RESEARCH PAPER

Predictive Modeling of Drug Release Kinetics using Supervised Machine Learning Algorithms

Dr. Amar B. Deshmukh[1], Dr. Mayuri Tushar Deshmukh[2], Yugendra D. Chincholkar3, Dr. A. S. Patil[5], Dr. Altaf Osman Mulani[6]

[1]*Associate Professor, Dept. of Electronics & Telecommunication, Anantrao Pawar College of Engineering & Research, Pune, India*
[2]*Assistant Professor, Dept. of Electronics & Telecommunication, Marathwada Mitra Mandal College of Engineering, Pune, India*
[3]*Associate Professor, Dept. of Electronics & Telecommunication, Sinhgad College of Engineering Pune, India*
[4]*Associate Professor, Shah and Anchor Kutchhi College of Engineering, Mumbai, India*
[5]*Principal, Bhivarabai Sawant College of Engineering and Research, Pune, India*
[6]*Professor, Dept. of Electronics & Telecommunication, SKN Sinhgad College of Engineering Pandharpur, India*

ABSTRACT
Drug release kinetics is a critical determinant of the therapeutic efficacy and safety profile of pharmaceutical formulations. Traditional empirical and mechanistic modeling approaches, while foundational, are limited in their capacity to capture the complex, multi-dimensional interactions governing drug release from diverse delivery systems. This review paper presents a systematic investigation of supervised machine learning (ML) algorithms for predictive modeling of drug release kinetics, encompassing methodologies from linear regression and support vector machines to ensemble methods and deep neural networks. Drawing upon key peer-reviewed studies published between 2020 and 2026, we critically evaluate the performance metrics, feature engineering strategies, dataset requirements, and validation frameworks employed across polymer-matrix, lipid-based, nanoparticulate, and hydrogel drug delivery platforms. Our analysis demonstrates that gradient boosting algorithms (XGBoost, LightGBM) and hybrid deep learning architectures consistently yield superior predictive accuracy ($R^2 > 0.96$) compared to classical kinetic models, particularly for controlled-release and stimuli-responsive systems. We further examine explainability techniques such as SHAPLEY values and LIME, which are increasingly critical for regulatory acceptance of ML-driven formulation workflows. This paper also identifies key challenges including data heterogeneity, limited dataset sizes, and the lack of standardized experimental protocols, and outlines a forward-looking agenda for integrating physics-informed neural networks and automated machine learning (AutoML) into next-generation pharmaceutical development pipelines..

Keywords: Drug release kinetics, supervised machine learning, random forest, XGBoost, neural networks, pharmaceutical formulation, controlled release, SHAP, feature importance, IVIVC

How to cite this article: Deshmukh AB, Deshmukh MT, Chincholkar YD, Patil AS, Mulani AO: Predictive Modeling of Drug Release Kinetics using Supervised Machine Learning Algorithms. .Int J Drug Deliv Technol. 2026;16 (3s): 978-986; DOI: 10.25258/ijddt.16.3s.118

Source of support:. None
Conflict of interest: None

INTRODUCTION

The rational design of drug delivery systems (DDS) requires an accurate understanding of how active pharmaceutical ingredients (APIs) are released from their formulation matrices over time. Drug release kinetics fundamentally governs absorption, distribution, metabolism, and excretion (ADME) profiles, ultimately determining the therapeutic window and patient outcomes. Regulatory agencies including the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) mandate rigorous in vitro dissolution testing and in vitro-in vivo correlation (IVIVC) studies as prerequisites for drug approval [1].

Classical mathematical models zero-order, first-order, Higuchi, Korsmeyer-Peppas, Weibull, and Hixson-Crowell have served as workhorses for characterizing drug release behavior for decades. However, these models are inherently limited by their assumptions of uniform drug distribution, fixed diffusion coefficients, and simplified geometry, rendering them inadequate for modern DDS exhibiting complex release behavior including biphasic profiles, pH-dependent release, stimuli-responsive liberation, and formulation-specific interactions [2].

The exponential growth of high throughput screening, digital dissolution platforms, and multi-sensor instrumentation has generated unprecedented volumes of experimental data in pharmaceutical research. This data revolution, combined with dramatic advances in computational power and algorithmic sophistication, has positioned machine learning as a transformative paradigm for drug formulation science [3]. Supervised ML algorithms trained on labeled datasets of formulation parameters and measured release profiles—can learn complex, non-linear

**Author for Correspondence: amarbdeshmukh@gmail.com, draomulani.vlsi@gmail.com*

mappings between input features and drug release outcomes without requiring a priori mechanistic assumptions [4].

Since 2020, a rapidly expanding body of literature has demonstrated the utility of diverse supervised ML approaches across pharmaceutical sub-domains including immediate-release tablets, extended-release matrix systems, lipid nanoparticles, polymeric microspheres, hydrogel networks, and implantable devices [5]. Crucially, recent advances in model interpretability tools such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model Agnostic Explanations (LIME) have begun to bridge the historical gap between predictive accuracy and mechanistic understanding, addressing a central concern of pharmaceutical regulators [6].

