Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

PREDICTING LONG-ACTING INJECTABLE RELEASE USING POLYMER DEGRADATION, DRUG LOADING, AND MICROSPHERE MORPHOLOGY

Ivan Petrov[1], Olga Ivanova[1]*, Dmitry Smirnov[2]

1. *Department of AI-Based Pharmaceutical Sciences, Faculty of Medicine, Lomonosov Moscow State University, Moscow, Russia.*
2. *Department of Computational Drug Engineering, Faculty of Pharmacy, Saint Petersburg State University, Saint Petersburg, Russia.*

ARTICLE INFO

Received:
16 October 2025
Received in revised form:
19 January 2026
Accepted:
21 January 2026
Available online:
28 February 2026

Keywords: Long-acting injectables, PLGA microspheres, Predictive modeling, Machine learning, Polymer degradation, Drug release

ABSTRACT

Long-acting injectable microspheres are an important platform for sustained drug delivery because they can maintain therapeutic exposure over extended dosing intervals. Their release behavior remains difficult to predict because polymer degradation, drug loading, and microsphere morphology interact across burst, lag, and erosion-controlled phases. Current formulation development often depends on empirical iteration, in which candidate batches are manufactured and tested before mechanistic understanding is complete. This trial-and-error workflow can slow development when small changes in polymer grade, drug distribution, or particle structure alter the full release profile. The objective of this predictive-model article is to describe a machine learning framework for forecasting the in-vitro release curve of long-acting injectable microspheres from formulation and morphology descriptors. The same framework could be extended conceptually to estimate in-vivo pharmacokinetic behavior when suitable bridging data are available. A gradient-boosted tree or multi-output regression model would be trained on curated long-acting injectable formulation records. Inputs would encode polymer chemistry and degradation properties, drug loading and physicochemical features, and quantitative microsphere morphology descriptors such as particle size, porosity, surface area, and internal structure. Conceptually, the model could predict the release profile under different formulation and processing conditions while ranking the relative influence of polymer, drug, and morphology features. Such predictions would be expected to support virtual screening of formulation variants before experimental confirmation. A model-informed formulation strategy could help shift long-acting injectable microsphere development from empirical testing toward rational selection of polymer, drug-loading, and morphology targets. This approach should be evaluated prospectively before being used to support high-impact development or regulatory decisions.

To Cite This Article: Petrov I, Ivanova O, Smirnov D. Predicting Long-Acting Injectable Release Using Polymer Degradation, Drug Loading, and Microsphere Morphology. Pharmacophore. 2026;17(1):33-42. https://doi.org/10.51847/MT3QwGFdub

Introduction

Long-acting injectable microspheres provide therapeutic and commercial advantages by reducing dosing frequency, improving adherence, and supporting sustained exposure for drugs that benefit from prolonged delivery [1]. PLGA-based long-acting injectable formulations are especially prominent because PLGA degradation can be tuned through copolymer composition, molecular weight, and end-group chemistry [2]. However, consistent release over weeks or months remains technically difficult because release is governed by coupled diffusion, polymer swelling, water uptake, pore formation, and erosion [3]. For predictive modeling, this means the target output should not be a single release value but a time-dependent curve whose shape reflects multiple overlapping mechanisms [4].

Current development practice still relies heavily on make-and-measure formulation screening, where batches with different polymer grades, drug loadings, process parameters, and particle-size distributions are produced and then evaluated in vitro [5]. This workflow is burdensome because changes in manufacturing conditions can shift particle morphology, which in turn can affect burst release and later-stage erosion-controlled release [6]. Experimental studies of PLGA microspheres show that particle size, polydispersity, and polymer degradation must be considered together when interpreting release behavior [7]. A

Corresponding Author: Olga Ivanova; Department of AI-Based Pharmaceutical Sciences, Faculty of Medicine, Lomonosov Moscow State University, Moscow, Russia. E-mail: olga.ivanova@outlook.com.

predictive model would therefore be most useful when it converts historical formulation data into a structured decision tool rather than simply replacing laboratory testing.

Machine learning has begun to enter drug delivery development through models that relate formulation descriptors to release or carrier performance [8]. Recent work has specifically framed polymeric long-acting injectables as a domain where machine learning could accelerate formulation design by learning non-linear relationships between composition and release behavior [9]. The availability of curated microparticle and nanoparticle formulation datasets further supports model development, because standardized descriptors can make cross-study learning more feasible [10]. Yet the application gap remains substantial for PLGA microspheres because release depends simultaneously on polymer degradation, drug loading, and microstructure rather than on formulation composition alone [11].

The central thesis of this article is that a machine learning model could integrate polymer degradation characteristics, drug-related properties, and microsphere morphology features to predict the complete release profile of long-acting injectable microspheres [12]. A physiology-based or pharmacokinetically informed framework could then connect in-vitro release predictions to in-vivo exposure descriptors when relevant data are available [13]. Such a model should be designed conceptually to estimate burst release, lag behavior, sustained-release rate, and derived pharmacokinetic tendencies without overstating performance before validation [14]. The proposed framework is therefore positioned as a predictive-model architecture for formulation decision support, not as a report of completed experimental outcomes.

