



Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



REVIEW

Machine learning empowered formulation design, optimization and characterization of nanoparticulate drug delivery systems: Current applications, challenges, and future perspectives



Chunyan Shen^{a,b,†}, Mengyan Zhang^{c,†}, Meiting Lu^{a,b},
Errong Chang^{a,b}, Ziting Gao^{a,b}, Weikang Ban^d, Qiang Liu^{a,b},
Zhong Zuo^{d,*}, Cuiping Jiang^{a,b,d,*}

^aGuangdong Provincial Key Laboratory of Chinese Medicine Pharmaceuticals, School of Traditional Chinese Medicine, Southern Medical University, Guangzhou 510515, China

^bGuangdong Basic Research Center of Excellence for Integrated Traditional and Western Medicine for Qingzhi Diseases, Guangzhou 510515, China

^cGuangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou 510515, China

^dSchool of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

Received 22 June 2025; received in revised form 19 August 2025; accepted 16 September 2025

KEY WORDS

Artificial intelligence;
Machine learning;
Nanoparticulate drug
delivery system;
Formulation development;
Formulation design and
optimization;
Nanomedicine;
Polymeric nanoparticles;

Abstract Nanoparticulate drug delivery systems (NDDS) have revolutionized modern medicine by significantly improving drug targeting, bioavailability, and therapeutic efficacy. Despite the clinical success of over 90 approved nanomedicines, the development of NDDS remains challenging due to the complexity of formulation design, optimization, and characterization processes. Artificial intelligence, particularly machine learning (ML), offers powerful data analytics and predictive capabilities that can address these challenges. This review systematically summarizes recent advances in ML applications across various NDDS formulations, including polymeric nanoparticles, lipid nanoparticles, liposomes, solid lipid nanoparticles, nanostructured lipid carriers, nanoemulsions, nanosuspensions, lipid-based hybrid NDDS, self-emulsifying drug delivery systems, niosomes, and nanocrystals. We also summarize how ML algorithms could help predict critical quality attributes of NDDS, such as particle size, shape,

This article is part of special column entitled: Advancing Pharmaceutics and Pharmacology published in Acta Pharmaceutica Sinica B.

*Corresponding authors.

E-mail addresses: joanzuo@cuhk.edu.hk (Zhong Zuo), cuipingjiangcpu@163.com (Cuiping Jiang).

[†]These authors made equal contributions to this work.

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

<https://doi.org/10.1016/j.apsb.2025.12.011>

2211-3835 © 2025 The Authors. Published by Elsevier B.V. on behalf of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

surface properties, drug encapsulation efficiency, drug loading efficiency, drug release behavior, and stability. Furthermore, we discuss existing challenges and prospects for the formulation development empowered by ML in NDDS. In conclusion, this review provides a comprehensive overview of the transformative potential of ML in improving the formulation development of nanomedicines, ultimately accelerating their clinical translation.

© 2025 The Authors. Published by Elsevier B.V. on behalf of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nanoparticulate drug delivery systems (NDDS) represent a revolutionary breakthrough in modern drug delivery technology. These systems employ carriers ranging in size from 1 to 1000 nm to precisely deliver drugs to target tissues or cells, thereby significantly enhancing the targeting ability, bioavailability, and therapeutic efficacy of the drugs¹. In recent years, various nanoparticulate platforms, including liposomes, polymeric nanoparticles, and nanoemulsions, have been widely utilized in the field of drug delivery². To date, over 90 nanomedicines have been approved for marketing globally, which fully demonstrates the immense clinical value of NDDS³. Nevertheless, the development of NDDS is accompanied by numerous challenges. The uncertain structure–function relationships impede the rational design of NDDS for targeted therapeutic outcomes. Besides, the optimization of the formulation and manufacturing process is complex, requiring the careful selection and fine-tuning of various components and preparation techniques such as emulsification, solvent evaporation, and nanoprecipitation, each with its specific applicable conditions and limitations. Moreover, the physicochemical heterogeneities inherent in these systems make their precise characterization particularly difficult⁴. Regardless, all the aforementioned formulation-related technical and characterization difficulties during NDDS development often led to a heavy reliance on traditional trial-and-error approaches, which consequently imposed substantial costs and time demands. Given these significant obstacles, there is a pressing need for innovative strategies to overcome these limitations⁵.