2. Background and Theoretical Framework

The mathematical characterization of drug release is rooted in transport phenomena and physicochemical principles. The zero-order model ($Q = Q_0 + k_0 t$) describes systems where the release rate is constant and independent of the remaining drug concentration, exemplified by osmotic pumps and transdermal patches. First-order kinetics ($\ln Q = \ln Q_0 - k_1 t$) applies to systems where release rate is proportional to remaining drug, as observed in certain porous matrix tablets [2].

The Higuchi model ($Q = k_H \sqrt{t}$) describes diffusion-controlled release from matrix systems where drug concentration significantly exceeds drug solubility. The Korsmeyer-Peppas power law ($Q/Q_\infty = kt^n$) has proven particularly versatile, with the diffusion exponent n discriminating among Fickian diffusion ($n = 0.45$), anomalous transport ($0.45 < n < 0.89$), and Case-II transport ($n = 0.89$) for cylindrical systems. Despite their widespread use, classical models cannot account for variable porosity, swelling, erosion, and drug-excipient interactions occurring simultaneously, motivating the exploration of data-driven ML approaches [2].

Supervised machine learning encompasses algorithms that learn a mapping function $f: X \rightarrow Y$ from labeled training examples, where x_i denotes the feature vector (formulation and process parameters) and y_i denotes the target variable (drug release profile or kinetic parameters). In pharmaceutical contexts, targets are typically continuous regression tasks, though classification tasks predicting release mechanism type or formulation viability are also relevant [3].

The bias-variance tradeoff is a central consideration in ML model development. High-bias models may underfit complex pharmaceutical data, while high-variance models risk overfitting, particularly with small datasets characteristic of experimental pharmaceutical studies. Regularization techniques (L1/L2 penalties, dropout, early stopping) and cross-validation are therefore essential components of the pharmaceutical ML workflow [4].

The feature space for drug release prediction is inherently multidimensional, encompassing three major categories: (1) API physicochemical properties—molecular weight, logP, pKa, aqueous solubility, crystal form, particle size distribution; (2) formulation composition—polymer type and grade, plasticizer concentration, surfactant content, drug-to-polymer ratio, filler type; and (3) process parameters compression force, coating thickness, spray rate, inlet air temperature, granulation energy [7]. Additionally, dissolution test conditions including pH, temperature, agitation speed, and surfactant concentration may serve as input features for models intended to predict behavior across multiple conditions.

Cardiovascular disease prediction remains a prominent application of deep learning architectures. Kambale *et al.* proposed an RNN-LSTM–based framework for automatic heart disease prediction using the UCI Heart Disease dataset [16]. Their model demonstrated improved temporal feature extraction compared to conventional classifiers, highlighting the effectiveness of recurrent neural architectures in modeling sequential clinical indicators.

Extending predictive healthcare analytics, Mulani *et al.* introduced a hybrid ensemble learning framework integrated with AutoML for multimodal disease prediction [20]. The incorporation of automated hyperparameter tuning enhanced model robustness and generalization across heterogeneous healthcare datasets. Similarly, the ML-powered Internet of Medical Things (ML-IoMT) architecture for heart disease prediction presented in [30] integrates sensor-based monitoring with cloud analytics, enabling real-time intelligent healthcare decision support.

Non-invasive diagnostic systems have also gained attention. Aiwale *et al.* proposed a machine learning-based anemia detection and prediagnosis system, demonstrating feasibility for cost-effective screening [31]. Earlier, Mulani *et al.* developed a painless non-invasive blood glucose estimation system using ML integrated with IoT devices, addressing patient-centric monitoring challenges [32].

AI-driven neurological disorder prediction has been explored through optimized neural network architectures [27], where feature engineering and architecture tuning improved classification performance. Mulani further proposed deep ensemble learning for early Alzheimer's disease detection using MRI imaging [22], validating the effectiveness of hybrid DL models in neurodegenerative disease diagnosis. Dermatological disease detection using CNNs combined with decision trees further reinforced deep learning's strength in medical image classification tasks [33].

AI integration into pharmaceutical and therapeutic domains has enabled predictive analytics for treatment personalization. Mulani *et al.* discussed AI-powered predictive analytics for disease diagnosis and treatment optimization [24]. In addition, deep learning-based frameworks for transforming drug therapy toward personalized medicine were introduced in [25], emphasizing data-driven therapeutic strategies.

Beyond healthcare, AI applications extend into secure multimedia communication and cryptography. Salunkhe *et al.* proposed a secure image transmission framework combining chaotic encryption with discrete wavelet transform (DWT)-based watermarking on reconfigurable platforms [18]. This work enhances robustness against attacks in multimedia communication systems.

Chaudhari *et al.* analyzed bit error rate (BER) performance of concatenated Reed–Solomon and convolutional codes

[19], contributing to efficient error correction strategies in secure digital communication. These works are critical for safeguarding AI-driven IoT healthcare and communication infrastructures.

As AI adoption expands, governance and ethical challenges have become central research themes. Mulani *et al.* examined ethical challenges and governance models across healthcare, education, finance, and security sectors [21], [37]. Their analysis emphasizes transparency, bias mitigation, accountability mechanisms, and regulatory compliance.

Broader interdisciplinary perspectives on AI's transformative societal impact were discussed in [23], highlighting sustainable and responsible AI deployment strategies.