Background

Mechanisms of Release from Polymeric Microspheres

Release from polymeric microspheres usually begins with an initial burst driven by surface-associated or near-surface drug, followed by diffusion through hydrated polymer domains and later release associated with polymer relaxation, swelling, and erosion [4]. PLGA and PLA systems can show complex internal water accumulation and swelling behavior, which makes the release curve sensitive to both polymer chemistry and particle structure [15]. Porosity, drug distribution, and surface morphology can create fast diffusion pathways that alter early release, while matrix densification or hollow internal structures may influence lag behavior [5]. A predictive model should therefore encode release as a multiphase process rather than assuming that a single kinetic equation explains all formulations.

Polymer Degradation and Its Impact on Release

PLGA degradation is commonly described as hydrolytic bulk erosion, but the observed release behavior depends on water uptake, autocatalytic acid accumulation, polymer molecular weight decline, and mass loss [3]. Molecular weight, inherent viscosity, lactide:glycolide ratio, and polymer source can modulate the degradation timescale and the resulting release profile [16]. Studies of injectable PLGA formulations also show that polymer grade and microsphere formation are interconnected, because processing can affect morphology while polymer chemistry governs later erosion [6]. In a machine learning framework, these polymer properties should be represented as mechanistically meaningful features rather than treated as interchangeable categorical labels.

Drug Loading, Drug Form, and Drug-Polymer Interactions

Drug loading can influence release by changing the concentration gradient, internal drug distribution, and probability of drug-rich domains near the particle surface [17]. Drug solubility and physical form are also important because amorphous, crystalline, or phase-separated drug regions may interact differently with the polymer matrix during hydration and degradation [18]. Long-acting injectable case studies show that drug-polymer compatibility and formulation composition can determine whether release is smooth, delayed, or burst-dominated [14]. Accordingly, a predictive model should include drug loading, drug-polymer ratio, solubility class, and physical-state descriptors as inputs that interact with polymer and morphology variables.

Microsphere Morphology and Its Quantitative Description

Microsphere morphology includes particle-size distribution, surface texture, internal porosity, and structural classes such as dense, porous, matrix-like, hollow, or core-shell particles [5]. Particle size and polydispersity have been directly linked to progesterone release from PLGA microparticles, demonstrating that morphology can alter both early and later release phases [7]. Image-based and instrument-derived morphology descriptors can be converted into features such as D50, span, qualitative surface class, porosity, and surface area [11]. For model development, these descriptors should be harmonized across SEM, micro-CT, laser diffraction, and gas-adsorption measurements so that morphology can be learned quantitatively rather than described only in narrative terms.

Machine Learning in Drug Delivery and Release Prediction

Machine learning has been applied to drug delivery problems ranging from polymeric long-acting injectable design to PLGA nanoparticle formulation and antiviral delivery systems [8, 19]. Several recent studies have explored machine learning for predicting drug release from polymeric delivery systems, showing that non-linear models may capture formulation-property relationships that are difficult to specify mechanistically [20]. Work on PLGA nanoparticles and broader polymeric systems suggests that formulation variables, process factors, and material descriptors can be integrated into predictive pipelines [21,

22]. The distinctive challenge for LAI microspheres is that the model should predict a time-varying release curve, not only a single endpoint such as total release or loading efficiency.

Model Development Overview

High-Level Prediction Framework

The proposed framework would assemble polymer descriptors, drug descriptors, manufacturing-linked morphology descriptors, and in-vitro release conditions into a structured feature vector for each LAI microsphere formulation [9]. The primary output would be a predicted percent-release curve over predefined sampling times, allowing the model to represent burst, lag, and sustained phases in one prediction task [12]. When in-vivo data are available, the model could include a secondary output layer for pharmacokinetic descriptors such as Cmax tendency, AUC tendency, and tmax tendency without claiming direct clinical equivalence from in-vitro data alone [13]. This architecture would support model-informed comparison of candidate formulations before experimental prioritization.

Figure 1 presents the proposed machine learning architecture for converting polymer degradation, drug-loading, microsphere morphology, and release-test descriptors into full-profile long-acting injectable release predictions.

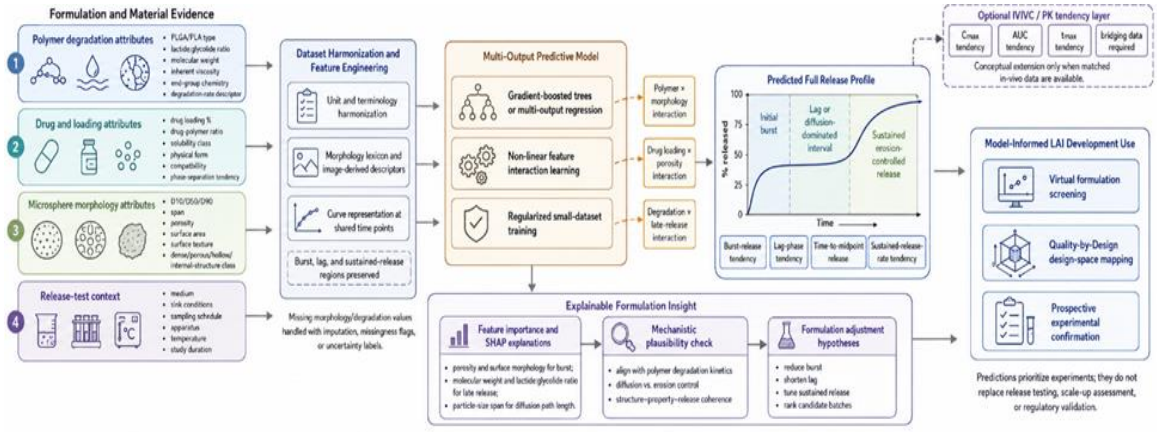

Figure 1. Machine Learning Framework for Predicting Long-Acting Injectable Microsphere Release from Polymer Degradation, Drug Loading, and Morphology

