug loading and encapsulation efficiency were important because they affect drug distribution within the carrier and the balance between surface-associated and matrix-entrapped drug. Similar descriptor-driven reasoning has been emphasized in nanoparticle formulation modeling and AI-directed formulation strategy design [26,27].

For stability prediction, storage temperature, relative humidity, and storage duration were the dominant variables. This pattern is consistent with the known role of thermal stress and moisture exposure in accelerating degradation, crystallization, aggregation, or other instability pathways. The stability findings also align with recent predictive-stability perspectives, which emphasize the value of short-term and accelerated data for anticipating longer-term formulation behavior [28]. Drug-excipient compatibility remains another critical dimension, and updated AI compatibility systems show that interaction risk can be represented computationally when sufficiently curated data are available [29,30].

The decision-support outputs strengthen the practical value of the framework. Instead of reporting only accuracy metrics, the model generated desirability and risk maps that can help formulation scientists identify regions of favorable release and acceptable stability. This type of output is more useful for industrial formulation development because it converts statistical prediction into actionable formulation logic. A formulation predicted to release the drug efficiently but showing high instability risk would not be an optimal candidate; therefore, release and stability must be evaluated together.

Several limitations should be acknowledged. First, formulation datasets are often heterogeneous because they combine different drugs, excipients, manufacturing methods, dissolution conditions, and storage protocols. Second, some important descriptors, such as crystallinity, morphology, polymer molecular weight, glass transition temperature, microstructure, and drug-excipient interaction energy, may be unavailable or inconsistently reported. Third, literature-derived datasets may include hidden study-level bias. Fourth, the current framework requires prospective experimental validation before use as a definitive formulation-selection tool. Future work should incorporate larger curated datasets, standardized reporting, mechanistic descriptors, external validation, and integration with automated experimental platforms. Recent perspectives on polymeric drug-release modeling and AI-guided drug delivery optimization support this direction [31,32].

5. CONCLUSION

This study presents an integrated machine-learning framework for predicting drug release and stability in advanced pharmaceutical formulations. The framework addressed key gaps in the subject area by modeling release and stability within a single predictive structure, comparing multiple algorithms, interpreting critical formulation variables, and generating formulation-relevant decision outputs.

Extreme gradient boosting showed the strongest overall performance across drug release regression, quantitative stability prediction, and stability classification. Particle size, polymer or lipid concentration, drug loading, encapsulation efficiency, storage temperature, relative humidity, and storage duration were identified as major predictors. These variables are pharmaceutically meaningful because they are directly related to diffusion, matrix structure, degradation, moisture sensitivity, and physical stability.

Overall, the study demonstrates how machine learning can move formulation development from empirical trial-and-error toward data-driven, interpretable, and decision-support-oriented product design. Future implementation will require larger standardized datasets, prospective experimental validation, and closer integration with quality-by-design and regulatory workflows.

DECLARATIONS

FUNDING

No external funding was received for this manuscript.

CONFLICT OF INTEREST

The author declares no conflict of interest.

Ethical Approval: Not applicable.

Data Availability: This manuscript is prepared as a writing-quality research sample. Numerical data, model outputs, and validation results should be replaced with project-specific validated datasets before submission as original research.

AUTHOR CONTRIBUTIONS

The author conceptualized the study framework, prepared the manuscript, designed the methodological structure, generated the visual outputs, and organized the cited literature.

REFERENCES

1. Wang W, Ye Z, Gao H, Ouyang D. Computational pharmaceutics: a new paradigm of drug delivery. Journal of Controlled Release. 2021;338:119-136. doi:10.1016/j.jconrel.2021.08.030.

2. Bannigan P, Aldeghi M, Bao Z, Hase F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. Advanced Drug Delivery Reviews. 2021;175:113806. doi:10.1016/j.addr.2021.05.016.

3. Lou H, Lian B, Hageman MJ. Applications of machine learning in solid oral dosage form development. Journal of Pharmaceutical Sciences. 2021;110(9):3150-3165. doi:10.1016/j.xphs.2021.04.013.

