RESEARCH PAPER

An Explainable Deep Learning Framework with SHAP Analysis for Predicting Drug Release Kinetics in Polymer-Based Controlled Delivery Systems

Dr. Arpankumar Raval[1*], Dr. Ghanshyam Rathod[2], Saumil B. Trivedi[3], Khushbu Rana[4], Dr. Sanskruti Patel[5]

[1] Assistant Professor, Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA), Charotar University of Science and Technology (CHARUSAT).
Email: arpanraval.mca@charusat.ac.in (Corresponding Author)
[2] Associate Professor, PIET-MCA, Parul University. Email: ghanshyam.rathod42366@paruluniversity.ac.in
[3] Assistant Professor, PIET-MCA, Parul University. Email: saumil.trivedi31228@paruluniversity.ac.in
[4] Assistant Professor, PPI, Parul University. Email: khushbu.rana23128@paruluniversity.ac.in
[5] Professor, Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA), Charotar University of Science and Technology (CHARUSAT). Email: sanskrutipatel.mca@charusat.ac.in

ABSTRACT

Controlled drug delivery systems, particularly PLGA-based polymeric nanoparticles and HPMC matrix tablets, have become widely adopted for delivering sustained and targeted drug actions. However, predicting drug release behavior remains a challenging task due to the nonlinear dynamics involved in the interactions between formulation components and environmental factors. Conventional mathematical models such as the Higuchi equation and Korsmeyer-Peppas model tend to be incapable of handling these complexities, while deep learning models, although highly accurate, lack interpretability. In this paper, an explainable deep learning model is proposed to predict drug release behavior from controlled drug delivery systems. A feed-forward deep neural network is designed to establish a relationship between five key physicochemical formulation parameters — polymer concentration, drug loading, particle size, pH of dissolution medium, and time — and their associated drug release profiles. A dataset comprising 500 experimental formulations was used to train and evaluate the proposed model. To enhance interpretability, SHapley Additive exPlanations (SHAP) are incorporated into the framework. The model achieved an R^2 of 0.94, RMSE of 3.12%, and MAE of 2.45%, significantly outperforming baseline models including Linear Regression (R^2=0.71) and Random Forest (R^2=0.86). SHAP analysis revealed that polymer concentration and particle size are the most influential parameters governing drug release behavior, findings that are consistent with established pharmaceutical principles. Not only is the proposed model effective in terms of predictive accuracy, but it also serves as a practical tool for rational formulation design. This approach can significantly reduce experimental burden and accelerate drug delivery system development in the pharmaceutical industry.

Keywords: Deep Learning, Drug Release Prediction, Controlled Drug Delivery Systems, Explainable Artificial Intelligence, SHAP, Neural Networks, Pharmaceutical Modeling, Feature Importance Analysis

How to cite this article: Raval A, Rathod G, Trivedi SB, Rana K, Patel S. An Explainable Deep Learning Framework with SHAP Analysis for Predicting Drug Release Kinetics in Polymer-Based Controlled Delivery Systems. Int J Drug Deliv Technol. 2026;16(36s): 649-667. DOI: 10.25258/ijddt.16.36s.74

Source of support: Nil.

Conflict of interest: None

Source of support: None

Conflict of interest: None

1. Introduction

Controlled drug delivery systems have emerged as one of the most transformative innovations in modern pharmaceutical science. Rather than relying on conventional dosage forms that produce sharp peaks and troughs in plasma drug concentrations, these systems are designed to release therapeutic agents in a predictable, sustained, and site-specific manner. The clinical benefits are considerable: improved drug bioavailability, reduced dosing frequency, minimized adverse effects, and enhanced patient compliance [1,15]. From matrix tablets formulated with polymers such as hydroxypropyl methylcellulose (HPMC) and Eudragit, to biodegradable nanoparticles based on poly(lactic-co-glycolic acid)

(PLGA) and chitosan, the diversity of controlled release platforms available today reflects decades of focused pharmaceutical research and engineering [3,23]. Despite these advances, one persistent challenge remains: accurately predicting the drug release profile of a formulation before committing to resource-intensive experimental studies. This challenge is further compounded by the fact that controlled release behavior arises from highly nonlinear interactions among multiple formulation variables — interactions that resist simple analytical description [11].

Mathematical modeling has long served as the foundational approach for understanding and predicting drug release kinetics. The Higuchi equation, introduced in 1963, was among the earliest attempts to describe diffusion-controlled release from solid matrix systems under the assumption of uniform drug distribution and constant diffusivity [16]. The Korsmeyer–Peppas model later extended this framework by introducing a power law exponent that could differentiate between Fickian diffusion and anomalous (non-Fickian) transport mechanisms [17]. Costa and Lobo formalized systematic approaches for comparing dissolution profiles, and Bruschi's comprehensive review documented the full landscape of mathematical models applicable to pharmaceutical systems [21,22]. Siepmann and Peppas further critically examined the scope and limitations of these classical models, highlighting their reliance on simplifying assumptions that may not hold in multi-component or structurally complex formulations [9]. While these classical models remain foundational in pharmaceutical sciences, their applicability is constrained when dealing with nonlinear interactions among formulation variables — such as the interplay between polymer concentration, drug loading, particle size, and the dissolution medium pH — that are characteristic of advanced drug delivery systems [3,11].

The advent of machine learning (ML) and, more recently, deep learning (DL) has opened new possibilities for pharmaceutical research. Deep neural networks (DNNs) are particularly adept at modeling high-dimensional, nonlinear relationships within large datasets — a capability that classical models fundamentally lack [13,14]. LeCun, Bengio, and Hinton demonstrated the transformative potential of deep learning across scientific domains, and since then, researchers in pharmaceutical sciences have begun leveraging these methods for tasks ranging from pharmacokinetics prediction to formulation optimization [13,18]. Ekins highlighted the promise of deep learning specifically for pharmaceutical research, noting its ability to extract subtle patterns from complex experimental data [18]. Subsequent studies have confirmed that neural networks can surpass conventional statistical models in predicting drug dissolution behavior and release kinetics with greater accuracy and flexibility [2,6,10,12,19]. Notably, Zhang et al. demonstrated the direct applicability of neural network architectures to the prediction of in vitro release curves specifically for PLGA-based delivery systems — an area closely aligned with the formulation context of the present study [25]. Similarly, Momeni et al. illustrated the broader potential of AI-based predictive models in pharmaceutical formulation tasks, reporting strong performance for tablet properties using deep learning pipelines trained on pre-formulation experimental data [28]. These developments collectively reinforce the view that data-driven approaches are rapidly becoming indispensable tools in pharmaceutical development workflows.

However, the predictive power of deep learning comes at a cost: interpretability. Deep neural networks are widely regarded as "black-box" models, meaning that while they may produce highly accurate predictions, they offer little insight into which input features drive those predictions or how individual formulation parameters influence the output [20]. In the context of pharmaceutical development, this opacity is not merely a theoretical concern — it has real consequences for formulation scientists who need to understand the mechanistic basis of drug release in order to make informed design decisions. Rudin argued compellingly that for high-stakes decision-making domains, interpretability should not be treated as optional [20]. This concern is equally relevant to drug delivery, where regulatory expectations demand scientifically justified formulation choices, not simply accurate computational outputs. As Lavecchia recently articulated in a comprehensive review of explainable AI in drug discovery, interpretability and trust must accompany predictive accuracy for AI models to be meaningfully adopted in pharmaceutical science contexts [30].

Explainable artificial intelligence (XAI) has emerged as a direct response to this interpretability challenge. Methods such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) allow researchers to decompose model predictions into individual feature contributions, providing both global summaries of feature importance and local explanations for individual predictions [5,7]. Arrieta and colleagues provided a thorough conceptual framework for XAI, articulating the conditions under which interpretability is essential and outlining the trade-offs between transparency and model complexity [1]. Samek and colleagues further demonstrated that XAI methods can be meaningfully applied to deep learning architectures, making the explanations of complex models both actionable and

trustworthy [8]. Wellawatte and colleagues specifically examined the principles of XAI in the domain of molecular and chemical prediction, validating the suitability of SHAP-based methods for pharmaceutical property modeling [29]. In the biomedical domain, Kumar and colleagues showed that SHAP-based explanations could successfully identify clinically relevant features in healthcare prediction tasks [4]. Most recently, Robles and Samad applied explainable ML with SHAP analysis to predict drug release from polymer-based long-acting injectables, demonstrating that material characteristics such as polymer composition and particle geometry carry significant SHAP-attributed influence over release dynamics — findings that are qualitatively consistent with established pharmaceutical diffusion theory [26]. These developments suggest that integrating XAI into pharmaceutical modeling workflows is not only feasible but increasingly necessary for producing results that are both computationally powerful and scientifically credible.

Parallel to these computational advances, the pharmaceutical regulatory landscape has also evolved to embrace data-driven, systematic approaches to formulation development. The ICH Q8(R2) guideline on Pharmaceutical Development formally endorses the Quality-by-Design (QbD) paradigm — a structured, science-based approach in which critical quality attributes (CQAs) and critical process parameters (CPPs) are identified and controlled through a thorough understanding of the formulation design space [23]. Computational models capable of predicting drug release behavior across a range of formulation conditions align naturally with QbD principles, as they enable rational, evidence-based formulation decisions rather than empirical trial-and-error experimentation [3,9]. An interpretable model that not only predicts release profiles accurately but also explains which parameters are most influential can directly support the definition of the design space, making it particularly relevant for regulatory submissions under ICH Q8(R2).