Faced with these challenges, the application of artificial intelligence (AI) technology offers new perspectives and methodologies for the development of NDDS^{6,7}. Machine learning (ML) is an application of AI that leverages statistical methods to uncover patterns in diverse data forms, including text, images, or any digitally stored information⁸. It focuses on enabling machines to learn and improve from experience, essentially using computational models to extract information and make predictions based on those patterns without being explicitly programmed^{9–11}. In the development of NDDS, the robust data processing and predictive capabilities of ML algorithms are reshaping the paradigms of formulation design, optimization, and characterization^{12,13}. For instance, ML can screen appropriate carrier materials and drugs by constructing quantitative structure–property relationship (QSPR) models for the design of NDDS^{14,15}. By analyzing vast amounts of experimental data, ML is capable of establishing correlation models between the formulation and process information of NDDS and their physicochemical properties, accelerating their optimization⁴. Additionally, ML algorithms can assist in characterizing nanoparticles (NPs) by analyzing their physicochemical

properties, such as size, distribution^{16–18}, and morphological structural features^{19–23}.

Although there are a few recent reports on the application of ML in the formulation development of liposomes²⁴ or lipid nanoparticles (LNPs)²⁵, a systematic review of ML utilization in the formulation development of diverse NDDS remains unavailable. Our current review aimed to update recent advances of ML application in the development of a broader range of NDDS, including not only liposomes and LNPs, but also polymeric nanoparticles (PLNs), solid lipid nanoparticles (SLNs), nanostructured lipid carriers (NLCs), nanoemulsions, nanosuspensions, lipid-based hybrid NDDS, self-emulsifying drug delivery systems (SMEDDS), niosomes, and nanocrystals. In addition, we summarize how ML is applied to formulation design, optimization, and characterization, highlighting ML-aided critical metrics predictions, such as particle size/distribution, shape, surface properties, encapsulation efficiency (EE), drug loading (DL), drug release behavior, and stability. Furthermore, this review discusses the existing challenges and prospects of AI-driven formulation development of NDDS, aiming to promote further clinical applications of nanomedicine technology.

To identify relevant literature, a systematic search was conducted on PubMed for publications within the last 15 years. The search terms employed were: (((ai artificial intelligence [MeSH Terms]) OR (machine learning [MeSH Terms]) OR (deep learning [MeSH Terms]) OR (artificial neural network [MeSH Terms]))) AND ((nanoparticles [MeSH Terms]) OR (drug delivery systems [MeSH Terms]) OR (nanomedicine [MeSH Terms]))) AND (“2010/01/01” [Date-Publication]: “2025/03/27” [Date-Publication]). As illustrated in Fig. 1, titles and abstracts of articles from PubMed were initially screened for their relevance to AI/ML/deep learning/ANN and NDDS. The eligible articles with full text that included the application of ML techniques in the design, preparation, or characterization of NDDS were eventually selected for the current review.

2. Overview of commonly used ML tools for NDDS formulation development

As one of the most prominent subsets of AI, ML encompasses major branches such as deep learning and artificial neural networks (ANN)²⁶, and enables systems to learn and improve from experience automatically²⁷. ML consists of four major procedures, including variable input, feature engineering, model training, and result outputs^{28,29}. Based on the learning approach, ML can be classified into supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning³⁰.

In the context of NDDS formulation development, supervised learning stands out as the most utilized ML technique. It is a task-

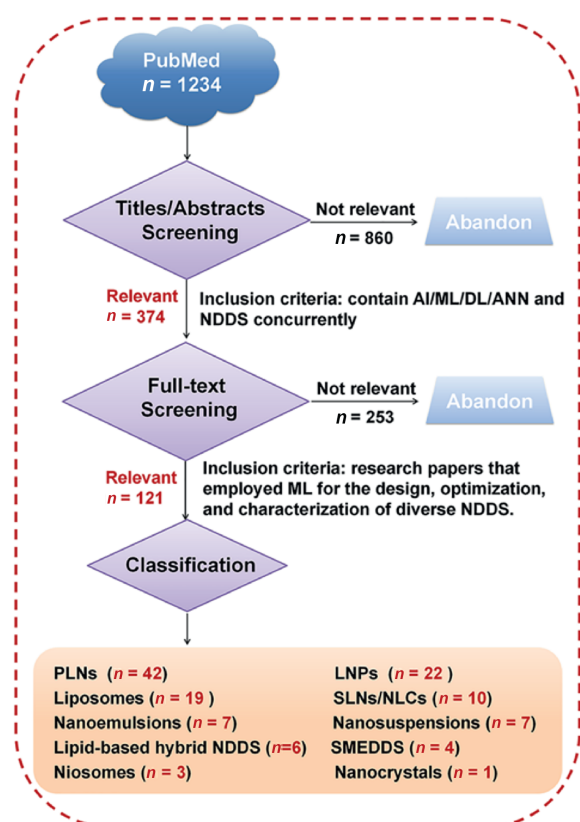


Figure 1 Flowchart of literature screening on ML empowered NDDS formulation development.