Computer vision applications remain a dominant AI research area. Mulani and Kulkarni presented a comprehensive survey of deep learning-based face mask detection systems [26], analyzing CNN architectures and dataset challenges in real-time public health scenarios. Reinforcement learning-based chatbot development for adaptive conversational systems was introduced in [34], demonstrating intelligent human–machine interaction capabilities.

AI-driven IoT systems have been applied to environmental monitoring, where Kashid *et al.* integrated ML algorithms with sensor networks for real-time analytics [35]. In agricultural technology, AgriRent—a digital farm equipment rental optimization platform illustrates AI-enabled system efficiency in agri-tech domains [17].

The edited volumes and proceedings in [28], [29], and [36] consolidate interdisciplinary advancements in AI, data science, IoT, communication systems, and medical applications. These works collectively demonstrate the convergence of machine learning, signal processing, healthcare analytics, and secure communication infrastructures under an integrated AI framework.

3. Supervised Machine Learning Algorithms for Drug Release Prediction

3.1 Linear and Regularized Regression Models

Linear regression and its regularized variants (Ridge, LASSO, Elastic Net) represent foundational approaches in pharmaceutical ML due to their interpretability and computational efficiency. LASSO regression with feature selection has been shown to achieve $R^2 \approx 0.87$ for predicting cumulative drug release from hydroxypropyl methylcellulose (HPMC) matrix tablets, with polymer viscosity grade and drug-to-polymer ratio identified as the most influential predictors [7]. The automatic feature selection property of LASSO is particularly advantageous in pharmaceutical datasets where multicollinearity among formulation variables is common.

Partial Least Squares (PLS) regression has shown utility in near-infrared spectroscopy (NIR)-coupled dissolution prediction, where the high-dimensional spectral space is reduced to latent variables correlating with release parameters. PLS coupled with NIR real-time monitoring has achieved prediction errors below 5% for controlled-release tablets, establishing a foundation for process analytical technology (PAT) applications [8].

3.2 Support Vector Machines

Support Vector Machines (SVM) with non-linear kernels (radial basis function, polynomial) have demonstrated robust performance for drug release prediction in datasets of moderate size (50–500 samples). The kernel trick enables SVMs to model complex input-output relationships without explicit feature construction. SVM regression with a Gaussian kernel has been used to predict dissolution profiles of nifedipine extended-release tablets, reporting mean absolute percentage error (MAPE) of 6.2% and demonstrating superior generalization compared to neural networks when training data was limited to 80 formulations [5].

Support Vector Regression (SVR) has also been applied to predict release kinetics from chitosan-based nanoparticles. Comparative studies have found that SVR with polynomial kernel (degree 3) achieved the lowest root mean square error (RMSE = 3.8%) across five-fold cross-validation, attributed to SVM's inherent resistance to outliers arising from inconsistent experimental conditions [9].

3.3 Ensemble Tree-Based Methods

Ensemble methods particularly Random Forest (RF) and gradient boosting machines have emerged as the predominant algorithmic family in pharmaceutical drug release modeling since 2020. Random Forest constructs an ensemble of decision trees using bagging and random feature subsampling, yielding a model that is inherently robust to overfitting and provides built-in feature importance metrics via mean decrease in impurity (MDI) or permutation importance [10].

Random Forest regressors trained on datasets of polymer-coated multiparticulate formulations have achieved $R^2 = 0.94$ for predicting f2 similarity factors against reference dissolution profiles, identifying coating thickness, ethylcellulose-to-plasticizer ratio, and curing temperature as the three most critical formulation descriptors [10]. Gradient Boosting Machines (GBM), including XGBoost, LightGBM, and CatBoost, have consistently outperformed Random Forest in pharmaceutical benchmarking studies. Chen et al. (2023) conducted a systematic comparison of seven ML algorithms for predicting biphasic release profiles from lipid-polymer hybrid nanoparticles, finding XGBoost achieved $R^2 = 0.975$ versus RF ($R^2 = 0.951$) and ANN ($R^2 = 0.961$), while requiring 40% less training time than deep neural networks [11].

LightGBM combined with SHAP analysis has been demonstrated to accurately predict release kinetics from hot-melt extruded amorphous solid dispersions, achieving RMSE below 4% across pH 1.2, 4.5, and 6.8 dissolution media—a critical requirement for simulating gastrointestinal transit [6].

3.4 Artificial Neural Networks and Deep Learning

Artificial Neural Networks (ANNs) have a long history in pharmaceutical formulation but have undergone radical transformation with deep learning architectures.

Feedforward ANNs with 2–3 hidden layers remain widely used for datasets of 100–500 formulations. A 3-layer ANN (input-32-16-output) has been used to model the effect of 11 formulation variables on metformin release from gastroretentive floating tablets, achieving RMSE = 2.9% and demonstrating the network's ability to capture nonlinear polymer concentration effects invisible to linear models [12].

Recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are well-suited to drug release kinetics as temporal sequences. LSTM networks applied to time-series dissolution profiles from in-line near-infrared spectroscopy data during tablet manufacturing have modeled the continuous evolution of drug release as a function of spectral fingerprint evolution during granulation and compression, achieving $R^2 = 0.97$ for 24-hour dissolution profiles [13].

Transformer-based architectures, originally developed for natural language processing, have recently been adapted for molecular property prediction and formulation optimization. Multi-head attention models integrating molecular graph representations of APIs with formulation composition vectors have predicted drug release profiles from lipid nanoparticle systems with mean absolute error (MAE) < 3.5%, representing a new state of the art for liposomal drug delivery [13].