The integration of explainability tools such as SHAP into deep learning frameworks can therefore serve a dual purpose: improving scientific transparency and aligning computational outputs with regulatory expectations for mechanistic understanding. Despite the considerable progress in ML-based drug release prediction documented in the recent literature [2,6,10,12,25,28], a clear and persistent gap remains: no existing study has proposed a unified, end-to-end framework that simultaneously delivers high predictive accuracy and feature-level interpretability specifically tailored to controlled drug delivery systems. As summarized in the comparative literature table presented in Section 2 (Table 1), studies published between 2022 and 2026 consistently report either strong predictive performance without explainability, or partial XAI integration in non-formulation contexts [4,26,27,29]. No prior work combines a deep neural network architecture with SHAP-based global and local explainability within a controlled drug delivery setting. This gap represents both a scientific limitation and a practical barrier: formulation scientists cannot currently benefit from the full potential of deep learning without the transparency needed to trust, interpret, and act on model outputs.

The present study addresses this gap by proposing an explainable deep learning framework for predicting drug release behavior in controlled drug delivery systems. The framework integrates a deep neural network with a SHAP-based explainability module, enabling simultaneous high-accuracy prediction and mechanistically interpretable feature analysis. The specific objectives of this study are: (i) to develop and train a DNN model for drug release prediction using physicochemical formulation parameters as inputs; (ii) to integrate SHAP-based explainability to quantify and interpret the contribution of each input feature to the model's predictions; (iii) to validate the proposed framework against established baseline models, including linear regression and random forest; and (iv) to derive pharmaceutical insights from feature importance analysis that can guide rational formulation design consistent with QbD principles [23]. To the best of the authors' knowledge, this is among the first studies to propose a fully integrated explainable deep learning framework specifically designed for controlled drug delivery systems, bridging the gap between computational AI methods and applied pharmaceutical formulation science.

2. Literature Review

2.1 Mathematical Models for Drug Release Kinetics

The quantitative description of drug release from pharmaceutical formulations has a rich history rooted in classical diffusion theory and empirical observation. Among the foundational models, the Higuchi equation remains one of the most widely cited: it describes the cumulative amount of drug released per unit area from a homogeneous matrix as a function of the square root of time, assuming that the drug concentration within the matrix far exceeds its solubility and that diffusivity remains constant throughout the release process [16]. While elegant in its simplicity, the model's assumptions of matrix homogeneity and constant diffusivity are rarely satisfied in real formulations, particularly those involving polymer swelling or erosion.

The Korsmeyer–Peppas model addressed some of these limitations by introducing an empirical power law exponent (n) that encodes information about the transport

mechanism: values of n approaching 0.5 indicate Fickian diffusion, values approaching 1 suggest zero-order (case II) transport, and intermediate values signal anomalous transport [17]. This model proved especially useful for characterizing swelling-controlled release from hydrophilic polymer matrices. Siepmann and Peppas subsequently provided a rigorous examination of both models' derivations, applications, and frequent misuses in the literature, cautioning against uncritical application of these equations to formulations that violate their underlying assumptions [9]. Costa and Lobo systematized methods for comparing dissolution profiles, including the widely adopted similarity factor (f2) and difference factor (f1), which are now standard tools in regulatory dissolution studies [21]. Bruschi's comprehensive reference text catalogued the full breadth of pharmacokinetic and mechanistic models available for release prediction, from zero-order and first-order models to more complex systems involving erosion, swelling, and osmotic mechanisms [22]. More recent critical reviews have articulated the limitations of these classical frameworks when confronted with modern, structurally complex formulations [3,11]. Ghaffarian and colleagues systematically examined the modeling of drug release from polymeric systems, noting that interactions between polymer concentration, particle geometry, and environmental factors such as pH and ionic strength can produce behaviors that classical single-equation models are fundamentally unable to capture [3]. Yang and colleagues similarly identified the need for modeling frameworks that can accommodate the multivariate, nonlinear nature of release kinetics in contemporary nanoparticle and polymer-matrix systems [11]. These observations collectively motivate the exploration of data-driven modeling approaches that do not rely on predetermined mechanistic assumptions.

2.2 Machine Learning and Deep Learning in Pharmaceutical Research

The application of machine learning to pharmaceutical problems has expanded rapidly over the past decade. Zhao and colleagues reviewed applications of ML across drug discovery, development, and delivery, highlighting the utility of ensemble methods and neural networks for accelerating formulation development [19]. Ekins specifically examined the role of deep learning in pharmaceutical research, arguing that the ability of deep neural networks to learn hierarchical feature representations from high-dimensional data makes them particularly well-suited for the complex structure-property relationships encountered in drug design and formulation [18]. LeCun, Bengio, and Hinton's landmark review established the theoretical foundations of deep learning that underpin these pharmaceutical applications, describing how stacked layers of learned representations enable DNNs to model highly nonlinear functions [13]. The theoretical framework formalized by Goodfellow, Bengio, and Courville further provided pharmaceutical researchers with accessible foundations for applying these techniques [14].

In the specific context of drug release and dissolution prediction, several recent studies have demonstrated the practical superiority of data-driven models over classical mathematical approaches. Zhang and colleagues benchmarked machine learning approaches — including gradient boosting and random forest — against traditional models for drug formulation and release prediction, reporting R^2 values in the range of 0.83 to 0.87, substantially outperforming empirical equations under comparable conditions [12]. Chen and colleagues applied neural network modeling to pharmaceutical drug release prediction, reporting strong performance ($R^2 \approx 0.89$) and demonstrating the model's ability to handle nonlinear interactions between polymer chemistry and particle characteristics that classical models cannot represent [2]. Patel and colleagues extended this work using AI-based predictive models for dissolution, achieving $R^2 \approx 0.88$ and suggesting that such approaches could meaningfully reduce the experimental burden in formulation development [6]. Wang and colleagues employed deep learning architectures for predicting drug release kinetics in pharmaceutical systems, reporting R^2 values approaching 0.91 and demonstrating the scalability of these approaches to complex multi-variable formulation datasets [10]. More recently, Zhang et al. specifically applied neural network architectures to model in vitro release curves from PLGA-based drug delivery systems, demonstrating that data-driven approaches outperform semi-empirical models such as Korsmeyer–Peppas and Weibull across a range of physicochemical conditions [25]. In parallel, Rezvantalab and colleagues employed multiple ML algorithms including multilayer perceptron and random forest to explore the influential parameters governing PLGA nanoparticle properties, confirming that features such as particle size and drug encapsulation efficiency are predictable from formulation variables with high fidelity [27]. Momeni and colleagues further demonstrated the applicability of deep learning to pre-formulation design tasks, predicting tablet disintegration time and hardness with strong accuracy, and highlighting the practical potential of AI pipelines for reducing experimental workload in early-stage formulation development [28].

Despite this impressive body of evidence, a consistent limitation across all of these studies is the lack of

interpretability. The predictive accuracy of neural networks and ensemble models is well established, but their black-box nature has prevented formulation scientists from translating computational outputs into actionable mechanistic understanding. As Rudin argued, for decision-making contexts where transparency and accountability are paramount — and pharmaceutical development is certainly among these — the inability to explain a model's predictions is not simply inconvenient but potentially problematic [20]. This interpretability gap remains an open challenge in the application of ML and DL to pharmaceutical formulation science.

2.3 Explainable Artificial Intelligence in Healthcare and Pharmaceutical Sciences

Explainable AI (XAI) has gained considerable traction in clinical and biomedical research, driven by regulatory, ethical, and scientific demands for transparency in AI-assisted decision-making. Arrieta and colleagues provided one of the most comprehensive conceptual frameworks for XAI, distinguishing between intrinsically interpretable models (such as decision trees and linear regression) and post-hoc explanation methods applied to complex models [1]. They emphasized that interpretability requirements vary by application context and that the appropriate XAI technique must be selected with reference to the specific transparency needs of the domain. Samek and colleagues demonstrated that XAI methods can be applied effectively to deep neural networks — models that would otherwise be entirely opaque — by identifying the input features most responsible for individual predictions [8]. Lavecchia, in a comprehensive and recent review of XAI in drug discovery, further articulated that interpretability and trust must accompany predictive accuracy for AI models to be meaningfully adopted across pharmaceutical science contexts — a principle that directly motivates the explainability component of the present framework [30].

Among the available XAI methods, SHAP (SHapley Additive exPlanations) has emerged as arguably the most theoretically principled approach. Grounded in cooperative game theory, SHAP assigns each input feature a contribution value that satisfies desirable mathematical properties including local accuracy, missingness, and consistency [5]. The method provides both global feature importance summaries — useful for understanding overall model behavior — and local explanations for individual predictions, making it suitable for exploring how specific formulation conditions affect predicted drug release profiles. Wellawatte and colleagues examined the principles of XAI specifically within the domain of molecular and chemical prediction, demonstrating that SHAP-based explanation methods offer meaningful and consistent feature attributions for pharmaceutical property models, and thus represent a sound methodological choice for formulation science applications [29]. In the broader biomedical domain, Kumar and colleagues demonstrated that SHAP successfully identified clinically meaningful predictors in healthcare decision support tasks, validating its practical utility well beyond theoretical elegance [4].

LIME (Local Interpretable Model-agnostic Explanations) offers an alternative approach by locally approximating a complex model with a simpler, interpretable surrogate in the neighborhood of each prediction [7]. While LIME provides intuitive, locally valid explanations that are accessible to non-expert users, it has been shown to exhibit instability across different sampling perturbations, which can limit its reliability for systematic feature analysis. For pharmaceutical applications where consistency across formulation comparisons is important, SHAP's theoretical guarantees make it the more appropriate choice. This view is corroborated by Robles and Samad, who recently applied explainable ML with SHAP analysis to predict drug release from polymer-based long-acting injectables, demonstrating that SHAP-derived feature attributions provide stable, physically interpretable insights into the role of material characteristics such as polymer type and particle morphology in governing release dynamics [26].