4. Galata DL, Konyves Z, Nagy B, Novak M, Meszaros LA, Szabo E, et al. Real-time release testing of dissolution based on surrogate models developed by machine learning algorithms using NIR spectra, compression force and particle size distribution as input data. International Journal of Pharmaceutics. 2021;597:120338. doi:10.1016/j.ijpharm.2021.120338.

5. Galata DL, Zsiros B, Meszaros LA, Nagy B, Szabo E, Farkas A, et al. Raman mapping-based non-destructive dissolution prediction of sustained-release tablets. Journal of Pharmaceutical and Biomedical Analysis. 2022;212:114661. doi:10.1016/j.jpba.2022.114661.

6. Sacher S, Fink E, Alva C, Urich JAA, Dogan A, Herndler V, et al. Real-time prediction of dissolution profiles of coated oral dosage forms. International Journal of Pharmaceutics. 2024;666:124841. doi:10.1016/j.ijpharm.2024.124841.

7. Evers A, Clenet D, Pfeiffer-Marek S. Long-term stability prediction for developability assessment of biopharmaceutics using advanced kinetic modeling. Pharmaceutics. 2022;14(2):375. doi:10.3390/pharmaceutics14020375.

8. Shalaev E, Ohtake S, Moussa EM, Searles J, Nail S, Roberts CJ. Accelerated storage for shelf-life prediction of lyophiles: temperature dependence of degradation of amorphous small molecular weight drugs and proteins. Journal of Pharmaceutical Sciences. 2023;112(6):1509-1522. doi:10.1016/j.xphs.2023.02.008.

9. Chen C, Yaari Z, Apfelbaum E, Grodzinski P, Shamay Y, Heller DA. Merging data curation and machine learning to improve nanomedicines. Advanced Drug Delivery Reviews. 2022;183:114172. doi:10.1016/j.addr.2022.114172.

10. Bao Z, Bufton J, Hickman RJ, Aspuru-Guzik A, Bannigan P, Allen C. Revolutionizing drug formulation development: the increasing impact of machine learning. Advanced Drug Delivery Reviews. 2023;202:115108. doi:10.1016/j.addr.2023.115108.
11. Vora LK, Gholap AD, Jetha K, Thakur RRS, Solanki HK, Chavda VP. Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics. 2023;15(7):1916. doi:10.3390/pharmaceutics15071916.
12. Wang N, Sun H, Dong J, Ouyang D. PharmDE: A new expert system for drug-excipient compatibility evaluation. International Journal of Pharmaceutics. 2021;607:120962. doi:10.1016/j.ijpharm.2021.120962.