Core Input Features

Core polymer inputs should include polymer type, lactide:glycolide ratio, molecular weight, inherent viscosity, end-group class, and estimated or measured degradation-rate descriptors because these features influence erosion and late-stage release [16]. Drug inputs should include loading percentage, drug-polymer ratio, solubility category, physical form, and indicators of drug-polymer compatibility or phase separation [14]. Morphology inputs should include D10, D50, D90, span, porosity, surface area, and an internal-structure label such as dense, porous, or hollow [7]. These features should be encoded in a way that preserves mechanistic meaning while allowing the algorithm to learn non-linear interactions among composition, degradation, and structure.

Design Principles

The model should be multi-output so that it can predict the entire release curve rather than force the formulation problem into isolated single-time-point predictions [23]. It should also be interpretable, because formulators need to understand whether predicted release is driven by polymer degradation, drug loading, particle size, or porosity [8]. Because LAI datasets are often small and heterogeneous, the modeling strategy should include regularization, robust validation, and careful handling of missing morphology or degradation measurements [20]. Finally, the framework should align with Quality-by-Design principles by linking controllable material and process attributes to a target release profile [24].

Table 1 shows the recommended characteristics and strategies for modeling long-acting injectable (LAI) drug release profiles.

Table 1. Key Considerations for Modeling LAI Drug Release

Aspect	Recommendation / Requirement	Rationale
Model Output	Multi-output prediction	Captures the full release curve instead of isolated time points
Interpretability	High	Helps formulators identify the contributions of polymer degradation, drug loading, particle size, and porosity
Data Strategy	Regularization, robust validation, careful handling of missing data	Addresses small and heterogeneous LAI datasets
Design Alignment	Link material and process attributes to target release profile	Ensures adherence to Quality-by-Design (QbD) principles

Data Sources and Feature Engineering

Compilation of an LAI Formulation Dataset

An LAI formulation dataset would be compiled from peer-reviewed PLGA and PLA microsphere studies, curated formulation datasets, and internal formulation records where available [10]. Each record should extract polymer grade, drug loading, particle-size descriptors, qualitative and quantitative morphology, release medium, sampling schedule, and in-vitro release method because inconsistent experimental conditions can confound release interpretation [4]. Studies of PLGA microparticles and LAI products show that release profiles are sensitive to both formulation composition and test context, so metadata on sink conditions and apparatus should be retained as covariates [1]. The dataset should be curated for conceptual model development without assuming that all literature-derived measurements are directly interchangeable.

Encoding Polymer Degradation and Drug Properties

Polymer molecular weight, lactide:glycolide ratio, inherent viscosity, and polymer source should be encoded as continuous or structured categorical variables because they influence degradation and release in related but non-identical ways [16]. Degradation rate could be represented by experimentally observed mass-loss behavior, molecular-weight decline, or an estimated descriptor derived from polymer chemistry when direct measurements are unavailable [3]. Drug properties should include loading, solubility, physical state, and formulation-specific interaction descriptors because drug-rich domains or phase separation may change burst release and diffusion pathways [18]. When exact drug solubility values are unavailable, categorical solubility classes may be used, but the model should track uncertainty in those assignments.

Quantifying Microsphere Morphology

Morphology features should be derived from laser-diffraction size distributions, SEM-based surface classification, image-analysis metrics, and porosity or surface-area measurements where available [5]. Particle-size descriptors such as D50 and distribution span are important because size and polydispersity can influence both diffusion distance and degradation-associated release [7]. Qualitative descriptions such as smooth, porous, collapsed, hollow, or dense should be mapped to a consistent morphology lexicon so that text-derived labels can be used alongside numeric measurements [11]. This feature-engineering step is essential because microsphere morphology often mediates the relationship between manufacturing process and release behavior.

Table 2 defines the mechanistically structured feature architecture required to represent polymer degradation, drug-loading behavior, microsphere morphology, and release-test context in a full-profile predictive model.