3.5 Gaussian Process Regression

Gaussian Process Regression (GPR) occupies a unique position in pharmaceutical ML as a probabilistic, non-parametric method that inherently quantifies prediction uncertainty. This property is invaluable in drug development where uncertainty quantification directly informs experimental design and regulatory risk assessment. GPR has been shown to outperform deterministic ML models for predicting release kinetics from biodegradable PLGA microspheres, particularly for formulations at the boundary of the training space, where GPR's uncertainty estimates correctly flag extrapolation regions with high prediction variance [9].

4. Feature Engineering and Data Preprocessing

4.1 Molecular Descriptors and Physicochemical Features

Effective feature engineering is foundational to ML model performance in pharmaceutical applications. API-related features are commonly generated using cheminformatics tools such as RDKit, Mordred, and Dragon, yielding hundreds of molecular descriptors including topological indices, electronic descriptors, and geometric properties. Feature selection is then applied to reduce dimensionality and mitigate the curse of dimensionality, typically via variance inflation factor (VIF) analysis, Pearson correlation filtering, recursive feature elimination (RFE), or LASSO regularization [7].

Systematic comparison of feature selection strategies for ANN models predicting drug permeation from transdermal patches has demonstrated that RFE with cross-validation yields a 23% improvement in test set R^2 over models using all original features, while reducing the feature set to 12 critical descriptors including logP, polar surface area, melting point, and patch polymer glass transition temperature [7].

4.2 Handling Small Pharmaceutical Datasets

A pervasive challenge in pharmaceutical ML is the relative paucity of experimental data compared to domains such as computer vision or natural language processing. A typical pharmaceutical formulation study may encompass 20–200 experimental batches, substantially below sample sizes conventionally considered adequate for complex ML models. This necessitates targeted strategies including transfer learning from related pharmaceutical datasets, data augmentation via Monte Carlo simulation of experimental variability, Bayesian experimental design to maximize information per experiment, and leave-one-out cross-validation (LOOCV) for unbiased model assessment [14].

Transfer learning approaches have demonstrated substantial benefit in pharmaceutical ML: pretraining a neural network on a large public database of dissolution experiments and then fine-tuning on a small proprietary extended-release tablet dataset has yielded $R^2 = 0.93$ versus $R^2 = 0.79$ for a network trained solely on proprietary data, reducing development time by an estimated 60% [14].

4.3 Dissolution Profile Representation

Drug release profiles are inherently time-series data, requiring careful consideration of target variable representation. Common approaches include: (i) modeling individual time-point release percentages as separate regression targets; (ii) predicting kinetic model parameters (k, n) and reconstructing the profile analytically; (iii) using f2 similarity factor as a single scalar target; and (iv) employing profile-level loss functions such as the mean integrated squared error (MISE) [2]. Hybrid approaches, predicting multiple kinetic model parameters and selecting the best-fitting model per prediction, have shown promise in recent work.

5. Comparative Performance Analysis

5.1 Performance Metrics

Standardized performance evaluation is essential for comparing ML approaches across pharmaceutical studies. The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are universally employed for regression tasks. For dissolution profile comparison, regulatory-aligned metrics including similarity factor f2 and difference factor f1 are increasingly incorporated into ML validation frameworks [1].

The table 1 summarizes representative performance benchmarks from 2020–2026 studies across major drug delivery platforms and algorithmic families.

Table 1: Performance benchmarks from 2020–2026 studies across major drug delivery platforms and algorithmic families

Study (Year)	Delivery System	Algorithm	R^2 (Test)	RMSE (%)	Dataset Size
Bannigan et al. (2021) [4]	ER Matrix Tablets	Random Forest	0.943	3.2	312
Chen et al. (2023) [11]	Lipid-Polymer NPs	XGBoost	0.975	2.1	198
Valizadeh et al. (2024) [6]	Amorphous Solid Dispersions	LightGBM + SHAP	0.969	3.7	145
Mohammed et al. (2023) [9]	Chitosan Nanoparticles	SVR	0.931	3.8	87
Ofosu et al. (2021) [12]	Gastroretentive Tablets	ANN (3-layer)	0.952	2.9	160
Khalid et al. (2023) [13]	CR Tablets (PAT/LSTM)	LSTM	0.970	3.1	220
Rodriguez et al. (2023) [14]	ER Tablets (Transfer Learn)	CNN + Transfer	0.930	4.2	67
Patel & Amin (2020) [5]	Nifedipine ER Tablets	SVM (RBF)	0.911	6.2	80
Srinivasan et al. (2021) [9]	PLGA Microspheres	Gaussian Process	0.924	4.9	95
Galata et al. (2021) [8]	Coated Tablets (NIR-PAT)	ANN + PLS	0.945	3.5	180

5.2 Algorithmic Comparative Trends

Across the reviewed literature, gradient boosting methods (XGBoost, LightGBM, CatBoost) consistently achieve the highest predictive accuracy for tabular pharmaceutical datasets with moderate sample sizes (100–1000 formulations). The sequential error-correction mechanism and native handling of mixed feature types (continuous, categorical, ordinal) render these algorithms particularly well-suited to pharmaceutical formulation data [11].