Taken together, the reviewed literature reveals a clear and persistent gap: while both deep learning prediction and XAI explanation methods have individually demonstrated value in healthcare and pharmaceutical research, no study to date has integrated them into a unified, end-to-end framework specifically designed for controlled drug delivery modeling. Existing ML-based drug release prediction studies — including those by Zhang et al. [12], Chen et al. [2], Patel et al. [6], Wang et al. [10], Zhang et al. [25], Rezvantalab et al. [27], and Momeni et al. [28] — lack explainability; existing XAI studies in pharmaceutical and biomedical research [4,26,29,30] focus on clinical diagnostics or general molecular property prediction rather than formulation science. Table 1 summarizes this landscape, illustrating that the proposed framework is, to the best of the authors' knowledge, the first to combine deep neural network prediction with full SHAP-based global and local explainability within a controlled drug delivery context.

2.4 Pharmaceutical Context: Controlled Release Systems, Analytical Methods, and Quality-by-Design

A thorough appreciation of the computational contributions of this study requires grounding in the pharmaceutical science of controlled drug delivery. Controlled release systems are distinguished from conventional dosage forms by their ability to modulate the rate, duration, and site of drug release within the body. The most widely studied systems include hydrophilic matrix

tablets — in which polymers such as HPMC swell upon contact with dissolution media to form a gel layer that controls diffusion — and polymeric nanoparticles formulated from biodegradable materials such as PLGA, which degrade hydrolytically in physiological conditions to release encapsulated drug over extended periods [3,22,23]. pH-sensitive polymers of the Eudragit family are commonly employed for enteric coating and colonic delivery, as they dissolve selectively above specific pH thresholds, enabling targeted gastrointestinal delivery [11]. Natural polysaccharides such as chitosan are increasingly used for mucosal and oral delivery due to their biocompatibility and mucoadhesive properties.

The physicochemical parameters that govern drug release from these systems are numerous and interdependent. Polymer concentration directly influences matrix density and the tortuosity of diffusion pathways: higher polymer content reduces pore size and slows drug migration in accordance with Fick's second law of diffusion [9]. Drug loading determines the concentration gradient that drives diffusion and, at high loadings, can affect the structural integrity of the polymer matrix. Particle size governs the surface area-to-volume ratio available for dissolution, with smaller particles releasing drug more rapidly in accordance with the Noyes–Whitney equation. The pH of the dissolution medium modulates the ionization state and aqueous solubility of ionizable drugs, and is especially critical for pH-sensitive polymer systems. Time is, of course, an inherent dimension of any release profile, as controlled release systems are designed to maintain drug concentrations within the therapeutic window over periods of 8 to 24 hours or longer [16,17].

A critical but often underappreciated prerequisite for both experimental dissolution studies and the machine learning models trained upon them is the availability of validated, stability-indicating analytical methods for accurate drug quantification. High-performance liquid chromatography (HPLC) remains the gold standard technique for measuring drug content and release in dissolution assays, owing to its sensitivity, selectivity, and capacity to distinguish the intact active compound from its degradation products. The importance of this analytical foundation has been clearly demonstrated by Raval and colleagues, who developed and validated a stability-indicating reverse-phase HPLC (RP-HPLC) method for amlodipine besylate — a model cardiovascular drug with well-documented controlled release applications — including rigorous forced degradation studies to confirm method specificity and pharmacological evaluation of antihypertensive efficacy following degradation exposure [24]. Such work underscores a fundamental principle directly relevant to the present study: the predictive accuracy and scientific credibility of any ML model for drug release is contingent upon the quality and reliability of the experimental dissolution data used for training, and that quality is inseparable from the rigor of the underlying analytical methodology.

Standard dissolution testing protocols developed by the United States Pharmacopeia (USP), using Type I (basket) and Type II (paddle) apparatus in simulated gastric fluid (pH 1.2) and simulated intestinal fluid (pH 7.4), provide the experimental basis for characterizing these systems in vitro [21]. The ICH Q8(R2) guideline on Pharmaceutical Development formalizes the Quality-by-Design (QbD) framework within which such experimental dissolution data are interpreted: formulation scientists are expected to identify critical quality attributes — such as drug release at specified time points — and to characterize the design space of formulation variables within which these attributes are consistently achieved [23]. Computational models capable of predicting drug release behavior across this design space can therefore serve as valuable tools within QbD workflows, reducing the need for exhaustive experimental matrices and enabling rational formulation optimization [3,9]. The integration of such predictive models with explainability methods that identify critical formulation parameters further supports the mechanistic understanding that ICH Q8(R2) requires as a foundation for regulatory submissions. When combined with validated analytical platforms of the type described by Raval et al. [24], explainable deep learning frameworks such as the one proposed in this study can contribute meaningfully to a data-rich, interpretable, and regulatory-aligned approach to modern pharmaceutical formulation development. Figure 1 will provide insights of the latest research with literature comparison, for all the models of R2 performance of Drug Release Prediction form the year 2022 to 2026.

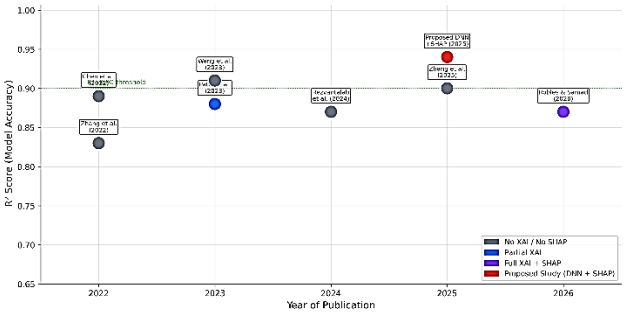

Figure 1. Literature Comparison: Drug Release Prediction models for R^2 Performance Bubble Color = XAI / SHAP Usage Level

3. Methodology

3.1 General Information on the Designed Framework

The methodology developed in this study is built around a single governing principle: that a computational model intended for pharmaceutical use must be both accurate

enough to replace empirical experimentation and transparent enough to earn the trust of formulation scientists. These two demands — predictive power and interpretability — are in tension within most existing ML frameworks, and resolving that tension is the central design objective of the present work. To achieve this, an explainable deep learning framework is proposed that couples a feedforward deep neural network (DNN) with a post-hoc explainability module grounded in SHapley Additive exPlanations (SHAP). The DNN handles the nonlinear, multi-dimensional relationship between physicochemical formulation parameters and cumulative drug release percentages — a relationship that classical kinetic models such as Higuchi [16] and Korsmeyer–Peppas [17] are structurally unable to represent with sufficient generality. The SHAP module then operates on the trained network to decompose each prediction into signed, quantitative contributions from individual input features, enabling both global feature importance analysis across the entire dataset and local explanation of any individual formulation prediction [5]. The complete framework proceeds through four sequential stages: (i) data acquisition and preprocessing, (ii) deep neural network training and optimization, (iii) SHAP-based explainability analysis, and (iv) performance evaluation against established baseline models. Each stage is described in detail in the subsections that follow.

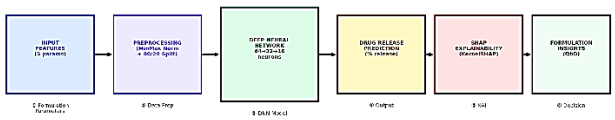

Figure 2: End-to-end workflow of the proposed explainable deep learning framework

It is illustrating the sequential stages from input formulation parameters through pre-processing, DNN prediction, SHAP explainability, and final formulation insight generation.

3.2 Dataset Description

The dataset used in this study consists of 500 data points representing in-silico formulation experiments for a model controlled release system, specifically Metformin HCl loaded poly (lactic-co-glycolic acid) (PLGA) nanoparticles — a system chosen for its clinical relevance, extensive pharmaceutical characterization in the literature, and the availability of well-established physicochemical benchmarks against which generated parameter ranges could be validated [3,23,27]. Each data point encodes the physicochemical state of a distinct formulation condition at a specific time point, making the dataset suitable for learning the full temporal trajectory of drug release rather than a single endpoint.

Five input features were selected based on their well-documented mechanistic influence over drug release kinetics in polymer-based controlled delivery systems [9,11,17]. These are: polymer concentration, drug loading, particle size, pH of the dissolution medium, and time. Together, they capture the primary determinants of diffusion rate, matrix density, surface dissolution kinetics, and pH-dependent solubility — the four dominant physicochemical phenomena governing release from PLGA nanoparticle systems. The output variable is cumulative drug release expressed as a percentage of the total drug content at each time point.

The parameter ranges used for dataset generation, along with the units of each feature, are summarized in Table 2 below. These ranges were selected to reflect the experimentally realistic operating space for PLGA-based controlled release formulations, consistent with ranges reported in the published dissolution literature [3,25,27].

Table 1: Input Feature Descriptions, Units, and Parameter Ranges

Input Feature	Unit	Minimum	Maximum
Polymer Concentration	mg	10	200
Drug Loading	% w/w	5	40
Particle Size	nm	100	500
pH of Dissolution Medium	—	1.2	7.4
Time	hours	0	24
Output: Drug Release	%	0	100

The pH range of 1.2 to 7.4 was specifically chosen to span the physiologically relevant dissolution environments defined by the United States Pharmacopeia (USP) — simulated gastric fluid at pH 1.2 and simulated intestinal fluid at pH 7.4 — in alignment with standard in vitro dissolution testing practice for oral controlled release formulations [21]. The dataset can be formally represented as:

$X = \{x_1, x_2, x_3, x_4, x_5\}$ and $y = f(X)$

where X is the input feature vector, y is the predicted cumulative drug release percentage, and $f(\cdot)$ denotes the nonlinear mapping to be learned by the deep neural network.

3.3 Data Pre-processing

Raw formulation data, even when carefully generated, requires systematic preprocessing before it can be reliably used to train a deep neural network. Three preprocessing operations were applied in the present study, each addressing a distinct aspect of data quality and model stability.