Table 2. Mechanistically Structured Feature Architecture for Predicting Long-Acting Injectable Microsphere Release Profiles

Feature domain	Specific variables to encode	Mechanistic relevance to release behavior	Expected strongest release-curve region	Modeling implication
Polymer identity and chemistry	PLGA or PLA type; lactide:glycolide ratio; end-group chemistry; polymer source	Determines hydrophobicity, water penetration, acid accumulation, and erosion tendency [2, 3, 16]	Middle-to-late release and erosion-controlled phase	Encode as structured categorical and continuous variables rather than simple polymer-name labels.
Polymer degradation descriptors	Molecular weight; inherent viscosity; molecular-weight decline; mass-loss rate; estimated degradation-rate class	Represents the time scale over which matrix relaxation, erosion, and pore evolution contribute to drug liberation [3, 16]	Lag phase and sustained-release phase	Include direct degradation measurements where available and estimated descriptors with uncertainty flags where unavailable.
Drug loading and drug-polymer ratio	Drug loading percentage; drug-polymer ratio; dose density within the matrix	Influences concentration gradients, drug-rich domains, near-surface exposure, and probability of burst release [14, 17]	Early burst and total release trajectory	Model as a continuous driver that interacts with porosity, particle size, and drug physical form.
Drug physicochemical and physical-state features	Solubility class; crystalline or amorphous form; compatibility; phase-separation tendency	Affects dissolution inside the hydrated polymer, diffusion pathways, and local drug-polymer interactions [14, 18]	Burst, lag, and sustained regions depending on matrix distribution	Preserve formulation-specific physical-state annotations instead of treating the drug as a generic active ingredient.
Particle-size distribution	D10; D50; D90; span; polydispersity	Controls diffusion distance, surface-area-to-volume relationships, and heterogeneity of release across particle populations [5, 7]	Early diffusion and sustained release	Use full size-distribution descriptors rather than mean size alone.
Surface morphology	Smooth, rough, porous, collapsed, cracked, or textured surface class; SEM-derived surface features	Indicates the extent of near-surface drug exposure and diffusion pathways that can accelerate initial release [5, 11]	Initial burst	Map qualitative descriptions to a standardized morphology lexicon for cross-study learning.

Internal microsphere structure	Dense, porous, hollow, matrix-like, core-shell, or phase-separated structure	Shapes water ingress, internal pore connectivity, and delayed release behavior [5]	Lag phase and transition into sustained release	Encode as categorical morphology states, supported by micro-CT, SEM cross-sections, or text-derived labels when available.
Porosity and surface area	Total porosity; pore-size class; BET surface area; image-derived pore fraction	Reflects internal access pathways for medium penetration and drug diffusion [5, 11]	Burst and mid-phase release	Treat as a key mediator between manufacturing process and release behavior.
Release-test conditions	Medium; pH; sink conditions; agitation; temperature; sampling schedule; apparatus	Experimental context can confound apparent formulation effects and alter measured release rates [1, 4]	Entire release curve	Retain as covariates so the model does not mistake test-method variation for formulation performance.
Curve-representation features	Harmonized time-point release values; burst descriptor; lag descriptor; sustained-release slope; kinetic-model parameters	Converts heterogeneous release profiles into a shared prediction target while preserving multiphase behavior [4, 23]	Entire release curve	Use multi-output targets or hybrid empirical-mechanistic descriptors instead of a single endpoint.

Predictive Model Architecture

Model Choice – Gradient-Boosted Trees with Multi-Output Regression

A gradient-boosted tree model with multi-output regression would be suitable because it can capture non-linear interactions among polymer properties, drug loading, and morphology descriptors while remaining more interpretable than many high-capacity neural networks [8]. Such a model could be configured to predict a vector of release percentages at standard sampling times, allowing the full release curve to be treated as the prediction target [12]. Prior machine learning work in polymeric drug delivery supports the use of flexible supervised models when formulation features interact in non-additive ways [20]. The architecture should be presented as a conceptual candidate and compared against simpler baselines during evaluation rather than assumed to be optimal.

Input Feature Vector and Pre-processing

The input feature vector should combine continuous variables such as molecular weight, drug loading, D50, porosity, and surface area with structured categorical variables such as polymer type, drug physical form, and morphology class [22]. Continuous variables would require scaling or normalization when used in algorithms sensitive to feature magnitude, while tree-based methods may require less scaling but still benefit from consistent units and harmonized feature definitions [25]. Missing morphology or degradation descriptors should be handled with imputation, missingness indicators, or model classes that can explicitly manage incomplete inputs [19]. Feature selection should be guided by formulation relevance so that dimensionality reduction does not remove variables known to influence burst or erosion behavior.

Output: Release Profile and Derived PK Parameters

The principal output would be the predicted in-vitro release profile, expressed as percent released at harmonized time points selected to capture burst, lag, and sustained-release regions [23]. Derived descriptors such as burst-release tendency, time-to-midpoint release tendency, lag-phase tendency, and sustained-release-rate tendency could be calculated from the predicted curve without presenting them as validated numerical outcomes [4]. If matched in-vivo pharmacokinetic data are available, a secondary mapping could estimate exposure-related tendencies while acknowledging that in-vitro release is only a surrogate for biological performance [13]. This output structure would allow formulators to compare candidates by expected profile shape rather than by a single endpoint.

Handling Small Datasets and Multi-Modal Release Data

Strategies for Data-Scarce Formulation Domains

LAI microsphere datasets are likely to remain small relative to the complexity of the formulation space, so the model should be evaluated using validation strategies that avoid overconfident conclusions from closely related batches [20]. Leave-one-formulation-out or leave-one-drug-out validation would be appropriate because it tests whether the model can generalize across meaningful formulation differences rather than memorize local composition trends [9]. Physical priors from zero-order, first-order, Higuchi, or erosion-informed release models could be used as regularizing features, especially when direct examples of a target polymer-drug combination are sparse. Simulation, bootstrapping, and constrained augmentation may support exploratory modeling, but such synthetic records should be treated as aids to model training rather than substitutes for experimentally measured release curves.