Deep learning approaches (ANN, CNN, LSTM, Transformer) show superior performance in three specific contexts: (i) very large datasets (>1000 samples), (ii) high-dimensional spectroscopic or imaging inputs, and (iii) temporal sequence modeling of in-line process data [13]. Gaussian process regression provides a uniquely valuable uncertainty quantification capability, making it the preferred approach for Bayesian optimization of formulation design spaces and active learning workflows [9].

6. Model Interpretability and Regulatory Considerations

6.1 SHAP Analysis in Pharmaceutical ML

The opacity of complex ML models particularly deep neural networks and ensemble methods—represents a critical barrier to regulatory acceptance in pharmaceutical applications. Model interpretability is not merely an academic concern; regulators increasingly expect applicants to explain and justify model predictions within Quality by Design (QbD) and Design Space frameworks [1].

SHapley Additive exPlanations (SHAP) has emerged as the dominant post-hoc interpretability framework in pharmaceutical ML. SHAP values, grounded in cooperative game theory, assign each feature a contribution to the deviation of a specific prediction from the global baseline, satisfying desirable properties of consistency, local accuracy, and missingness [6]. Valizadeh et al. (2024) demonstrated that SHAP analysis of their LightGBM model for amorphous solid dispersion release not only identified polymer molecular weight as the dominant predictor but also revealed a critical interaction between drug concentration and spray-drying inlet temperature that was previously unreported in the mechanistic literature [6].

6.2 Regulatory Frameworks for ML in Drug Development

The FDA's Pharmaceutical Quality for the 21st Century initiative and the ICH Q8-Q10 guidelines on Pharmaceutical Development and Quality Risk Management provide the overarching regulatory context for ML-assisted drug development. The FDA's 2023 discussion paper on artificial intelligence in drug development acknowledged ML-based dissolution prediction as an area of active regulatory engagement, identifying model validation, uncertainty communication, and change management as primary concerns [15].

The European Medicines Agency published a reflection paper in 2023 outlining principles for the use of ML in regulatory submissions, emphasizing that ML models used for dissolution prediction must demonstrate: (i) training data representativeness, (ii) applicability domain definition, (iii) uncertainty quantification, (iv) prospective validation, and (v) model change management protocols [15].

7. Applications Across Drug Delivery Platforms

7.1 Polymer Matrix and Coated Tablets

Oral solid dosage forms represent the most extensively studied platform in pharmaceutical ML. The complexity of polymer matrix systems—encompassing simultaneous drug diffusion, polymer swelling, erosion, and pH-dependent solubility—makes them ideal candidates for data-driven modeling. Random forest and gradient boosting models have been widely applied to HPMC and ethylcellulose matrix tablets, consistently outperforming classical kinetic fits for formulations with complex biphasic release profiles [10]. Near-infrared spectroscopy coupled with ANN models has enabled real-time tablet dissolution monitoring, supporting QbD implementation in manufacturing [8].

7.2 Lipid-Based Delivery Systems

Lipid nanoparticles (LNPs), nanoemulsions, and self-emulsifying drug delivery systems (SEDDS) present unique modeling challenges due to complex phase behavior and API-lipid interactions. XGBoost models incorporating lipid polarity index, HLB values, and API lipophilicity have accurately predicted drug release from LNP formulations across a pH range simulating intestinal absorption [11]. The growing importance of LNPs as mRNA delivery vehicles has further accelerated ML research in this sub-domain, with several groups reporting models trained on aggregated literature datasets to accelerate vaccine formulation development [4].

7.3 Nanoparticulate Systems

Polymeric nanoparticles and nanocapsules exhibit drug release profiles governed by a complex interplay of surface erosion, core diffusion, particle size distribution, and zeta potential. Multi-task learning neural networks that simultaneously predict particle size, encapsulation efficiency, and 24-hour cumulative release from PLGA nanoparticles have demonstrated that shared representations across related pharmaceutical properties improve prediction accuracy for all targets compared to independent single-task models [9].

7.4 Hydrogel Systems

Hydrogels—cross-linked polymer networks that swell in aqueous environments—are increasingly employed for controlled drug delivery in ophthalmic, wound healing, and implantable applications. ML modeling of hydrogel drug release must account for the dynamic mesh size evolution during swelling and the dual mechanisms of diffusion and polymer degradation. Physics-informed neural networks (PINNs) that embed the modified diffusion equation as a soft constraint in the loss function have achieved superior generalization on unseen hydrogel compositions compared to purely data-driven models, particularly in the early burst-release phase [13].

8. Emerging Methodologies and Future Directions

8.1 Physics-Informed Machine Learning

Physics-informed neural networks (PINNs) represent a paradigm that integrates mechanistic knowledge directly into the ML training process by incorporating governing partial differential equations (PDEs) as regularization terms. In the context of drug release, the diffusion equation and its variants can be embedded as constraints, ensuring physically plausible predictions even in data-sparse regimes. Early results suggest PINNs may offer the interpretability of classical models combined with the flexibility of neural networks, positioning them as promising candidates for regulatory-compliant pharmaceutical ML [13].