driven ML approach whose core process involves training mathematical models using labeled datasets and employing the trained models to predict outputs for new data³¹. Such a method is commonly applied to address two types of problems: classification and regression. In classification tasks, models are trained to assign new variables to predefined categories based on labeled training data. The performance of classification models is typically assessed by accuracy, precision, recall, F1 score, area under the receiver operating characteristic curve, and confusion matrices. For regression tasks, algorithms aim to estimate associations between different variables by predicting continuous outcomes. Their performance is typically evaluated using mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2)³². Based on their underlying principles, supervised learning algorithms can be broadly categorized into several distinct groups: linear models (e.g., linear regression, multiple linear regression (MLR), logistic regression), tree-based models (e.g., decision tree (DT), random forest (RF), eXtreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), least squares boosting (LSBoost)), kernel-based methods (e.g., support vector machine (SVM), gaussian processes (GP)), instance-based methods (e.g., k-nearest neighbor (KNN)), and neural network-based models (e.g., artificial neural network (ANN), multilayer perceptron (MLP), deep neural network (DNN)). Detailed descriptions of these specific algorithms are well-documented in numerous reviews^{28,33-35} and briefly summarized in Table 1. Moreover, genetic algorithms (GA) and Bayesian optimization (BO) are commonly used to enhance ML algorithms by optimizing hyperparameters, with GA broadly

exploring potential settings through evolutionary mimicry and BO more efficiently selecting subsequent settings *via* a predictive model^{36,37}.

3. Application of ML approach in NDDS formulation development

Traditionally, developing NDDS has often been a slow and expensive process reliant on time-consuming trial-and-error experimentation. Recognizing the powerful data processing and predictive capabilities of ML, researchers are increasingly turning to ML algorithms. These algorithms are now being leveraged to assist in the design, optimization and characterization of NDDS formulations, thereby accelerating the overall development cycle. Based on the number of retrievals for different NDDS shown in Fig. 1, the following section summarizes and discusses the applications of ML in the development of various NDDS in the following order: PLNs, LNPs, liposomes, SLNs, NLCs, nanoemulsions, nanosuspensions, lipid-based hybrid NPs, SMEDDS, niosomes, and nanocrystals.

3.1. Polymeric nanoparticles (PLNs)

PLNs are typically prepared from natural or synthetic polymers using physical or chemical methods³⁸. They can achieve controlled drug release and targeted delivery through the degradation of the polymer matrix and surface modifications. Based on the origin of the polymeric materials, PLNs are classified into synthetic polymer-based PLNs and natural biopolymer-based PLNs as described below³⁹.

3.1.1. Synthetic polymer-based PLNs

Among synthetic polymers, poly(lactic-co-glycolic acid) (PLGA) has gained significant attention in drug delivery due to its outstanding biodegradability, biocompatibility, and controlled drug release behavior, making it a preferred nanocarrier⁴⁰. Nanoprecipitation is a popular and scalable method for the preparation of PLGA-based PLNs (termed PLGA-NPs), involving rapid mixing of an organic PLGA phase with an aqueous surfactant solution to induce NPs formation *via* solvent diffusion and reduced polymer solubility⁴¹. The size of PLGA-NPs in nanoprecipitation is critically influenced by factors such as the types and concentrations of anti-solvents (e.g., surfactants) and the polymer itself⁴². Building on this understanding, Seegobin et al.⁴³ developed five supervised ML models (i.e., XGBoost, RF, KNN, SVM, MLP) to predict the size and zeta potential of PLGA-NPs using the types and concentrations of PLGA and anti-solvent as inputs. It was found that XGBoost outperformed other models in predicting particle size but showed higher errors for zeta potential (Table 2⁴³⁻⁷⁴). The dichotomy between size and zeta potential predictability motivated Huwaimel et al.⁴⁴ to pioneer a hybrid framework integrating advanced preprocessing (e.g., isolation forest outlier detection, min-max normalization, one-hot encoding) with ensemble learning. Comparative R^2 analysis revealed that bagging-enhanced support vector regression (BAG-SVR) achieved unprecedented accuracy for particle size prediction, while bagging-enhanced polynomial regression (BAG-PR) excelled in zeta potential modeling, thereby addressing the limitations of single-algorithm models like XGBoost⁴³. Capitalizing on these methodological advancements, Almansour and colleagues⁴⁵ further refined the paradigm by incorporating advanced

Table 1 Summaries of commonly used supervised learning tools for NDDS formulation development.