Integrating Diverse Release Curves into a Common Representation

Published release profiles often differ in sampling frequency, total study duration, medium composition, and reporting format, so the first modeling task should be to represent heterogeneous curves in a shared structure [4]. One option is to interpolate release values at predefined time points chosen to reflect burst, lag, and sustained-release phases, while another is to fit each profile to a common kinetic equation and use the fitted parameters as intermediate targets [7]. Korsmeyer-Peppas, Higuchi,

and related release models may help summarize curve shape, but they should not be assumed to capture every microsphere mechanism in isolation. A hybrid representation, combining sampled release percentages with kinetic descriptors, would allow the machine learning model to learn both empirical profile shape and mechanistically interpretable release tendencies.

Multi-Fidelity Modeling

A multi-fidelity strategy would allow the model to use inexpensive early observations, formulation descriptors, or approximate morphology features to prioritize formulations before full release testing is completed [21]. Early burst measurements may provide information about surface-associated drug and porosity, while polymer degradation descriptors would be expected to contribute more strongly to later sustained release [3]. Lower-fidelity records, such as incomplete release curves or qualitative morphology descriptions, could be incorporated with explicit uncertainty flags so they do not carry the same evidentiary weight as fully characterized LAI studies [11]. This approach would be especially useful in early development, where the objective is to rank candidate formulations for experimental follow-up rather than claim definitive release performance. Figure 2 illustrates how multi-fidelity modeling can integrate early release observations, formulation descriptors, morphology evidence, degradation-related features, and uncertainty flags to prioritize candidate long-acting injectable formulations before full release testing is completed.

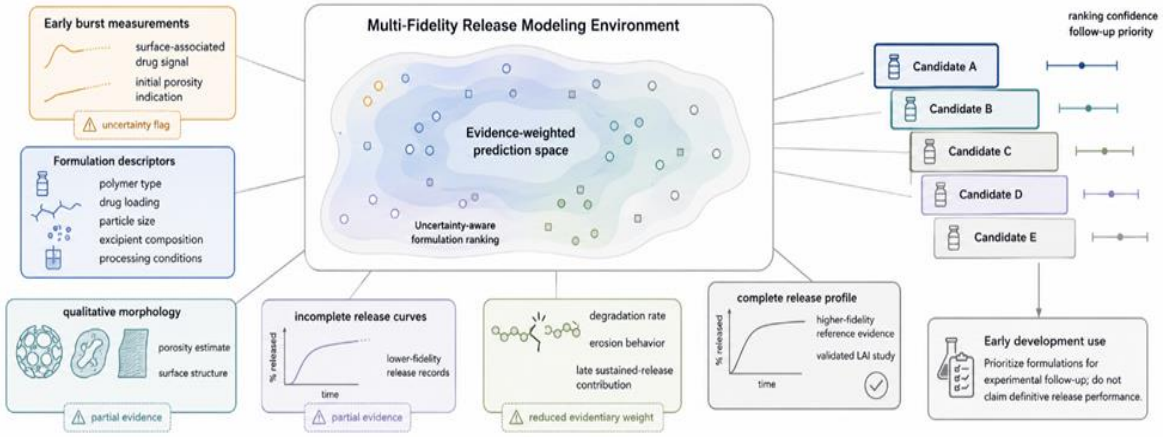

Figure 2. Multi-Fidelity Modeling for Early Prioritization of Long-Acting Injectable Formulations

Model Interpretability and Formulation Insights

SHAP Analysis of Release Drivers

Interpretability methods such as SHAP analysis would be valuable because the model must provide formulation insight, not only predicted release curves [8]. Conceptually, SHAP summary and dependence plots could identify whether porosity, surface morphology, drug loading, or particle-size span is most associated with burst-release behavior [12]. Later release regions would be expected to depend more strongly on molecular weight, lactide:glycolide ratio, polymer source, and degradation descriptors, consistent with studies showing that PLGA properties influence erosion-driven release [16]. These explanations should be treated as hypothesis-generating maps that guide formulation thinking and should be checked against mechanistic understanding before being used for development decisions.

Translating Model Logic into Design Decisions

A model that predicts excessive burst release could guide formulators toward lower drug loading, reduced surface porosity, altered solvent-removal conditions, or a morphology target that decreases near-surface drug exposure [5]. If the predicted sustained phase is too slow, the model could suggest a polymer with a faster degradation tendency, a lower molecular weight, or a different lactide:glycolide ratio, while still requiring experimental confirmation [2]. Case studies of sustained-release microspheres show that optimization decisions often involve simultaneous adjustment of drug loading, polymer composition, and process-linked morphology rather than isolated single-factor changes [17, 24]. The model's main value would therefore be to translate complex feature interactions into actionable formulation hypotheses.

Integration Into Lai Formulation Development

Early-Stage Feasibility Screening

In early development, the model could be used to rank candidate polymers, drug loadings, and morphology targets according to their expected ability to approach a desired release profile before microspheres are synthesized [13]. This virtual screening role is consistent with recent machine learning frameworks for polymeric long-acting injectables, which position predictive models as tools for accelerating formulation selection [8]. Curated formulation datasets could support this screening by providing historical examples of how polymer type, particle properties, and release outcomes vary across published systems

[10]. The model should be used to prioritize candidates for testing, not to eliminate experimental verification of promising or clinically important formulations.
Table 3 shows the potential applications and data considerations for predictive modeling in early LAI formulation development.