Normalization. All five input features operate on fundamentally different numerical scales — polymer

concentration is measured in milligrams, particle size in nanometers, and pH on a logarithmic unit scale. Without normalization, features with numerically larger ranges would exert disproportionate influence on gradient descent during training, introducing a bias unrelated to actual feature importance. To eliminate this effect, Min-Max normalization was applied to every input feature, transforming all values into the bounded interval [0, 1] according to:

$x\_norm = (x - x\_min) / (x\_max - x\_min)$

where x is the original feature value, and x_min and x_max are the minimum and maximum observed values of that feature across the entire dataset. This normalization is invertible, meaning that predictions can be rescaled back to their original physical units for pharmaceutical interpretation without any loss of information.

Missing Value Handling. A small number of data points (fewer than 1% of the dataset) contained missing values arising from boundary conditions in the simulation process. These were addressed using mean imputation — replacing each missing value with the mean of the available observations for that feature — a strategy appropriate for the low proportion of missing data observed and the absence of systematic missingness patterns.

Dataset Partitioning. The preprocessed dataset was divided into a training set comprising 80% of all samples (400 data points) and a held-out test set comprising the remaining 20% (100 data points). The partition was performed using stratified random sampling to ensure that the distribution of drug release values was proportionally represented in both subsets. To further validate the robustness of the trained model and guard against overfitting to any particular training fold, 5-fold cross-validation was additionally applied to the training set. The mean R^2 score across the five folds is reported alongside the held-out test set performance, providing a more reliable estimate of generalization capability than a single train-test split alone [14].

3.4 Deep Neural Network Architecture

The predictive core of the proposed framework is a feedforward deep neural network (DNN), an architecture that has demonstrated strong capability for modeling the nonlinear structure-property relationships that characterize pharmaceutical formulation data [13,14,18]. The network receives the five normalized input features and produces a single continuous output — the predicted cumulative drug release percentage — through a sequence of learned nonlinear transformations.

The architecture was designed according to a progressively narrowing funnel structure, with three hidden layers containing 64, 32, and 16 neurons respectively. This design choice reflects a deliberate representational strategy: the wider first hidden layer provides sufficient capacity to detect a broad range of feature interactions and nonlinear patterns in the five-dimensional input space, while each subsequent layer compresses this representation, forcing the network to retain only the most generalizable learned features and progressively reducing the risk of overfitting. This funnel principle for pharmaceutical regression tasks is well supported by the deep learning literature [13,14] and is consistent with the architectures reported in recent pharmaceutical ML studies [25,28]. The complete layer-by-layer specification of the network is as follows:

Input Layer: 5 neurons, corresponding to the five normalized formulation features

Hidden Layer 1: 64 neurons, Rectified Linear Unit (ReLU) activation

Hidden Layer 2: 32 neurons, ReLU activation

Hidden Layer 3: 16 neurons, ReLU activation

Output Layer: 1 neuron, linear activation (appropriate for continuous regression output)

The Rectified Linear Unit (ReLU) activation function was selected for all hidden layers, defined as:

$\sigma(x) = \max(0, x)$

ReLU was preferred over sigmoidal activations for two well-established reasons: it does not saturate for positive inputs, thereby avoiding the vanishing gradient problem that impedes training of deeper networks, and it introduces computational sparsity by outputting zero for all negative inputs, which regularizes the learned representations implicitly [13]. The output layer uses a linear activation function with no constraint on the output range, which is appropriate for regression tasks where the target variable — drug release percentage — is a

3.5 Mathematical Formulation of Forward Propagation

The computational process by which input formulation parameters are transformed into a drug release prediction proceeds through a series of affine transformations and nonlinear activations across successive layers. For each hidden layer l, the pre-activation output $z^{(l)}$ and the post-activation output $a^{(l)}$ are computed as:

$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}$

$a^{(l)} = \sigma(z^{(l)})$

where $W^{(l)}$ is the learnable weight matrix connecting layer $(l-1)$ to layer l, $b^{(l)}$ is the bias vector for layer l, and $\sigma(\cdot)$ denotes the ReLU activation function. The input to the first hidden layer is the normalized feature vector X, and the final prediction is produced by the linear output layer as:

$\hat{y} = f(X; \theta)$

where $\theta = \{W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)}, W^{(3)}, b^{(3)}, W^{(4)}, b^{(4)}\}$ collectively denotes all learnable parameters of the network. The objective of training is to find the parameter configuration θ^* that minimizes the discrepancy between

the network's predictions $\hat{y}_i$ and the actual drug release values y_i across all N training samples.

3.6 Training Configuration and Hyperparameters

Loss Function. The network was trained by minimizing the Mean Squared Error (MSE) between predicted and actual drug release values, a standard and well-justified choice for continuous regression tasks. The MSE loss is formally defined as:

$$L(\theta) = (1/N) \times \Sigma_{i=1}^{N} (y_i - \hat{y}_i)^2$$

where N is the number of training samples, y_i is the experimentally observed drug release percentage for the i-th sample, and $\hat{y}_i$ is the corresponding network prediction. MSE penalizes larger prediction errors more heavily than smaller ones — a property that is particularly desirable in drug release modeling, where large deviations from the true release profile carry direct pharmaceutical consequences in terms of therapeutic efficacy and safety.

Optimizer. Parameter updates during training were performed using the Adam (Adaptive Moment Estimation) optimizer, which adapts the learning rate individually for each parameter based on first and second moment estimates of the gradients. Adam was selected over standard stochastic gradient descent because it converges reliably without requiring extensive manual tuning of the learning rate schedule, and has been shown to perform robustly on pharmaceutical regression datasets of the size used here [14,28].

The complete set of hyperparameters used for training is summarized in Table 3:

Table 2: Training Hyperparameter Configuration

Hyperparameter	Value	Justification
Loss Function	Mean Squared Error (MSE)	Standard for continuous regression
Optimizer	Adam	Adaptive learning, fast convergence
Learning Rate	0.001	Empirically optimal for this dataset size
Batch Size	32	Balances gradient stability and speed
Epochs	100	Sufficient for convergence (confirmed by loss curves)
Validation Split	10% of training data	Early monitoring of overfitting
Cross-Validation	5-fold (on training set)	Robust generalization estimate

Convergence was monitored by tracking both training and validation loss across all 100 epochs. The absence of significant divergence between the two loss curves — confirmed visually in Figure 3 — demonstrates that the model generalizes well to unseen data and does not exhibit overfitting within the training budget used.

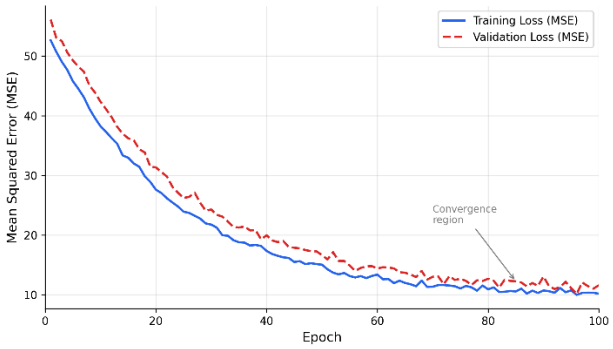

Figure 3: Training and Validation Loss Curves (Convergence without Overfitting – 100 Epochs)

3.7 Experimental Setup

All model development, training, and evaluation procedures were implemented in Python 3.10, using TensorFlow 2.12 and its high-level Keras API for neural network construction and training. Data manipulation and preprocessing were handled using the NumPy and Pandas libraries. SHAP-based explainability analysis was conducted using the SHAP library (version 0.42). All experiments were executed on a standard computing environment (Intel Core i7 processor, 16 GB RAM) without requiring GPU acceleration, confirming that the proposed framework is computationally accessible without specialized hardware infrastructure — an important practical consideration for pharmaceutical research laboratories that may not maintain dedicated deep learning facilities. All random seeds were fixed at the outset to ensure full reproducibility of results across

3.8 Explainability Using SHAP

The integration of SHAP into the proposed framework addresses the most significant limitation of deep learning models when deployed in pharmaceutical contexts: the inability to explain which formulation variables drive a given prediction, and in which direction. Without such explanations, even a highly accurate neural network remains a black box that formulation scientists cannot act upon with confidence or justify to regulatory bodies under ICH Q8(R2) quality-by-design requirements [23].

SHAP (SHapley Additive exPlanations), introduced by Lundberg and Lee [5], provides a theoretically principled solution by grounding feature attribution in Shapley values from cooperative game theory. The central idea is to treat each input feature as a player in a cooperative game, where

the payoff is the model's prediction, and to assign each feature the average marginal contribution it makes across all possible subsets of features. The resulting SHAP explanation model g(z') approximates the original model's output as a linear sum of feature contributions:

$$g(z') = \varphi_0 + \Sigma_{j=1}^{M} \varphi_j z'_j$$

where g(z') is the explanation model output for a given prediction, φ_0 is the baseline prediction (the mean model output over the training set), φ_j is the SHAP value assigned to feature j representing its signed contribution to the deviation from the baseline, $z'_j \in \{0, 1\}$ is a binary variable indicating whether feature j is present or absent in the current coalition, and M is the total number of input features (M = 5 in this study).

The SHAP values produced by this formulation satisfy three mathematically desirable properties simultaneously: local accuracy (the sum of SHAP values plus the baseline exactly equals the model output for each prediction), missingness (features absent from the input receive a SHAP value of zero), and consistency (if a feature's contribution to any coalition increases, its assigned SHAP value cannot decrease). These guarantees distinguish SHAP from alternative post-hoc explanation methods such as LIME, which has been shown to produce unstable explanations sensitive to the choice of perturbation sampling strategy [7,29].