Type	ML tools	Description	Advantages	Disadvantages	Task
Linear	Linear regression	A linear model for predicting continuous values by fitting a line to minimize prediction errors.	Easy to understand, implement, and interpret, fast training and prediction speed.	Only suitable for linear relationships.	Regression
	MLR	A statistical method that models the linear relationship between a dependent variable and two or more independent variables.	Easy to understand and implement, capable of handling multiple variables.	Sensitivity to outliers, poor at capturing nonlinearity.	Regression
	Logistic regression	A linear model for binary classification that estimates probabilities using a logistic function to make predictions.	Simple, interpretable, and efficient in handling linear problems.	Only suitable for linear modeling, sensitive to feature scaling issues.	Classification
Tree-based	DT	A model based on a hierarchical tree structure, designed to make predictions <i>via</i> recursive data partitioning.	Easy to understand, interpret, and manage non-linearities, robust against noise.	Unstable, prone to overfitting. Not suitable for big datasets.	Classification regression
	RF	An ensemble of decision trees that uses voting or averaging for large datasets.	Analogous to decision trees, offers resistant to overfitting.	Complex, long training time, poor interpretability.	Classification regression
	XGBoost	An ensemble learning algorithm based on gradient boosting that optimizes model efficiency and performance.	Efficient for large-scale data, high prediction accuracy, offers resistant to overfitting.	Complex parameter tuning, sensitive to outliers, poor interpretability.	Classification regression
	LightGBM	An efficient ensemble gradient boosting framework based on decision trees for large datasets.	Fast training, high efficiency, supports distributed computing.	Complex, poor interpretability, prone to overfitting.	Classification regression
	LSBoost	An ensemble learning algorithm that combines multiple weak regression models to create a stronger predictive model.	High accuracy, simple and easy to implement, robust to outliers.	Sensitive to noisy data, outliers, and overfitting.	Regression
Kernel-based	SVM	An algorithm that separates different classes of data by finding the optimal decision boundary with the maximum margin for small datasets.	Handles high-dimensional data, generalizes well, robust to outliers, resists overfitting.	Sensitive to parameters, long training, poor interpretability, not for large datasets.	Classification regression
	GP	A non-parametric Bayesian model that provides uncertainty estimates for predictions.	Provides uncertainty estimates, interpretable for small to medium datasets.	High computational cost, sensitive to hyperparameters, slow training time.	Classification regression
Instance-based	KNN	An algorithm that finds the K closest neighbors to a new data point and uses their outputs to determine its prediction.	Simple and easy to implement, robust to outliers, suitable for small datasets.	Complex, poor interpretability, intolerant of noise.	Classification regression
Neural network-based	ANN	A neuron-mimicking network models learn data patterns from large datasets by adjusting connection weights.	Capable of complex non-linear relationships, strong generalization and adaptive learning abilities.	High data demand, slow training, complex tuning, poor interpretability.	Classification regression
	MLP	A feedforward neural network that learns complex patterns through multiple layers of neurons	Capable of complex non-linear relationships, strong generalization.	Prone to overfitting, high data demand, complex tuning, poor interpretability.	Classification regression
	DNN	A neural network with multiple layers, incorporating feedback and feedforward loops.	Capable of complex non-linear relationships, excels in complex tasks like image, speech, and text.	High data demand, slow training, complex tuning, poor interpretability.	Classification regression

MLR, multiple linear regression; DT, decision tree; RF, random forest (RF); XGBoost, eXtreme gradient boosting; LightGBM, light gradient boosting machine; LSBoost, least squares boosting; SVM, support vector machine; GP, gaussian processes; KNN, k-nearest neighbor; ANN, artificial neural network; MLP, multilayer perceptron; DNN, deep neural network.