Table 3. Early-Stage Modeling for LAI Formulation Selection

Aspect	Recommendation / Approach	Rationale / Benefit
Candidate Ranking	Evaluate polymers, drug loadings, and morphology targets	Identifies formulations likely to achieve target release before synthesis
Virtual Screening	Use predictive models to accelerate formulation selection	Speeds up decision-making in early development
Dataset Use	Curated historical formulation data	Provides examples of polymer types, particle properties, and release outcomes for training and validation
Experimental Verification	Prioritize candidates, not replace testing	Ensures promising or clinically relevant formulations are experimentally confirmed

Supporting Quality-by-Design and Regulatory Submissions
For Quality-by-Design use, the model could define a conceptual design space linking polymer attributes, drug loading, and morphology descriptors to a target release-profile region [24]. Such a design space would be expected to support risk assessment by identifying which material or process attributes are most likely to affect burst, lag, or sustained-release behavior [1]. Predictive models may also help organize evidence across in-vitro release testing, formulation characterization, and in-vivo pharmacokinetic bridging when these data are available [13]. In a regulatory context, the model would need transparent feature definitions, documented limitations, and prospective confirmation before being used as supportive evidence.

Evaluation Strategy
Prediction Accuracy for Release Profiles
Evaluation should focus on whether the model could reproduce the shape of the release curve across early, middle, and late sampling regions, rather than only whether it matches one endpoint [12]. Candidate metrics may include time-point-level error, curve similarity, burst-region agreement, and sustained-phase agreement, but this article does not report numerical performance values because the framework is conceptual [23]. Comparisons against kinetic-model baselines such as Higuchi, first-order, or Korsmeyer-Peppas representations would help determine whether machine learning adds value beyond conventional curve fitting. The evaluation should also examine whether predicted derived descriptors are consistent with formulation mechanisms such as porosity-driven burst release and degradation-driven late release [3].

Temporal and External Validation
Temporal validation would test the model on formulations published or developed after the training period, which is important because formulation practices and characterization methods evolve over time [9]. External validation should use data from different laboratories, polymer suppliers, or manufacturing processes to determine whether the model generalizes beyond local experimental conventions [16]. Published examples of injectable PLGA microspheres with in-vitro and in-vivo characterization could provide conceptually useful external cases for testing whether release predictions align with pharmacokinetic tendencies [14, 18]. This validation design would be more informative than random splitting alone because random splits may place highly similar formulations in both training and test partitions.

Utility in Formulation Development
The model's practical utility should be assessed through a prospective workflow in which formulation scientists use predictions to choose a small set of candidate microsphere designs for experimental testing [8]. The key question would be whether model-guided selection reduces unnecessary formulation iteration while still identifying candidates with release profiles close to the target [25]. Prior work using supervised machine learning to predict release from polymeric matrices and related drug delivery systems supports the idea that predictive tools can guide formulation decisions when paired with careful validation. Utility should therefore be evaluated not only by profile agreement but also by whether the model improves decision quality in realistic LAI development settings.
Table 4 provides a validation, interpretability, and development-utility framework for determining whether the proposed model can support formulation decisions beyond statistical release-curve prediction.

Table 4. Validation, Interpretability, and Formulation-Decision Framework for Model-Informed LAI Microsphere Development

Evaluation or decision layer	Purpose	Recommended analytical approach	What the result should clarify	Development-use interpretation
Time-point prediction accuracy	Determine whether the model predicts release percentages at clinically and	Mean absolute error or root mean squared error at harmonized early, middle, and late time points	Whether prediction error is concentrated in burst, lag, or sustained-release regions	Identifies which part of the release profile requires more data, better descriptors, or mechanistic constraints.