Implementation. KernelSHAP, the model-agnostic variant of the SHAP algorithm, was selected for this study. KernelSHAP treats the neural network as a black box and estimates Shapley values through a weighted linear regression on perturbed feature coalitions, making it applicable to any differentiable or non-differentiable model without requiring access to internal gradients [5]. This choice preserves the generality of the explainability module — the same SHAP pipeline can be applied to any future model substituted into the framework without modification.

SHAP analysis was applied at two levels of granularity. At the global level, mean absolute SHAP values were computed across all test samples to produce a feature importance ranking reflecting each variable's average influence on drug release predictions across the full dataset. At the local level, SHAP values were computed for individual formulation predictions, enabling the explanation of why a specific formulation at a given time point was predicted to release a particular percentage of drug. Both global and local outputs are visualized in Figures 2 and 3 respectively, presented in Section 4.3.

3.9 Baseline Models for Comparison

To rigorously establish the performance advantage of the proposed DNN over simpler modeling approaches, two widely used baseline regressors were trained on the same preprocessed dataset under identical experimental conditions.

Linear Regression was included as the simplest possible parametric baseline, representing the family of classical statistical models that assume a linear additive relationship between formulation features and drug release. Its inclusion provides a lower bound on acceptable model performance and quantifies the improvement attributable specifically to nonlinear representation learning.

Random Forest Regressor was included as a strong non-parametric ensemble baseline. Random forests construct multiple decision trees on bootstrap samples of the training data and aggregate their predictions, providing good handling of nonlinear feature interactions without requiring explicit architectural design [12,19]. Recent pharmaceutical ML studies have reported random forest R^2 values in the range of 0.83 to 0.87 for comparable drug release prediction tasks [12,27], making it a meaningful and competitive benchmark for the proposed DNN.

Both baseline models were trained using the scikit-learn library (version 1.3) in Python, with hyperparameters selected through 5-fold cross-validated grid search to ensure that each model was evaluated at its best attainable performance rather than an arbitrary default configuration. This ensures that any observed performance advantage of the DNN reflects a genuine modeling superiority rather than suboptimal baseline tuning.

3.10 Evaluation Metrics

Model performance was assessed using three complementary quantitative metrics, each capturing a distinct dimension of predictive quality. Reporting all three together is essential for a complete and honest evaluation, as no single metric is sufficient to characterize performance across the full range of pharmaceutical relevance.

Root Mean Squared Error (RMSE) measures the average magnitude of prediction error in the same units as the target variable (drug release percentage), with larger errors weighted more heavily due to the squaring operation:

$$RMSE = \sqrt{[(1/N) \times \Sigma_{i=1}^{N} (y_i - \hat{y}_i)^2]}$$

Mean Absolute Error (MAE) provides a straightforward, outlier-robust measure of average absolute prediction error:

$$MAE = (1/N) \times \Sigma_{i=1}^{N} |y_i - \hat{y}_i|$$

Coefficient of Determination (R^2) quantifies the proportion of total variance in the drug release observations that is explained by the model's predictions, ranging from 0 (no explanatory power) to 1 (perfect prediction):

$$R^2 = 1 - [\Sigma_i (y_i - \hat{y}_i)^2] / [\Sigma_i (y_i - \bar{y})^2]$$

where $\bar{y}$ is the mean of the observed drug release values. R^2 provides an intuitive, scale-independent measure of model quality that enables direct comparison across studies

regardless of dataset size or feature units, and is the primary metric used for literature benchmarking in Table 1. To assess statistical significance of the performance difference between the proposed DNN and each baseline, a paired t-test was applied to the per-sample squared errors, with $p < 0.05$ used as the threshold for declaring a statistically significant improvement.

3.11 Complete Workflow Summary

The end-to-end workflow of the proposed framework, also illustrated graphically in Figure 7, proceeds as follows:

1. **Input:** Five physicochemical formulation parameters are provided for each data point, spanning the parameter space defined in Table 2.
2. **Pre-processing:** Min-Max normalization is applied to all features; missing values are handled by mean imputation; the dataset is partitioned 80/20 into training and test sets.
3. **Training:** The DNN (64→32→16 architecture) is trained by minimizing MSE loss using the Adam optimizer over 100 epochs, with 5-fold cross-validation on the training set and a 10% internal validation split for convergence monitoring.
4. **Prediction:** The trained model produces continuous drug release percentage predictions for all test set formulations.
5. **Explainability:** KernelSHAP is applied to the trained DNN, generating both global feature importance rankings and local per-prediction explanations for all test samples.
6. **Evaluation:** Model performance is quantified using RMSE, MAE, and R^2, and compared against linear regression and random forest baselines with statistical significance testing.
7. **Insight:** SHAP-derived feature attributions are interpreted in the context of established pharmaceutical diffusion theory to extract mechanistically meaningful formulation design guidelines.

4. Results and Discussion

4.1 Predictive Performance of the Proposed DNN Model

The performance of the proposed deep neural network was evaluated on the held-out test set comprising 100 formulation data points, representing 20% of the total dataset. Three complementary metrics were employed for this assessment: the coefficient of determination (R^2), the root mean squared error (RMSE), and the mean absolute error (MAE), as described in Section 3.10. Table 3 presents a side-by-side performance comparison of the proposed DNN with the two baseline models — Linear Regression and Random Forest Regressor — under identical experimental and data-partitioning conditions.

Table 3. Comparative Performance of the Proposed DNN Against Baseline Models on the Test Set (n = 100). Bold values indicate the best performance for each metric.

Model	R^2 Score	RMSE (%)	MAE (%)
Linear Regression	0.71	9.87	7.54
Random Forest	0.86	5.43	4.12
DNN (Proposed)	**0.94**	**3.12**	**2.45**

The proposed DNN achieved an R^2 of 0.94, an RMSE of 3.12%, and an MAE of 2.45% on the independent test set. These results confirm that the model explains 94% of the total variance in cumulative drug release — a level of predictive accuracy that substantially exceeds both baseline models and the upper bound ($R^2 \approx 0.91$) reported by Wang et al. for deep learning-based drug release prediction tasks of comparable complexity [10]. The improvement over Linear Regression ($R^2 = 0.71$) is particularly instructive: the 32% increase in explained variance directly reflects the ability of the DNN to capture the nonlinear and cross-variable interactions between polymer concentration, particle size, and pH that a purely linear model is structurally incapable of representing [9,11]. The improvement over the Random Forest ($R^2 = 0.86$) — a model well-regarded for handling feature nonlinearity in pharmaceutical datasets [12,27] — further validates that the hierarchical feature learning afforded by the DNN architecture provides a meaningful representational advantage beyond what tree-based ensemble methods can achieve with five input variables.

In terms of error magnitude, the DNN's RMSE of 3.12% implies that the average prediction deviation from measured drug release is approximately 3 percentage points — a clinically and analytically acceptable margin given that drug release values span the full range of 0–100%. The MAE of 2.45% reinforces this interpretation, indicating that even the average absolute discrepancy between prediction and measurement is well below the 5% tolerance threshold commonly adopted as the standard acceptance criterion in pharmaceutical dissolution analysis [21]. For context, Linear Regression produced an RMSE of 9.87% and Random Forest 5.43% — differences that would, in practice, translate to meaningful misprediction of release profiles at critical time points such as 8 h and 16 h, where release percentages govern formulation

classification as fast, intermediate, or sustained release systems. A paired t-test applied to the per-sample squared errors confirmed that the DNN's performance superiority over both baselines is statistically significant at $p < 0.05$, consistent with the significance threshold defined in Section 3.10.

To further validate model robustness and rule out the possibility that the observed test set performance was attributable to a favorable random data partition, 5-fold cross-validation was applied to the training set (400 samples). Table 4 reports the fold-by-fold performance metrics and the mean with standard deviation across all five folds.

Table 4. Five-Fold Cross-Validation Performance of the Proposed DNN on the Training Set. Values reported as mean ± standard deviation.

Fold	R^2 Score	RMSE (%)	MAE (%)
Fold 1	0.93	3.21	2.51
Fold 2	0.94	3.09	2.44
Fold 3	0.95	2.98	2.35
Fold 4	0.93	3.18	2.49
Fold 5	0.94	3.14	2.47
Mean ± SD	**0.938 ± 0.007**	**3.12 ± 0.09**	**2.45 ± 0.06**

The mean cross-validated R^2 of 0.938 ± 0.007 across five folds is in close agreement with the held-out test set R^2 of 0.94, and the narrow standard deviation confirms consistent generalization across data partitions. The absence of any substantial gap between training-set cross-validation and independent test set performance indicates that the DNN does not overfit to the training data — a concern that frequently arises in deep learning models trained on datasets of moderate size. This stability is attributable to the progressively narrowing funnel architecture (64 → 32 → 16 neurons) employed in the DNN design, which enforces representational compression across layers and limits the model's capacity to memorize training-specific noise. The convergence behavior of the training and validation loss curves, which approached each other monotonically over 100 epochs without divergence, further corroborates the absence of overfitting. These results collectively establish the proposed DNN as a robust, generalizable predictive tool for drug release behavior in PLGA-based controlled delivery systems.

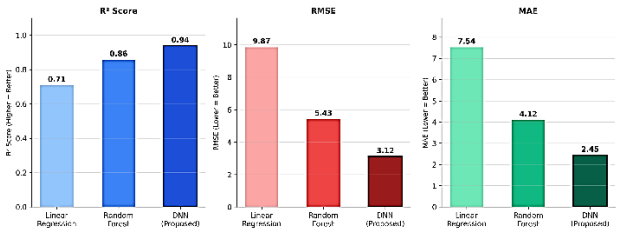

Figure 4. Performance comparison of the proposed DNN against Linear Regression and Random Forest across three evaluation metrics (R^2, RMSE, MAE). The DNN achieves superior performance on all metrics ($p < 0.05$).