Table 2 (continued)

Type of polymer	Type of PLNs	Preparation method	Main ML tool	Input feature	Output feature	Ref.
Natural polymer	PLA-PEG-PLA-NPs	Nanoprecipitation	ANN	Polymer conc., amount of drug, solvent ratio and mixing rate	Size	59
	Poly(beta-amino ester)-NPs	Not mentioned	RF	Polymer properties	Transfection efficiency	60
	Antibiotics-derived polymers-NPs	Not mentioned	SVR	Polymer properties	Transfection efficiency	61
	Poly(allyl methacrylamide)-NPs	Not mentioned	BO, SHAP	Polymer properties	Transfection efficiency	62
	Chitosan-NPs	Ionotropic gelation	ANN	Cs conc., TPP conc., Cs: TPP mass ratio	Size, zeta potential, yield %	65
			ANN	PEG chain length, chitosan/PEG ratio, pH of solution	Size, zeta potential	67
		Polyelectrolyte complexation	ANN	Chitosan and albumin conc., pH, and reaction time	Size, DL, cytotoxicity	68
		Ultrasonication	ANN	Chitosan conc., buffer pH, amplitude and time of sonication	Size	69
		FNP	Bagging	The numbers of reactive group and monomer in cationic/anionic polymer, the conc. of cationic/anionic polymer, flow speed	Size, PDI	70
	Agar nanospheres	Natural polymer coacervation	ANN, GA	CaCl ₂ conc., agar conc., HPβCD conc., homogenizer speed	Size, PDI, zeta potential, DL, drug release	71
	Fucoidan-NPs	Electrostatically driven complexation	ANN-DOE	Fucoidan: polyethyleneimine ratio	Size, EE	72
	Gelatin-NPs	Nanoprecipitation and two-step desolvation	ANN, MLR	Gelatin conc., glutaraldehyde conc., poloxamer 407 conc.	Size, EE	73
	HA-derivative NPs	Microfluidics	ANN	Flow rates, polymer conc.	Size	74

^aGCN, graph convolutional network; FCNN, fully connected neuronal network.

preprocessing strategies (*i.e.*, leave-one-out encoding and local outlier factor) alongside the Bat optimization algorithm to enhance Bagging and adaptive Boosting-based KNN (*i.e.*, Bag-KNN, ADA-KNN), and Bat-optimized small-size KNN regression (SBNKR) models. Results demonstrated that ADA-KNN outperformed other models in particle size prediction, while SBNKR achieved exceptional accuracy for zeta potential by integrating generative adversarial networks and deep feature extraction to address data sparsity. Collectively, these studies highlight the key role of integrating advanced preprocessing, algorithmic optimization, and ensemble techniques to enhance the predictive efficacy of ML models in PLNs formulation optimization.

Beyond optimizing PLGA-NPs prepared *via* nanoprecipitation, multiple ML techniques, including linear regression, ANN, GP, SVR, and DT, have been employed to accelerate the formulation and process optimization of PLGA-NPs through methods such as the modified supercritical antisolvent technique⁴⁶, electro-spraying⁴⁷, double emulsion⁴⁸, microfluidics methods^{49,50}, and etc.⁵¹ (Table 2). Besides treating physicochemical parameters (*e.g.*, size, zeta potential, PDI, drug encapsulation efficiency (EE), drug loading efficiency (DL)) as output parameters, therapeutic outcome was also regarded as an output indicator. For example, Dong et al.⁴⁸ developed GP algorithms to reveal the quantitative relationships between formulation variables and IC₅₀ in ovarian cancer cells for double-emulsion-derived PLGA NPs. To address data scarcity, Mihandoost and coworkers⁴⁹ innovatively constructed a DT algorithm with a tabular generative adversarial network (TGAN) to augment experimental datasets by synthetically generating training and testing subsets (Fig. 2)⁴⁹. This approach aided in uncovering the factors affecting the particle size of PLGA-NPs synthesized *via* microfluidics. Feature importance analysis revealed that the concentration of poly(vinyl alcohol) (PVA) and the ratio of lactide to glycolide (LA/GA) in the PLGA copolymer are the main determinants.