	formulation-relevant sampling times			
Curve-shape agreement	Evaluate whether the model captures the overall profile rather than isolated time points	Curve similarity metrics; area-between-curves analysis; dynamic time-aligned comparison where appropriate	Whether the predicted profile has the correct burst-lag-sustained shape [4, 23]	Supports formulation ranking by expected release behavior rather than by a single endpoint.
Burst-region validation	Test whether the model captures early release driven by near-surface drug, porosity, and particle structure	Early time-window error; burst-release classification; threshold agreement for excessive burst	Whether morphology and loading descriptors adequately explain initial release [5, 17]	Guides decisions to modify surface porosity, drug loading, solvent-removal conditions, or particle-size distribution.
Sustained-phase validation	Test whether the model captures erosion-controlled and degradation-associated release	Late time-window slope error; sustained-release-rate agreement; comparison with degradation-informed baselines	Whether polymer degradation descriptors explain late release [3, 16]	Guides polymer-grade selection, molecular-weight adjustment, or lactide:glycolide ratio tuning.
Baseline comparison	Determine whether machine learning adds value beyond conventional release-curve fitting	Compare against Higuchi, first-order, Korsmeyer-Peppas, or erosion-informed kinetic models	Whether non-linear ML interactions improve prediction beyond standard mechanistic summaries	Prevents overclaiming if simpler kinetic approaches perform similarly.
Leave-one-formulation-out validation	Assess generalization across formulation variants rather than memorization of closely related batches	Hold out complete formulation families or candidate batches during validation	Whether the model can predict release for meaningfully distinct formulation designs [9, 20]	More relevant to formulation screening than random-split validation.
Leave-one-drug-out validation	Test extrapolation to drugs not seen during training	Hold out all records for a given active ingredient	Whether learned polymer and morphology effects transfer across drug properties	Defines the boundary between interpolation within known systems and higher-risk extrapolation.
External laboratory validation	Assess transferability across laboratories, polymer suppliers, and release-test conventions	Validate using records from independent sources, suppliers, or manufacturing processes [6, 16]	Whether predictions remain stable outside the training-data environment	Essential before the model is used for high-impact development or regulatory support.
Interpretability analysis	Convert predictions into formulation insight	SHAP summary plots, dependence analysis, interaction analysis, and feature-ablation checks [8]	Whether burst, lag, and sustained release are driven by plausible formulation factors	Supports hypothesis generation but should be checked against drug-delivery mechanisms.
Prospective formulation utility	Test whether model-guided selection improves real development decisions	Compare model-guided candidate selection against conventional empirical screening	Whether the model reduces unnecessary formulation iteration while preserving release-profile quality [8, 25]	Establishes practical value as a decision-support tool rather than only a statistical predictor.
Quality-by-Design alignment	Link controllable attributes to target release-profile regions	Map critical material attributes and morphology targets to acceptable predicted release space [24]	Which polymer, loading, and morphology variables define the conceptual design space	Supports structured risk assessment and rational formulation optimization.
Regulatory-readiness boundary	Prevent premature use of unvalidated predictions as evidence	Document feature definitions, missing-data handling, validation limits, external testing, and prospective confirmation	Whether the model is transparent, reproducible, and bounded by known limitations	Positions the framework as supportive evidence only after validation, not as a replacement for experimental release testing.

Limitations

Batch-to-Batch Variability and Scale-Up Effects

A major limitation is that the model may not capture batch-to-batch polymer variability, residual solvent differences, mixing history, or scale-up effects that alter microsphere morphology after the formulation has nominally the same composition [6]. Polymer source has been shown to affect in-vitro drug release from PLGA microspheres, so a model trained without supplier or lot descriptors may miss important sources of variability [16]. Manufacturing scale can also change particle-size distribution, porosity, and internal structure, which means periodic recalibration with process-specific data would be necessary [5]. The model should therefore be deployed as a continuously updated decision-support tool rather than a fixed universal predictor.

Limited Chemical Diversity in Training Data

Another limitation is that available data are likely to be concentrated in PLGA and PLA systems with small-molecule drugs, making predictions less reliable for novel polymers, biologics, peptides, or unusual drug-polymer interaction regimes [2]. Machine learning models trained on narrow chemical domains can appear plausible within familiar formulation space but extrapolate poorly when polymer chemistry or drug physical behavior changes substantially [20]. Broader polymeric delivery studies, including nanofibers and nanoparticle datasets, may inform feature design but cannot automatically substitute for LAI

microsphere evidence. Future model development should therefore expand chemical diversity while preserving careful annotation of polymer, drug, morphology, and release-testing conditions.

Conclusion

A machine learning model for long-acting injectable microspheres could integrate polymer degradation characteristics, drug-loading variables, and microsphere morphology descriptors to predict the full in-vitro release profile. By treating release as a curve rather than a single endpoint, the framework could represent burst, lag, and sustained-release phases within one decision-support system. When suitable bridging evidence is available, the same framework could be extended conceptually toward pharmacokinetic tendency prediction.

The main strength of this approach is its ability to connect formulation inputs with mechanistically meaningful release behavior. A full-profile prediction model could help formulators understand whether a candidate formulation is likely to fail because of excessive burst release, prolonged lag, or insufficient sustained release. Interpretable outputs could further support rational adjustment of polymer grade, loading level, particle-size distribution, or porosity target. This makes the framework compatible with model-informed Quality-by-Design development.

Important challenges remain before such a model could be used confidently in practice. LAI formulation datasets are often small, heterogeneous, and incomplete, particularly with respect to morphology and degradation measurements. Batch-to-batch variability, supplier differences, and scale-up effects may also limit transferability across laboratories and manufacturing sites. Prospective validation in a realistic formulation workflow would therefore be essential.