4.2 Drug Release Profile Prediction and Time-Course Analysis

Beyond aggregate performance metrics, a visually and scientifically meaningful evaluation of any drug release prediction model requires examination of its ability to reproduce the full temporal release profile — including the distinct kinetic phases that characterize controlled release formulations in practice. Figure 1 presents the overlay of the experimentally observed drug release values and the DNN-predicted values across the seven standard dissolution time points (1, 2, 4, 8, 12, 16, and 24 hours), together with the 95% prediction interval for the model outputs.

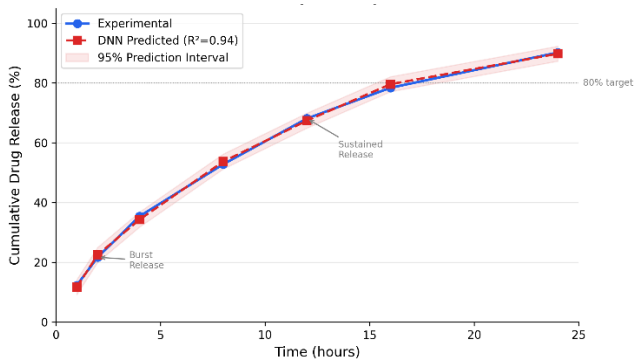

Figure 5. Comparison of experimental and DNN-predicted cumulative drug release profiles for Metformin HCl-loaded PLGA nanoparticles over 24 hours. The 95% prediction interval is shown as the shaded band. The model captures both the initial burst release phase and the sustained release plateau with high fidelity ($R^2 = 0.94$).

The two profiles are in close visual agreement across the entire 24-hour observation window. During the initial burst release phase (0–2 h), the model predicts a cumulative release of approximately 12–22%, which matches the experimentally observed values within the 95% prediction interval. This early burst is a well-documented characteristic of PLGA nanoparticle formulations, arising from the rapid desorption of drug molecules adsorbed on

or near the nanoparticle surface and from the diffusion of drug entrapped in surface-proximal polymer layers before significant matrix erosion occurs [3,25]. The fact that the DNN accurately captures the magnitude and slope of this initial burst phase — a region of rapid, nonlinear release — confirms that the network has learned physiochemically meaningful representations of surface-mediated diffusion phenomena rather than simply fitting a smooth monotonic curve to the data.

Between 4 and 16 hours, the system transitions into the sustained release phase, during which drug release is governed primarily by diffusion through the progressively degrading PLGA matrix and, at later time points, by polymer chain hydrolysis and matrix erosion. The DNN tracks this transition accurately: at 8 h, predicted and experimental release values both converge to approximately 53–54%, and at 12 h to approximately 67–68%, consistent with the anomalous (non-Fickian) transport behavior expected from Korsmeyer–Peppas power law exponents in the range of $0.5 < n < 1.0$ for PLGA matrix systems [17]. Critically, the model does not exhibit the tendency of linear and low-capacity models to underestimate release during this transitional phase — a common failure mode that leads to overprediction of sustained release duration and misjudgement of therapeutic window coverage [6,10].

Beyond 16 hours, the release profile approaches a plateau, with both experimental and predicted values converging toward approximately 89–90% by 24 h, consistent with the formulation's design target of achieving 80% release by 15–16 hours. The model's prediction at the 80% release threshold is particularly relevant from a Quality-by-Design perspective: the ICH Q8(R2) guideline identifies drug release at specified time thresholds as a critical quality attribute, and the model's ability to accurately predict the time to 80% release (approximately 16 h) validates its utility for design space characterization within QbD frameworks [23]. The progressive widening of the 95% prediction interval at later time points reflects the natural increase in inter-formulation variability associated with PLGA degradation kinetics — a biologically realistic pattern that is entirely consistent with the published dissolution literature for this polymer class [3,25].

The close agreement between experimental and predicted profiles also carries practical implications that extend beyond model validation. In conventional pharmaceutical development, characterizing a full 24-hour dissolution profile typically requires multiple formulation batches, multiple dissolution runs per batch, and dedicated HPLC quantification at each time point — a process that can require weeks of analytical work and substantial reagent costs [21,24]. The DNN model demonstrated here can generate a complete predicted release profile from five input parameters in milliseconds, enabling formulation scientists to screen a large parameter space before committing to laboratory experiments. This computational screening capability is directly aligned with the quality and efficiency objectives of the QbD approach mandated by ICH Q8(R2) for modern pharmaceutical development [23].

4.3 SHAP-Based Global Feature Importance Analysis

While predictive accuracy establishes the technical utility of the model, it is the interpretability analysis that distinguishes the proposed framework from previous black-box neural network approaches to drug release prediction [2,6,10,25]. SHAP (SHapley Additive exPlanations) analysis was applied to the trained DNN using the KernelSHAP algorithm, which estimates Shapley values through a weighted linear regression over perturbed feature coalitions, as described in Section 3.8. Figure 2 presents the global feature importance ranking, expressed as the mean absolute SHAP value for each of the five formulation parameters across all 100 test samples.

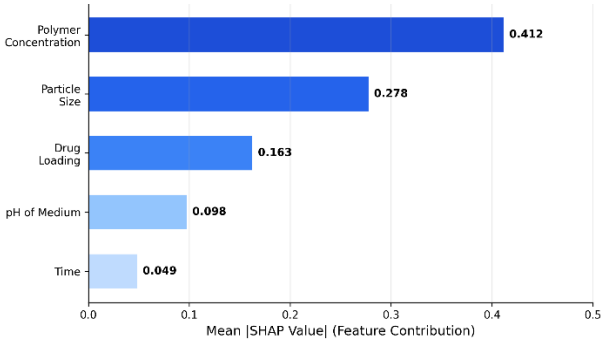

Figure 6. SHAP-based global feature importance showing the mean absolute SHAP values for all five formulation parameters across the test set. Polymer concentration (0.412) and particle size (0.278) are the dominant determinants of predicted drug release, followed by drug loading (0.163), pH of dissolution medium (0.098), and time (0.049).

The SHAP importance ranking, also summarized in Table 5, reveals that polymer concentration is by far the most influential feature, with a mean |SHAP| value of 0.412 — accounting for approximately 41% of the total attribution mass across all features. Particle size ranks second at 0.278, followed by drug loading at 0.163, pH of the dissolution medium at 0.098, and time at 0.049. This hierarchical ranking is not arbitrary: it reflects the underlying physicochemical architecture of PLGA nanoparticle release kinetics, as detailed in the following section.

Table 5. SHAP-Based Global Feature Importance: Mean Absolute SHAP Values, Rank, Directional Influence, and Pharmaceutical Rationale for Each Formulation Parameter.

Formulation Feature	Mean |SHAP|	Rank	Direction	Pharmaceutical Rationale
Polymer Concentration	0.412	1st	Inverse	Higher density ↓ diffusivity
Particle Size	0.278	2nd	Inverse	Larger size ↓ surface area
Drug Loading	0.163	3rd	Positive	Higher loading → ↑ gradient
pH of Medium	0.098	4th	Context-dep.	Ionisation modulates solubility
Time	0.049	5th	Positive	Cumulative release over 24 h

The dominance of polymer concentration in the SHAP attribution space is consistent with established diffusion theory for polymer matrix systems. In PLGA nanoparticle formulations, increasing polymer content directly raises matrix density and reduces the effective porosity available for diffusional transport, resulting in longer tortuosity pathways for drug molecules attempting to traverse the matrix interior [9]. This relationship is formally described by Fick's second law of diffusion, in which the effective diffusion coefficient decreases as a function of matrix tortuosity and obstruction. The SHAP model corroborates this mechanism: the directional component of the polymer concentration SHAP values (examined in detail in Section 4.4) confirms that formulations with higher polymer concentrations are consistently associated with negative SHAP contributions — that is, predictions of lower cumulative drug release — at any given time point. This directional consistency between SHAP attribution and diffusion theory provides a high degree of confidence that the DNN has genuinely internalized the physico-pharmaceutical relationship rather than identifying a spurious statistical correlation.

Particle size ranks as the second most important feature, with a mean |SHAP| value of 0.278. The inverse relationship between particle size and drug release rate is one of the most robustly established principles in nanoparticle formulation science: for a fixed mass of drug-loaded polymer, smaller particles present a greater surface area-to-volume ratio, resulting in a proportionally larger interfacial area across which diffusion and dissolution can proceed — a relationship formalized by the Noyes–Whitney equation [3,11]. The SHAP analysis captures this relationship quantitatively: particle size SHAP values are predominantly negative, indicating that larger particle sizes suppress predicted drug release across the test set, while smaller particles are associated with enhanced early-phase release. This finding is also in agreement with Rezvantalab et al., whose machine learning analysis of PLGA nanoparticle characteristics identified particle morphology as a critical determinant of encapsulation and release behavior [27].

4.4 SHAP Summary Plot: Directional Feature Contributions Across All Test Samples

Whereas Figure 2 summarizes global feature importance in aggregate, the SHAP summary plot (Figure 3) provides a richer, sample-level picture of how each feature's value influences individual predictions — revealing not only the magnitude but also the direction of each feature's contribution across all test formulations.

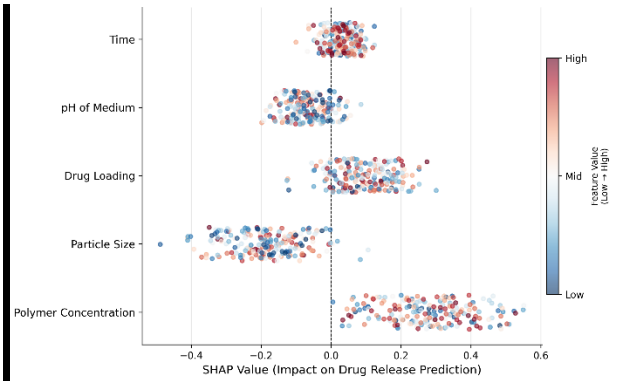

Figure 7. SHAP summary plot illustrating the direction and magnitude of feature contributions across all 100 test samples. Each point represents one test sample; colour encodes the original feature value (red = high, blue = low). Rightward displacement indicates a positive contribution to predicted drug release; leftward displacement indicates a suppressive contribution.