In addition to PLGA-NPs formulation, ML has aided optimization of other PLNs, including polycaprolactone (PCL)-NPs⁵², polymethacrylate-NPs^{53,54}, polydimethylsiloxane-NPs⁵⁵, and poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate) (PHBV)-NPs⁵⁶, as summarized in Table 2. Among the ML tools used, ANN is the

most widely used model^{53,56-59}. Studies have shown that the ANN model accurately predicts particle size for various PLNs, with polymer surfactant activity being key for PLNs⁵⁷. Additionally, the polymer molecular weight and concentration are key for block/tri-block copolymer NPs^{58,59}. Furthermore, beyond their application in small-molecule drug delivery, PLNs have emerged as promising platforms for the delivery of nucleic acids. Multiple ML methods were used to explore the effect of polymer properties on transfection efficiency for the rational design of gene vectors, including RF⁶⁰, SVR⁶¹, and BO with Shapley Additive exPlanations (SHAP) analysis⁶². For example, to elucidate the structure-function relationship of the poly(beta-amino ester) polymer, Gong et al.⁶⁰ developed an RF algorithm to predict the *in vitro* DNA transfection efficiency by vectorizing its structure in a machine-readable format, achieving moderate correlation (0.57–0.66) with experimental results in RAW264.7 macrophages and Hep3B liver cancer cells.

3.1.2. Natural biopolymer-based PLNs

Natural biopolymer-based PLNs represent a diverse class of delivery systems, including polysaccharide-based NPs (*e.g.*, chitosan derivatives, starch, cellulose, carrageenan, and alginate NPs), protein-based NPs (*e.g.*, gelatin NPs), and lignin NPs, etc.⁶³. In the context of ML-driven formulation design of natural biopolymer-based PLNs, chitosan and its derivative-based NPs are the most extensively studied. Ionotropic gelation is the most commonly used method for their preparation *via* electrostatic interactions between chitosan and polyanions (*e.g.*, tripolyphosphate (TPP)) under acidic conditions⁶⁴. ANN-based models demonstrated that TPP concentration had the most significant impact on both particle size and yield for chitosan-NPs⁶⁵. Given that unmodified chitosan-NPs tend to aggregate under neutral conditions⁶⁶, Bozuyuk et al.⁶⁷ adopted a PEGylation strategy to modify chitosan-NPs. ANN models equipped with a single-layer Levenberg-Marquardt algorithm and a four-layer gradient descent method were selected to optimize PEGylated chitosan-NPs due to their strong predictive capabilities for size and zeta potential, respectively. Besides ionotropic gelation, ANN models were also used to optimize the formulation of chitosan-NPs prepared by polyelectrolyte

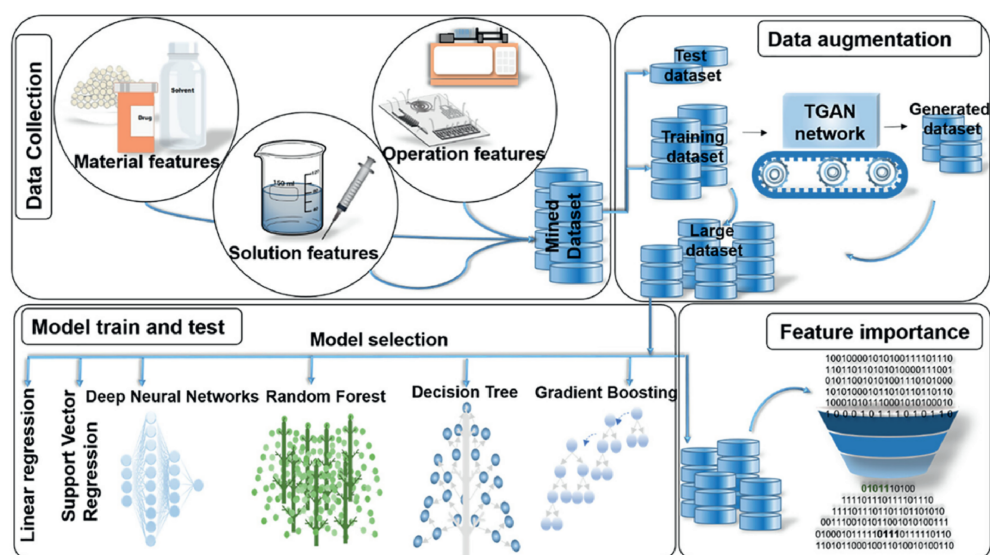


Figure 2 Workflow for ML-aided formulation optimization of PLGA-NPs synthesized *via* microfluidics. Reprinted with the permission from Ref. 49. Copyright © 2025 American Chemical Society.

fine-tuning strategies to screen ionizable lipids. Unlike other studies, they introduced the BalMol module in the framework by balancing the similarity between adjacent targets to mitigate data imbalance. Consequently, TransLNP reduced the MSE value to 1.47 on the AGILE dataset.