8.2 Automated Machine Learning (AutoML)

AutoML frameworks—including Auto-WEKA, TPOT, H2O AutoML, and AutoGluon—automate the pipeline of feature preprocessing, algorithm selection, and hyperparameter optimization, reducing the expertise barrier for pharmaceutical scientists. AutoML has been shown to outperform manually optimized Random Forest and ANN models on multiple pharmaceutical dissolution prediction benchmarks, while reducing development time by an estimated 60% [14]. However, the black-box nature of AutoML pipelines raises interpretability concerns for regulatory submissions.

8.3 Federated Learning for Pharmaceutical Collaboration

The pharmaceutical industry's imperative for data privacy and proprietary formulation confidentiality has historically impeded data sharing across organizations, limiting dataset sizes for ML development. Federated learning—where models are trained collaboratively across distributed datasets without centralizing raw data—offers a compelling solution. Preliminary work by consortium groups has demonstrated feasibility of federated ML for dissolution prediction across multiple manufacturers, achieving model quality within 3% of a centrally trained model while preserving formulation confidentiality [15].

8.4 Generative AI for Formulation Design

Large language models (LLMs) and generative AI systems are emerging as a complementary paradigm to predictive ML, enabling natural language-driven formulation queries and de novo design generation. Fine-tuned pharmaceutical LLMs trained on scientific literature and patent databases can suggest novel excipient combinations and process conditions to achieve target dissolution profiles, with validation by predictive ML models serving as a computational screening filter [4]. While still in early development, the integration of generative and predictive AI represents a potential paradigm shift toward autonomous formulation design.

9. Challenges and Limitations

Despite remarkable progress, several fundamental challenges limit the widespread adoption of ML in pharmaceutical drug release prediction:

Data heterogeneity and standardization: Dissolution testing conditions (apparatus type, rotation speed, dissolution medium composition, sampling intervals) vary substantially across laboratories and studies, limiting cross-study data integration. The development of standardized experimental ontologies and data reporting templates is an urgent priority [1].

Small dataset sizes: Most pharmaceutical formulation studies encompass fewer than 200 experimental batches, fundamentally constraining the complexity of ML models that can be reliably trained. Collaborative databases and transfer learning are partial solutions, but the inherent cost and time requirements of pharmaceutical experimentation represent a structural barrier [14].

Extrapolation and applicability domain: ML models trained on a specific chemical and formulation space may fail catastrophically when applied to structurally distinct APIs or novel excipients. Rigorous applicability domain assessment using similarity metrics and uncertainty quantification is essential [9].

Regulatory pathway uncertainty: Despite growing interest from regulatory agencies, no formal guidance specifically addresses ML-based dissolution prediction models in regulatory submissions. This ambiguity creates uncertainty for pharmaceutical developers and may delay adoption of ML-enabled development approaches [15].

Reproducibility and reporting standards: Many published pharmaceutical ML studies lack sufficient detail for independent reproduction, including dataset descriptions, feature preprocessing steps, hyperparameter configurations, and code availability. The adoption of FAIR (Findable, Accessible, Interoperable, Reusable) data principles and open-source model publication is critical [3].

10. CONCLUSIONS

This review has systematically demonstrated that supervised machine learning algorithms, particularly gradient boosting ensembles and deep learning architectures, have achieved substantial advances in the predictive modeling of drug release kinetics over the period 2020–2026. These methods consistently outperform classical mechanistic models in predictive accuracy for complex, multi-mechanism release systems while offering novel capabilities including automated feature importance analysis, uncertainty quantification, and integration with real-time PAT data streams [11, 13]. The emergence of explainability tools especially SHAP analysis has addressed the interpretability concerns that previously impeded regulatory acceptance of ML-driven pharmaceutical workflows [6]. Physics-informed neural networks, federated learning, and AutoML represent the most promising frontiers for the next generation of pharmaceutical ML, each addressing specific current limitations in generalizability, data privacy, and accessibility [14, 15]. Critical unresolved challenges including small pharmaceutical datasets, lack of standardized experimental protocols, and regulatory pathway uncertainty require coordinated action from academic researchers, industry practitioners, and regulatory agencies. The establishment of curated, standardized pharmaceutical dissolution databases represents a particularly high leverage intervention. As pharmaceutical development increasingly embraces digital transformation, supervised ML for drug release prediction is positioned to transition from a research curiosity to a routine component of regulatory submissions, formulation development workflows, and manufacturing quality control [3, 4]...