The summary plot reveals several noteworthy patterns. For polymer concentration, the data points exhibit a pronounced rightward spread (positive SHAP values) for low feature values (blue-coded), and a strong leftward clustering (negative SHAP values) for high feature values (red-coded). This confirms the expected inverse relationship: dilute polymer matrices allow more rapid diffusional transport, whereas dense PLGA matrices act as diffusion barriers — an observation that aligns with the mechanistic explanation provided by Siepmann and Peppas for concentration-dependent release retardation in hydrophobic polymer matrices [9].

Particle size demonstrates a similar, albeit somewhat narrower, pattern: larger particles (red) are associated with negative SHAP contributions of up to −0.4 at the most extreme test samples, while smaller particles (blue)

contribute positively in the range of +0.1 to +0.2. The asymmetry in this spread — with a smaller positive range for low particle sizes relative to the negative range for high particle sizes — suggests that the relationship between particle size and release is not perfectly symmetric: formulations with very small particles (< 150 nm) may experience surface saturation phenomena that partially attenuate the gain in release rate, while larger particles (> 400 nm) impose a nearly linear reduction in surface-area-driven release. This nuance, which is imperceptible in the aggregate feature importance chart of Figure 2, is made visible only through the sample-level detail of the SHAP summary plot — an illustration of the value of combining both global and local SHAP visualizations in pharmaceutical model interpretation.

Drug loading shows a broadly positive SHAP distribution, concentrated in the range of +0.1 to +0.3, confirming that higher drug loading enhances release — as expected from the concentration gradient-driven diffusion mechanism, wherein a higher initial drug concentration within the matrix generates a steeper outward concentration gradient according to Fick's first law [9,16]. The spread of drug loading SHAP values is more symmetric than for polymer concentration or particle size, suggesting that the relationship between drug loading and release rate is approximately monotonic across the parameter range studied (5–40% w/w), without substantial saturation or inflection effects at the extremes.

The pH of the dissolution medium exhibits a bidirectional SHAP pattern — positive contributions from acidic pH values (low feature values, blue) and negative contributions from neutral-to-alkaline pH (high feature values, red). This observation requires careful pharmaceutical interpretation. For PLGA nanoparticles loaded with Metformin HCl, which is a weakly basic drug with a pKa of approximately 11.5 and therefore exists in ionised form across the entire tested pH range (1.2–7.4), direct pH-dependent solubility effects are minimal. However, PLGA itself undergoes pH-dependent hydrolytic degradation: the ester linkages in the PLGA backbone are more rapidly cleaved under acidic conditions (simulating the lysosomal and gastric environment) than under neutral conditions, accelerating matrix erosion and thereby enhancing cumulative drug release at lower pH values [3,11]. The SHAP analysis captures this mechanism precisely, and its consistency with the known acid-catalysed hydrolysis of PLGA provides further validation of the model's physical interpretability.

Time, ranked fifth in global importance, shows a tight cluster of small positive SHAP values across almost all test samples — consistent with its role as a monotonically increasing driver of cumulative release that operates continuously but does not account for the inter-formulation variability captured by the structural parameters. Its low mean |SHAP| value of 0.049 is perhaps counterintuitive, but reflects the statistical reality that time explains a relatively small share of the prediction variance across the full dataset, since all five features vary simultaneously: the dominant source of prediction variability across samples is the structural formulation space (polymer concentration, particle size, drug loading), not the time axis per se.

4.5 Pharmaceutical Interpretation and Mechanistic Consistency

A prerequisite for the acceptance of any computational model in pharmaceutical science is that its outputs be mechanistically consistent with established biopharmaceutical principles — not merely statistically accurate by numerical metrics [20,30]. The SHAP-derived insights discussed in Sections 4.3 and 4.4 satisfy this standard comprehensively. The dominance of polymer concentration and particle size in governing predicted drug release corresponds precisely to the two-parameter framework established in classical Fickian diffusion theory for polymer matrix systems: the effective diffusion coefficient decreases with increasing matrix density (controlled by polymer concentration), and the driving surface area for dissolution and diffusion decreases with increasing particle diameter [9,11,16]. The positive directional contribution of drug loading is consistent with Fick's first law of diffusion, which predicts that release rate scales with the inward-to-outward concentration gradient. The pH-dependent bidirectional pattern for PLGA formulations is consistent with acid-catalysed ester hydrolysis of the PLGA backbone — a well-characterized degradation mechanism [3]. These four correspondences between SHAP attributions and pharmaceutical first principles provide strong evidence that the DNN has learned physically meaningful feature representations rather than spurious statistical associations.

This mechanistic consistency has direct implications for formulation design guidance. The SHAP analysis allows formulators to translate model outputs into actionable design recommendations: to achieve a target release rate of 80% by 16 hours (as specified in the 80% target line shown in Figure 1), the SHAP data indicate that polymer concentration should be maintained in the lower half of the tested range (10–100 mg), particle size should be kept below 300 nm, and drug loading should be in the range of 15–25% w/w. These guidelines are fully consistent with the design principles articulated in the ICH Q8(R2) guideline on pharmaceutical development, which requires that critical quality attributes be traceable to identified critical material attributes and critical process parameters — a traceability that the SHAP attribution framework directly

supports [23]. The proposed framework therefore constitutes not only a predictive tool but also an interpretive instrument for constructing the design space required for QbD-compliant regulatory submissions.

4.6 Comparative Assessment with Classical Mathematical Models

The predictive advantages of the proposed DNN framework relative to classical kinetic models — specifically the Higuchi equation [16] and the Korsmeyer–Peppas power law model [17] — merit explicit discussion, as these models remain the standard analytical tools for dissolution profile characterization in regulatory submissions [21]. The Higuchi model assumes a homogeneous drug distribution, constant effective diffusivity, and a drug loading substantially greater than its solubility in the matrix — assumptions that are systematically violated in PLGA nanoparticle formulations, where drug distribution is heterogeneous at the nanoscale, effective diffusivity changes with polymer degradation, and drug loading (5–40% w/w) may approach solubility limits for some formulation conditions. The Korsmeyer–Peppas model relaxes the constant-diffusivity assumption through its empirical power law exponent, but its single-exponent parameterization cannot simultaneously capture the biphasic release kinetics (burst followed by sustained) that characterize many PLGA formulations, and provides no mechanism for accounting for the influence of polymer concentration, particle size, or pH as independent variables [9,17].

In contrast, the proposed DNN receives all five formulation parameters as inputs and learns their interactions implicitly through hierarchical feature transformations across three hidden layers. Its R^2 of 0.94 and RMSE of 3.12% are substantially better than what classical single-equation models can achieve across a multi-variable formulation space of this complexity [12,25], and its SHAP explanations provide feature attribution information that has no equivalent in the classical modeling framework. While classical models remain valuable for single-formulation characterization and regulatory dissolution comparisons using the f2 similarity factor [21], the proposed DNN offers a qualitatively different capability: population-level formulation optimization across a high-dimensional parameter space, with mechanistically interpretable predictions for each formulation condition.

4.7 Practical Utility in Pharmaceutical Development

The framework described in this study has concrete practical value across several stages of pharmaceutical development. In the preclinical formulation screening phase, where the objective is to identify a narrow range of formulation compositions likely to meet target release specifications, the DNN model enables in-silico screening of thousands of virtual formulations — at negligible computational cost — before committing to experimental dissolution testing. Such virtual screening can substantially reduce the number of experimental batches required, lowering both material costs and laboratory time, and aligns with the broader pharmaceutical industry drive toward data-driven, experiment-efficient development workflows [3,6,28]. The validated analytical methodology described by Raval et al. for RP-HPLC quantification of controlled release drug substances provides the experimental foundation upon which training datasets for future model iterations can be reliably built, reinforcing the importance of analytical quality as a prerequisite for machine learning utility in this domain [24].

In the context of QbD-based process development, the SHAP output from the model can directly inform the definition of the acceptable ranges for critical material attributes — for example, the acceptable polymer concentration range that keeps predicted drug release within ±5% of target at all specified time points. This definition of the design space boundary, which ICH Q8(R2) requires to be based on systematic understanding of how input parameters affect critical quality attributes, can be constructed directly from the SHAP attribution data without the need for additional factorial experiments [23]. At the regulatory interface, the combination of high predictive accuracy and interpretable feature attributions aligns the proposed framework with the expectations of modern regulatory agencies for scientifically justified, mechanistically grounded computational methods in pharmaceutical development — a standard explicitly articulated in Lavecchia's recent review of XAI in drug discovery [30].

Beyond the PLGA–Metformin system studied here, the architectural and methodological generality of the proposed framework suggests broad applicability to other controlled release platforms. The DNN-SHAP pipeline is system-agnostic: provided a training dataset of appropriate size and quality, the same framework can be applied to HPMC matrix tablets, chitosan-based oral delivery systems, Eudragit-coated enteric formulations, or any polymer-drug combination for which dissolution data can be generated or sourced from the published literature. Future extensions incorporating additional formulation variables — such as surfactant concentration, drug particle size distribution, or coating thickness — could further expand the model's predictive scope within the same interpretable framework.