Significant progress has been made on ionizable lipids, but the roles of other LNP components (*e.g.*, PEGylated lipids, cholesterol) and their complex interactions are still poorly understood, hindering rational LNP design. Evidence suggests these components also impact mRNA transfection efficiency^{87,88}. For instance, Bae et al.⁸⁷ utilized an RF algorithm to explore the impact of the type of ionizable lipids, molar ratios of ionizable lipids/neutral helper lipids (DOPE, DSPC)/cholesterol, and N/P ratio on mRNA transfection efficiency. Feature importance analysis revealed that, in addition to ionizable lipids, the molar ratio of the saturated phospholipid DSPC also plays a critical role in mRNA transfection efficiency. By integrating a high-throughput screening platform with ML algorithms, Cheng and colleagues⁸⁸ systematically evaluated the impact of chemical structure and molar percentage of neutral helper lipids on gene transfection efficiency across six cell lines (Fig. 3C)⁸⁸. Findings revealed that lightGBM showed more accurate prediction than RF algorithm, notably for B16F10, HepG2, and PC3 cells as evidenced by low MAE values (<0.06). Most importantly, they determined that fine-tuning helper lipid parameters is necessary for achieving cell-type-preferential transfection.

The versatility of LNPs also stems from their tunable surface properties. To enhance mRNA delivery to microglial subtypes, Rafiei et al.⁸⁹ constructed HA-conjugated LNPs to enhance mRNA therapy for neuroinflammatory disorders. Four supervised ML classifiers (*i.e.*, RF, SVM, MLP, naïve Bayes) were employed to predict the impact of the design parameters on transfection efficiency and microglial phenotypic polarization. Among them, MLP demonstrated the best performance, with an F1-score of ≥ 0.8 . Subsequent SHAP analysis highlighted the significance of HA modification in improving transfection efficiency and driving microglia towards an anti-inflammatory state. The transfection time of LNP-mRNA influences cellular state. Harrison et al.⁹⁰ developed a pipeline to explore the temporal response of cells to RNA therapeutics. This pipeline utilized automated microscopy, deep learning, and gradient boosting machine models on time-series data to assess LNP-mRNA effects in HepG2 cells. Beyond using transfection efficiency as an output parameter, Ouyang's group⁹¹ employed LightGBM to predict the LNP formulation for mRNA vaccines based on the IgG titer. Due to LightGBM's high accuracy in classification and regression, they also used it to predict the *in vitro* and *in vivo* gene knockdown rates of siRNA-LNPs⁹².

Beyond the lipid composition and surface modifications, LNPs' transfection efficiency is also largely influenced by physicochemical parameters. Consequently, employing ML to predict physicochemical parameters (*e.g.*, size, PDI, EE, DL) is crucial for accelerating LNPs formulation development. Tree-based methods like XGBoost and RF have proven particularly effective for screening LNPs formulations^{93,94}, probably because their ordered binary decision nodes enable them to more accurately identify and fit non-linear and complex patterns compared to other algorithms⁸⁸. Furthermore, research by Jeong's group⁹⁵ showed that a combinatorial ANN-DOE model outperformed XGBoost in predicting size, PDI, zeta potential, and EE for mRNA-LNPs with a minimum number of experiments. Optimizing LNPs formulations to meet multiple outputs simultaneously is challenging. Maharjan et al.⁹⁶ addressed this by initially using principal component

analysis (PCA) to reduce eight critical parameters. They then compared BO-optimized XGBoost against a self-validated ensemble model (SVEM) for optimizing mRNA-LNP. It was found that the SVEM model proved superior predictive accuracy (>97%) owing to its weighted bootstrapping that eliminated the need for separate training/validation datasets.

Taken together, ML has become a critical driver in optimizing and accelerating LNPs formulation development. By predicting the properties of key lipids, particularly ionizable cationic lipids, understanding complex component interactions, and optimizing physicochemical parameters, ML significantly advances the design and development of non-viral gene delivery systems.

3.3. Liposomes

Liposomes are spherical vesicular systems composed of phospholipids and cholesterol that self-assemble into unilamellar or multilamellar architectures⁹⁷. Owing to their structural resemblance to natural biological membranes and tunable surface properties, liposomes have been extensively utilized for drug delivery⁹⁸. Since the US Food and Drug Administration (FDA) first approved the liposomal product Doxil® in 1995, dozens of liposome-based formulations have been clinically approved⁹⁹. To further advance the development of liposomes, researchers are increasingly adopting ML technologies to optimize liposomal formulations and manufacturing processes. Previous review on the application of ML in liposome development focused on the refinement of key parameters to accelerate the design and optimization process of liposome²⁴. We have further summarized the detailed ML techniques utilized, as shown in Table 3¹⁰⁰⁻¹¹⁴ and illustrated below.