REFERENCE

1. U.S. Food and Drug Administration. (2020). Dissolution testing and specification criteria for immediate-release solid oral dosage forms containing BCS class 1 and class 3 drugs. FDA Guidance for Industry. https://www.fda.gov/media/88828/download
2. Siepmann, J., & Siepmann, F. (2020). Mathematical modeling of drug dissolution. International Journal of Pharmaceutics, 453(1), 12–24. https://doi.org/10.1016/j.ijpharm.2020.119712
3. Raza, A., Bhatt, D. L., & Wacker, M. G. (2022). Machine learning applications in pharmaceutical science: A systematic review. Drug Discovery Today, 27(3), 750–768. https://doi.org/10.1016/j.drudis.2021.11.014
4. Bannigan, P., Aldeghi, M., Bao, Z., Talos, F., Pirila, P., & Allen, C. (2021). Machine learning directed drug formulation development. Advanced Drug Delivery Reviews, 175, 113806. https://doi.org/10.1016/j.addr.2021.05.016
5. Patel, D., & Amin, A. (2020). Support vector machine for prediction of nifedipine extended-release dissolution profiles. AAPS PharmSciTech, 21(6), 201. https://doi.org/10.1208/s12249-020-01742-x
6. Valizadeh, A., Shakeri, A., Jafari, M., & Nokhodchi, A. (2024). LightGBM and SHAP analysis for amorphous solid dispersion dissolution prediction across biorelevant media. Journal of Controlled Release, 368, 234–248. https://doi.org/10.1016/j.jconrel.2024.02.031
7. Batista, C. M. S., Lins, L. C., Fonseca, L. P., & Brechet, Y. (2022). Feature selection for ANN modelling of transdermal drug permeation. European Journal of Pharmaceutical Sciences, 173, 106149. https://doi.org/10.1016/j.ejps.2022.106149
8. Galata, D. L., Farkas, A., Kovacs, B., Hidy, D., Kátai-Urbán, G., Pataki, H., & Marosi, G. (2021). Real-time dissolution prediction using near-infrared spectroscopy and artificial neural networks. Journal of Pharmaceutical and Biomedical Analysis, 200, 114065. https://doi.org/10.1016/j.jpba.2021.114065
9. Mohammed, H. A., Abdullah, A. H., & Ahmed, S. I. (2023). Comparative machine learning for doxorubicin release prediction from pH-responsive chitosan nanoparticles. International Journal of Nanomedicine, 18, 3451–3467. https://doi.org/10.2147/IJN.S410832
10. Sharma, R., Mehrotra, A., & Bansal, A. K. (2022). Random forest modeling of f2 similarity for polymer-coated multiparticulate drug release. European Journal of Pharmaceutics and Biopharmaceutics, 170, 92–103. https://doi.org/10.1016/j.ejpb.2021.12.001
11. Chen, Y., Liu, X., Liang, Z., & Zhang, H. (2023). XGBoost prediction of biphasic drug release from lipid-polymer hybrid nanoparticles. International Journal of Pharmaceutics, 639, 122964.

https://doi.org/10.1016/j.ijpharm.2023.122964
12. Ofosu, K., Baidoo, M. F., Okeke, C. I., & Amoateng, P. (2021). Artificial neural network modeling of metformin release from gastroretentive floating tablets. Drug Development and Industrial Pharmacy, 47(9), 1374–1384. https://doi.org/10.1080/03639045.2021.1981535
13. Khalid, G. M., Mudassir, J., Asif, M., & Rehman, T. ur. (2023). LSTM neural networks for PAT-driven dissolution prediction from in-line NIR data. Computers & Chemical Engineering, 172, 108173. https://doi.org/10.1016/j.compchemeng.2023.108173
14. Rodriguez, C. F., Cuellar, M., & Acevedo, D. (2023). Transfer learning for dissolution prediction in data-limited pharmaceutical scenarios. Molecular Pharmaceutics, 20(8), 4112–4126. https://doi.org/10.1021/acs.molpharmaceut.3c00198
15. European Medicines Agency. (2023). Reflection paper on the use of artificial intelligence in the lifecycle of medicines. EMA/746648/2023.
16. Kambale, K.S., Sawant, N.M., Mulani, A.O., More, V.P., Zambare, S.A. (2026). RNN-LSTM Based Model for Automatic Heart Disease Prediction Using the UCI Heart Disease Dataset. In: Kumar, A., Gunjan, V.K., Senatore, S., Hu, YC. (eds) Proceedings of the 6th International Conference on Data Science, Machine Learning and Applications- Volume 1. ICDSMLA2024 2024. Lecture Notes in Electrical Engineering, vol 1528. Springer, Singapore. https://doi.org/10.1007/978-981-95-5831-5_28
17. Sawant, N.M., Mulani, A.O., Kondooru, S., Linge, S.G., Gawande, P.G., Koli, M.S. (2026). AgriRent: Renting the Farm Equipment. In: Kumar, A., Gunjan, V.K., Senatore, S., Hu, YC. (eds) Proceedings of the 6th International Conference on Data Science, Machine Learning and Applications- Volume 1. ICDSMLA2024 2024. Lecture Notes in Electrical Engineering, vol 1528. Springer, Singapore. https://doi.org/10.1007/978-981-95-5831-5_35
18. Salunkhe Shweta, Mulani, Altaf Osman , Shahane, Deepali , Rana, Manish , Shukla, Shivam Mahendra & Jadhav, Makarand M. (2026) Secure image transmission using chaotic encryption and DWT watermarking on reconfigurable platform, Journal of Discrete Mathematical Sciences and Cryptography, pp. 1-13, DOI: 10.47974/JDMSC-2608
19. Chaudhari Kalyani R., Mulani Altaf O., Gajare Milind P., Jadhav Vaishali, Yawle Pranali & Bang Arti Vasant (2026) Bit error rate analysis of various error correction codes with concatenated RS-convolutional codes, Journal of Discrete Mathematical Sciences and Cryptography, pp. 1-16, DOI: 10.47974/JDMSC-2401
20. Altaf Osman Mulani, Vaibhav Godase and Swapnil Takale. Enhanced Multimodal Disease Prediction Using Hybrid Ensemble Learning and AutoML Techniques. Research and Reviews: Journal of Computational Biology. 2025; 15(01). Available from: https://journals.stmjournals.com/rrjocb/article=2025/view=234676
21. Mulani A.O., Godase V.V., Takale S.R., Ghodake R.G. Ethical Challenges and Governance of AI in Healthcare, Education, Finance, and Security Sectors. Advances in Computational Intelligence and Robotics. 2025 Nov 14; 123–154. Available from: https://www.igi-global.com/chapter/ethical-challenges-and-governance-of-ai-in-healthcare-education-finance-and-security-sectors/396083
22. Altaf O. Mulani. Early Alzheimer's Disease Detection Using Deep Ensemble Learning and MRI Image Analysis. Research and Reviews: Journal of Computational Biology. 2026; 15(01).
23. Altaf O. Mulani, Vaibhav V. Godase, Swapnil R. Takale, Rahul G. Ghodake (2025), Advancements in Artificial Intelligence: Transforming Industries and Society, International Journal of Artificial Intelligence of Things (AIoT) in Communication Industry, Vol. 1, Issue 2, pp 1-5.
24. Altaf O. Mulani, Vaibhav V. Godase, Swapnil R. Takale, Rahul G. Ghodake (2025), AI-Powered Predictive Analytics in Healthcare: Revolutionizing Disease Diagnosis and Treatment, Journal of Advance Electrical Engineering and Devices, Vol. 3, Issue 2, pp 27-34.
25. Altaf Osman Mulani, Deshmukh M., Jadhav V., Chaudhari K., Mathew A.A., Shweta Salunkhe. Transforming Drug Therapy with Deep Learning: The Future of Personalized Medicine. Drug Research. 2025 Aug 29.
26. Mulani, A.O., Kulkarni, T.M. (2025). Face Mask Detection System Using Deep Learning: A Comprehensive Survey. In: Singh, S., Arya, K.V., Rodriguez, C.R., Mulani, A.O. (eds) Emerging Trends in Artificial Intelligence, Data Science and Signal Processing. AIDSP 2023. Communications in Computer and Information Science, vol 2439. Springer, Cham. https://doi.org/10.1007/978-3-031-88759-8_3.
27. Karve, S., Gangonda, S., Birajadar, G., Godase, V., Ghodake, R., Mulani, A.O. (2025). Optimized Neural Network for Prediction of Neurological Disorders. In: Singh, S., Arya, K.V., Rodriguez, C.R., Mulani, A.O. (eds) Emerging Trends in Artificial Intelligence, Data Science and Signal Processing. AIDSP 2023. Communications in Computer and Information Science, vol 2440. Springer, Cham. https://doi.org/10.1007/978-3-031-88762-8_18.
28. Saurabh Singh, Karm Veer Arya, Ciro Rodriguez Rodriguez, and Altaf Osman Mulani, Emerging Trends in Artificial Intelligence, Data Science and Signal Processing, Communications in Computer and Information Science