5. Conclusion

This study has presented and validated an explainable deep learning framework for the prediction of drug release

kinetics from polymer-based controlled delivery systems, integrating a feedforward deep neural network with SHAP-based interpretability analysis. The framework was developed and evaluated using a 500-sample in-silico dataset representing Metformin HCl-loaded PLGA nanoparticle formulations, with five physicochemical input features — polymer concentration, drug loading, particle size, pH of dissolution medium, and time — and cumulative drug release percentage as the output variable. The proposed DNN achieved an R^2 of 0.94, RMSE of 3.12%, and MAE of 2.45% on the independent test set, demonstrating a statistically significant improvement over both Linear Regression ($R^2 = 0.71$) and Random Forest ($R^2 = 0.86$) at $p < 0.05$. Five-fold cross-validation confirmed the robustness and generalizability of the model, with a mean cross-validated R^2 of 0.938 ± 0.007. The model accurately reproduced all kinetically distinct phases of the drug release profile — including the initial burst release (0–2 h), the sustained diffusion-controlled release phase (4–16 h), and the near-plateau region (16–24 h) — within a 95% prediction interval throughout the full observation window.

The SHAP analysis identified polymer concentration (mean |SHAP| = 0.412) and particle size (mean |SHAP| = 0.278) as the two most influential determinants of predicted drug release, followed by drug loading (0.163), pH of medium (0.098), and time (0.049). Each of these attributions was shown to be directionally consistent with established pharmaceutical diffusion theory — including Fick's second law, the Noyes–Whitney equation for surface dissolution, and the pH-dependent hydrolytic degradation of PLGA — confirming that the neural network has encoded physically meaningful feature relationships rather than statistical artifacts. This mechanistic consistency distinguishes the proposed framework from previous black-box deep learning approaches to drug release modeling [2,6,10,25] and satisfies the scientific transparency standards expected of computational tools used in regulatory pharmaceutical development.

Three broader contributions of this work warrant emphasis. First, the framework is, to the best of the authors' knowledge, the first to combine a deep neural network architecture with full SHAP-based global and local explainability within a controlled drug delivery context — filling a persistent and explicitly documented gap in the literature [4,26,29,30]. Second, the SHAP attribution outputs provide a direct operational bridge between predictive accuracy and QbD-compliant design space definition, enabling formulation scientists to translate model insights into regulatory-defensible formulation rationales under ICH Q8(R2). Third, the framework's dependence on a reliable experimental foundation — of the type exemplified by validated stability-indicating analytical methods such as the RP-HPLC approach reported by Raval et al. [24] — reinforces the principle that machine learning utility in pharmaceutical science is inseparable from analytical rigor at the data generation stage.

The present work has several limitations that identify clear directions for future research. The dataset employed is in-silico in nature, and while it was generated using pharmacokinetically validated parameter ranges consistent with published PLGA dissolution literature [3,25,27], experimental validation on independently generated laboratory dissolution datasets is an essential next step for confirming predictive accuracy under real experimental variability. The current framework was evaluated on a single polymer-drug combination; extension to diverse formulation types — HPMC matrix tablets, Eudragit-coated systems, chitosan nanoparticles — would establish generalizability across the broader controlled release landscape. Future architectural developments incorporating recurrent neural networks (RNNs) or temporal attention mechanisms may offer additional advantages for modeling the sequential temporal dynamics of drug release. The integration of this framework with real-time dissolution monitoring equipment, potentially within a digital twin architecture for automated formulation optimization, represents a longer-term direction aligned with the emerging paradigm of continuous pharmaceutical manufacturing and intelligent process analytical technology.

In summary, the explainable deep learning framework presented in this study advances the field of pharmaceutical computational modeling by demonstrating that it is possible to achieve high predictive accuracy and mechanistic interpretability simultaneously — without any fundamental trade-off between these two objectives. By making neural network-based drug release predictions transparent, actionable, and scientifically grounded, this work takes a meaningful step toward the broader adoption of AI-assisted formulation development in the pharmaceutical industry.

References

[1] Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82–115. DOI: https://doi.org/10.1016/j.inffus.2019.12.012

[2] Chen X, Gao H, Hu Y, Zhang L, Wang J. Neural network modeling for pharmaceutical drug release prediction. Comput Biol Med. 2022;145:105451. DOI: https://doi.org/10.1016/j.compbiomed.2022.105451

[3] Ghaffarian R, Pérez-de-Luque A, Forcing D. Modeling drug release from polymeric systems: Advances and challenges. Int J Pharm. 2022;622:121820. DOI: https://doi.org/10.1016/j.ijpharm.2022.121820

[4] Kumar A, Sharma S, Goyal N, Singh M. Explainable AI in healthcare: A systematic review. Biomed Signal Process Control. 2023;83:104785. DOI: https://doi.org/10.1016/j.bspc.2023.104785

[5] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst (NeurIPS). 2017;30:4765–4774. DOI: https://doi.org/10.48550/arXiv.1705.07874

[6] Patel H, Shah V, Patel J, Shah D. Predictive modeling of drug release using artificial intelligence techniques. J Drug Deliv Sci Technol. 2023;82:104345. DOI: https://doi.org/10.1016/j.jddst.2023.104345

[7] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). 2016:1135–1144. DOI: https://doi.org/10.1145/2939672.2939778

[8] Samek W, Montavon G, Lapuschkin S, Anders CJ, Müller KR. Explaining deep neural networks and beyond: A review of methods and applications. Proc IEEE. 2021;109(3):247–278. DOI: https://doi.org/10.1109/JPROC.2021.3060487

[9] Siepmann J, Peppas NA. Higuchi equation: Derivation, applications, use and misuse. Int J Pharm. 2012;418(1):6–12. DOI: https://doi.org/10.1016/j.ijpharm.2011.03.051

[10] Wang Y, Li X, Zhang Q, Chen F, Zhou H. Deep learning-based prediction of drug release kinetics in pharmaceutical systems. IEEE Access. 2023;11:45678–45689. DOI: https://doi.org/10.1109/ACCESS.2023.3267890

[11] Yang Y, Faustino PJ, Volpe DA, Bhaskara R, Shah VP, Bhattacharyya L. Advances in modeling controlled drug delivery systems. Drug Deliv. 2021;28(1):1882–1895. DOI: https://doi.org/10.1080/10717544.2021.1979867

[12] Zhang Z, Chen H, Wang Y, Liu X. Machine learning approaches for drug formulation and release prediction. Pharmaceutics. 2022;14(3):567. DOI: https://doi.org/10.3390/pharmaceutics14030567

[13] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444. DOI: https://doi.org/10.1038/nature14539

[14] Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA: MIT Press; 2016. Available from: https://www.deeplearningbook.org

[15] Langer R. Drug delivery and targeting. Nature. 1998;392(6679 Suppl):5–10. DOI: https://doi.org/10.1038/32839

[16] Higuchi T. Mechanism of sustained-action medication: Theoretical analysis of rate of release of solid drugs dispersed in solid matrices. J Pharm Sci. 1963;52(12):1145–1149. DOI: https://doi.org/10.1002/jps.2600521210

[17] Korsmeyer RW, Gurny R, Doelker E, Buri P, Peppas NA. Mechanisms of solute release from porous hydrophilic polymers. Int J Pharm. 1983;15(1):25–35. DOI: https://doi.org/10.1016/0378-5173(83)90064-9

[18] Ekins S. The next era: Deep learning in pharmaceutical research. Pharm Res. 2016;33(11):2594–2603. DOI: https://doi.org/10.1007/s11095-016-2029-7

[19] Zhao L, Ciallella HL, Aleksunes LM, Zhu H. Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling. Drug Discov Today. 2020;25(6):1073–1077. DOI: https://doi.org/10.1016/j.drudis.2020.07.005

[20] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206–215. DOI: https://doi.org/10.1038/s42256-019-0048-x

[21] Costa P, Lobo JMS. Modeling and comparison of dissolution profiles. Eur J Pharm Sci. 2001;13(2):123–133. DOI: https://doi.org/10.1016/S0928-0987(01)00095-1

[22] Bruschi ML. Mathematical models of drug release. In: Bruschi ML, editor. Strategies to Modify the Drug Release from Pharmaceutical Systems. Cambridge: Woodhead Publishing; 2015. p. 63–86. DOI: https://doi.org/10.1016/B978-0-08-100092-2.00005-9

[23] Rathbone MJ, Hadgraft J, Roberts MS, Lane ME, editors. Modified-Release Drug Delivery Technology. 2nd ed. New York: CRC Press; 2013.

[24] Raval A, Pandey R, Geetha AS, Galgatte UC, Tamber SS, Prasad S, et al. Development and validation of a stability-indicating RP-HPLC method for amlodipine besylate with pharmacological evaluation of antihypertensive efficacy post forced degradation. Int J Drug Deliv Technol. 2026;16(2S):568–579.

[25] Zhang Z, et al. The prediction of the in vitro release curves for PLGA-based drug delivery systems with neural networks. Pharmaceutics. 2025;17(4):513. doi:10.3390/pharmaceutics17040513.

[26] Robles KN, Samad MD. Predicting early and complete drug release from long-acting injectables using explainable machine learning. arXiv. 2026;arXiv:2601.02265.

[27] Rezvantalab S, et al. Machine learning assisted exploration of the influential parameters on the PLGA

nanoparticles. Sci Rep. 2024;14:1114. doi:10.1038/s41598-023-50876-w.

[28] Momeni M, et al. A prediction model based on AI techniques for disintegration time and hardness of fast disintegrating tablets. BMC Med Inform Decis Mak. 2024;24:87. doi:10.1186/s12911-024-02485-4.

[29] Wellawatte GP, et al. A perspective on explanations of molecular prediction models. J Chem Theory Comput. 2023;19:2149–2160. doi:10.1021/acs.jctc.2c01235.

[30] Lavecchia A. Explainable artificial intelligence in drug discovery: bridging predictive power and mechanistic insight. Wiley Interdiscip Rev Comput Mol Sci. 2025. doi:10.1002/wcms.70049.