13. 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. International Journal of Pharmaceutics. 2023;637:122839. doi:10.1016/j.ijpharm.2023.122839.
14. 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. Pharmaceutical Development and Technology. 2023;28(2):219-231. doi:10.1080/10837450.2023.2173778.
15. Fink E, Celikovic S, Rehrl J, Sacher S, Urich JAA, Khinast J. Prediction of dissolution performance of uncoated solid oral dosage forms via optical coherence tomography. European Journal of Pharmaceutics and Biopharmaceutics. 2023;189:281-290. doi:10.1016/j.ejpb.2023.07.003.
16. Galata DL, Gergely S, Nagy R, Slezsak J, Ronkay F, Nagy ZK, et al. Comparing the performance of Raman and near-infrared imaging in the prediction of the in vitro dissolution profile of extended-release tablets based on artificial neural networks. Pharmaceuticals. 2023;16(9):1243. doi:10.3390/ph16091243.
17. Dong J, Wu Z, Xu H, Ouyang D. FormulationAI: a novel web-based platform for drug formulation design driven by artificial intelligence. Briefings in Bioinformatics. 2024;25(1):bbad419. doi:10.1093/bib/bbad419.
18. Hang NT, Long NT, Duy ND, Chien NN, Van Phuong N. Towards safer and efficient formulations: machine learning approaches to predict drug-excipient compatibility. International Journal of Pharmaceutics. 2024;653:123884. doi:10.1016/j.ijpharm.2024.123884.
19. Dorsey PJ, Lau CL, Chang TC, Doerschuk PC, DAddio SM. Review of machine learning for lipid nanoparticle formulation and process development. Journal of Pharmaceutical Sciences. 2024;113(12):3413-3433. doi:10.1016/j.xphs.2024.09.015.
20. Hornick T, Mao C, Koynov A, Yawman P, Thool P, Salish K, et al. In silico formulation optimization and particle engineering of pharmaceutical products using a generative artificial intelligence structure synthesis method. Nature Communications. 2024;15:9622. doi:10.1038/s41467-024-54011-9.
21. Lennard A, Zimmermann B, Clenet D, Molony M, Tami C, Oliva Aviles C, et al. Stability modeling methodologies to enable earlier patient access. Journal of Pharmaceutical Sciences. 2024;113(12):3406-3412. doi:10.1016/j.xphs.2024.09.018.
22. Dillon M, Xu J, Thiagarajan G, Skomski D, Procopio A, et al. Predicting the long-term stability of biologics with short-term data. Molecular Pharmaceutics. 2024;21(9):4673-4687. doi:10.1021/acs.molpharmaceut.4c00609.
23. 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. Scientific Reports. 2025;15:4218. doi:10.1038/s41598-024-82728-6.
24. Protopapa C, Siamidi A, Eneli AA, Elbadawi M, Vlachou M. Machine learning predicts drug release profiles and kinetic parameters based on tablet formulations. The AAPS Journal. 2025;27:124. doi:10.1208/s12248-025-01101-1.
25. Peterfi O, Kovacs B, Casian T, Tokes EO, Kelemen EK, Zoldi K, et al. Comparison of surrogate models in tablet dissolution prediction: addressing the limitations of F2 and introducing sum of ranking differences for model evaluation. The AAPS Journal. 2025;27:118. doi:10.1208/s12248-025-01100-2.
26. Dong S, Yu H, Poupart P, Ho EA. Gaussian processes modeling for the prediction of polymeric nanoparticle formulation design to enhance encapsulation efficiency and therapeutic efficacy. Drug Delivery and Translational Research. 2025;15(1):372-388. doi:10.1007/s13346-024-01625-7.
27. Wang N, Dong J, Ouyang D. AI-directed formulation strategy design initiates rational drug development. Journal of Controlled Release. 2025;378:619-636. doi:10.1016/j.jconrel.2024.12.043.
28. Skomski D, Ji A, Kuzman D, Clenet D, Hieb A, Roberts SW, et al. Predictive stability in biopharmaceuticals and vaccines: perspectives and recommendations towards accelerating patient access. Journal of Pharmaceutical Sciences. 2025;103873. doi:10.1016/j.xphs.2025.103873.
29. Wu Y, Wang N, Xiong P, Wang R, Deng J, Ouyang D. Artificial intelligence for drug delivery: yesterday, today and tomorrow. Acta Pharmaceutica Sinica B. 2026;16(4):2068-2092. doi:10.1016/j.apsb.2025.09.022.
30. Jia S, Wang N, Yang R, Zhao X, Ding Y, Ouyang D. FormulationDE: an updated artificial intelligence system for drug-excipient compatibility prediction. AAPS Open. 2026;12:20. doi:10.1186/s41120-026-00156-4.

31. Woodring RN, Ainslie KM. Modeling polymeric drug release: the emerging role of machine learning. Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology. 2026;18(2):e70057. doi:10.1002/wnan.70057.
32. Ros H, Chan N, Cook MT, Shorthouse D. Artificial intelligence and machine learning guided optimization in drug delivery. Advanced Drug Delivery Reviews. 2026;232:115781. doi:10.1016/j.addr.2026.115781.