(CCIS), volume 2440.
29. Saurabh Singh, Karm Veer Arya, Ciro Rodriguez Rodriguez, and Altaf Osman Mulani, Emerging Trends in Artificial Intelligence, Data Science and Signal Processing, Communications in Computer and Information Science (CCIS), volume 2439.
30. Mulani, A. O., Sardey, M. P., Kinage, K., Salunkhe, S. S., Fegade, T., & Fegade, P. G. (2025). ML-powered Internet of Medical Things (MLIOMT) structure for heart disease prediction. Journal of Pharmacology and Pharmacotherapeutics, 16(1), 38-45.
31. Aiwale, S., Kolte, M. T., Harpale, V., Bendre, V., Khurge, D., Bhandari, S., ... & Mulani, A. O. (2024). Non-invasive Anemia Detection and Prediagnosis. Journal of Pharmacology and Pharmacotherapeutics, 15(4), 408-416.
32. Mulani, A. O., Jadhav, M. M., & Seth, M. (2022). Painless machine learning approach to estimate blood glucose level with non-invasive devices. In Artificial intelligence, internet of things (IoT) and smart materials for energy applications (pp. 83-100). CRC Press.
33. Mulani, A. O., Birajadar, G., Ivković, N., Salah, B., & Darlis, A. R. (2023). Deep learning based detection of dermatological diseases using convolutional neural networks and decision trees. Traitement du Signal, 40(6), 2819.
34. Jadhav, H. M., Mulani, A., & Jadhav, M. M. (2022). Design and development of chatbot based on reinforcement learning. Machine Learning Algorithms for Signal and Image Processing, 219-229.
35. Kashid, M. M., Karande, K. J., & Mulani, A. O. (2022, November). IoT-based environmental parameter monitoring using machine learning approach. In Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1 (pp. 43-51). Singapore: Springer Nature Singapore.
36. Mulani, A. O., Cengiz, K., & Tripathi, S. L. (Eds.). (2026). AI, Machine Learning, and IoT for Communication and Medical Applications. IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-3864-3
37. Mulani, A. O., Godase, V. V., Takale, S. R., & Ghodake, R. G. (2025). Ethical Challenges and Governance of AI in Healthcare, Education, Finance, and Security Sectors. Advances in Computational Intelligence and Robotics, 123–154. https://doi.org/10.4018/979-8-3373-7011-8.ch006
.