The field would benefit from collaborative efforts to build shared, well-annotated LAI formulation databases containing polymer characteristics, drug properties, morphology metrics, release profiles, and pharmacokinetic context. Such databases would make it possible to evaluate predictive models more rigorously and compare model architectures on common evidence. Pilot deployments in industrial formulation groups could then test whether model-guided design reduces empirical iteration while preserving formulation quality. With careful validation and transparent limitations, predictive modeling could become a practical accelerator for long-acting injectable development.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. O'Brien MN, Jiang W, Wang Y, Loffredo DM. Challenges and opportunities in the development of complex generic long-acting injectable drug products. J Control Release. 2021;336:144-58.
2. Park K. PLGA-based long-acting injectable (LAI) formulations. J Control Release. 2025;382:113758.
3. Bode C, Kranz H, Fivez A, Siepmann F, Siepmann J. Often neglected: PLGA/PLA swelling orchestrates drug release: HME implants. J Control Release. 2019;306:97-107.
4. de Azevedo CR, von Stosch M, Costa MS, Ramos AM, Cardoso MM, Danhier F, et al. Modeling of the burst release from PLGA micro- and nanoparticles as function of physicochemical parameters and formulation characteristics. Int J Pharm. 2017;532(1):229-40.
5. Park K, Otte A, Sharifi F, Garner J, Skidmore S, Park H, et al. Formulation composition, manufacturing process, and characterization of poly(lactide-co-glycolide) microparticles. J Control Release. 2021;329:1150-61.
6. Park K, Skidmore S, Hadar J, Garner J, Park H, Otte A, et al. Injectable, long-acting PLGA formulations: Analyzing PLGA and understanding microparticle formation. J Control Release. 2019;304:125-34.
7. Busatto C, Pesoa J, Helbling I, Luna J, Estenoz D. Effect of particle size, polydispersity and polymer degradation on progesterone release from PLGA microparticles: Experimental and mathematical modeling. Int J Pharm. 2018;536(1):360-9.
8. Bannigan P, Bao Z, Hickman RJ, Aldeghi M, Häse F, Aspuru-Guzik A, et al. Machine learning models to accelerate the design of polymeric long-acting injectables. Nat Commun. 2023;14(1):35.
9. Deng J, Ye Z, Zheng W, Chen J, Gao H, Wu Z, et al. Machine learning in accelerating microsphere formulation development. Drug Deliv Transl Res. 2023;13(4):966-82.
10. Bao Z, Kim J, Kwok C, Le Devedec F, Allen C. A dataset on formulation parameters and characteristics of drug-loaded PLGA microparticles. Sci Data. 2025;12(1):364.
11. Bao Z, Kim J, Le Devedec F, Clasky A, Allen C. Polymer microparticles in an evolving drug delivery landscape: challenges and the role of machine learning. Int J Pharm. 2025;682:125906.

12. Robles KN, Samad MD. Predicting early and complete drug release from long-acting injectables using explainable machine learning. arXiv. 2026;2601.02265.
13. Xiong P, Zhu J, Zhong H, Wang W, Wang N, Lu T, et al. FormulationLAI: A physiology-based machine learning framework for accelerated development of long-acting injectable formulations. J Control Release. 2025:114418.
14. Nakhla DS, Mekkawy AI, Naguib YW, Silva AD, Gao D, Kim JA, et al. Injectable long-acting ivacaftor-loaded poly(lactide-co-glycolide) microparticle formulations for the treatment of cystic fibrosis: in vitro characterization and in vivo pharmacokinetics in mice. Int J Pharm. 2024;650:123693.
15. Rapier CE, Shea KJ, Lee AP. Investigating PLGA microparticle swelling behavior reveals an interplay of expansive intermolecular forces. Sci Rep. 2021;11(1):14512.
16. Wan B, Andhariya JV, Bao Q, Wang Y, Zou Y, Burgess DJ. Effect of polymer source on in vitro drug release from PLGA microspheres. Int J Pharm. 2021;607:120907.
17. Sun Z, Gu X, Hao T, Liu J, Gao R, Li Y, et al. Intra-articular injection PLGA blends sustained-release microspheres loaded with meloxicam: preparation, optimization, evaluation in vitro and in vivo. Drug Deliv. 2022;29(1):3317-27.
18. Avendaño-Godoy J, Miranda A, Mennickent S, Gómez-Gaete C. Intramuscularly administered PLGA microparticles for sustained release of rivastigmine: In vitro, in vivo and histological evaluation. J Pharm Sci. 2023;112(12):3175-84.
19. Noorain L, Nguyen V, Kim HW, Nguyen LT. A machine learning approach for PLGA nanoparticles in antiviral drug delivery. Pharmaceutics. 2023;15(2):495.
20. Aghajanpour S, Amiriara H, Esfandyari-Manesh M, Ebrahimnejad P, Jeelani H, Henschel A, et al. Utilizing machine learning for predicting drug release from polymeric drug delivery systems. Comput Biol Med. 2025;188:109756.
21. Sun Y, Qin S, Li Y, Hasan N, Li YV, Liu J. Machine learning integrated with in vitro experiments for study of drug release from PLGA nanoparticles. Sci Rep. 2025;15(1):4218.
22. Rezvantalab S, Mihandoost S, Rezaiee M. Machine learning assisted exploration of the influential parameters on the PLGA nanoparticles. Sci Rep. 2024;14(1):1114.
23. Catapano AF, Zheng L, Yuan X, Yuan K. Machine learning for predicting drug release behavior of PLGA microspheres. In: AMIA Annu Symp Proc. 2025;2024:212-7.
24. Longre S, Rana D, Rangra S, Jindal AB, Salave S, Vitore J, et al. Quality-by-design based development of doxycycline hyclate-loaded polymeric microspheres for prolonged drug release. AAPS PharmSciTech. 2024;25(3):49.
25. Seegobin N, Abdalla Y, Li G, Murdan S, Shorthouse D, Basit AW. Optimising the production of PLGA nanoparticles by combining design of experiment and machine learning. Int J Pharm. 2024;667:124905.