Since ANN demonstrates exceptional capability in identifying complex patterns and establishing nonlinear relationships between independent variables and dependent outcomes, it has emerged as the most widely utilized approach in liposomal formulation design¹⁰⁰⁻¹⁰⁵ (Table 3). For example, Azubuikwe's team¹⁰⁰ employed ANN to investigate the critical process parameters affecting curcumin-loaded liposomes. The ANN-driven optimization method revealed that maximum EE was achieved under the following conditions: lipid/curcumin ratio = 4.35, sonication time = 15 min, and hydration volume = 25 mL. Liposomes exhibit a dual drug loading mechanism, allowing for the incorporation of hydrophobic drugs into the phospholipid bilayer and the effective encapsulation of hydrophilic therapeutics within the aqueous core¹⁰⁶. However, the limited aqueous core capacity of conventional liposomes often results in low EE for hydrophilic therapeutics. To address this limitation, Metwally's team¹⁰¹ utilized an ANN to markedly improve the EE of sodium salicylate by adjusting the amounts of gelatin and cholesterol in gelatinized-core liposomes. Numerous studies have demonstrated that ANN exhibits superior predictive performance over response surface methodology (RSM)^{103,104}, and MLR¹⁰⁵ in liposomal formulation design and characterization. However, ANN can be limited by its restricted exploration of the factor space and tendency towards local optima, preventing deep insights into complex systems. Molecular descriptors, which numerically represent molecular properties, are well-suited for predictive modeling¹⁰⁷. Sansare et al.¹⁰⁸ illustrated that integrating these descriptors with ANN can more accurately predict the relationships between liposome processing parameters (*e.g.*, hydrocarbon tail length, cholesterol percentage, buffer type, solvent temperature, and flow rate) and key quality attributes (*e.g.*, particle size and PDI) of liposomes.

Table 3 Summary of common ML tools used in liposomal formulation development.

ML tool	Input feature	Output feature	Ref.
ANN	Sonication time, hydration volume, lipid: curcumin ratio	EE, flux	100
ANN	Amount of gelatin/cholesterol	DL	101
ANN	Cholesterol/PEGylated lipid concentration, TFR, FRR	Size, PDI	102
ANN, RSM	Lipid content, ultrasound power/time	EE, lipid absolute loading	103
ANN, RSM	Pilocarpine hydrochloride, sodium deoxycholate, water content	EE	104
ANN, MLR	Amount of cholesterol/edge activator, the location of the drug into the vesicle, addition of stearylamine, type of edge activator	Size, PDI, zeta potential, EE	105
ANN, molecular descriptors	Hydrocarbon tail length, cholesterol content, osmolarity, temperature, flow rate	Size, PDI	108
ANN, SVM	TFR, aqueous: Organic volume ratio, total lipid conc.	Size, dispersity, stability	109
LSBoost	Sonication time, extrusion temperature, compositions	Size, PDI	110
LSBoost	Lipids molar ratios, particle size, sonication time, pH, PDI	EE	111
XGBoost	Amount of PEGylated lipids/cholesterol, chip geometry, FRR	Size	112
Lasso regression	Lipid composition/conc., FRR, TFR, solvent, buffer, temperature	Size, PDI	113
LightGBM, RF	Preparation time, sonication, filter diameter, preparation temperature, drug: lipid ratio, weight percent of lipid, lipid composition	Size, PDI, zeta potential, EE	114

Conventional top-down liposome preparation methods (*e.g.*, thin film hydration, solvent injection) often result in solvent residues, high PDI, and poor reproducibility¹⁰². Microfluidics offers a better, bottom-up approach for uniform vesicles with improved reproducibility and lower PDI, controlled by parameters like total flow rate (TFR) and aqueous-to-organic flow rate ratio (FRR)¹⁰². Recent advancements have integrated ANNs with microfluidics to accelerate liposome optimization^{102,109}. For example, Boso's group¹⁰⁹ developed an ANN model to investigate the influence of lipid concentration, TFR, and FRR on the particle size of curcumin-loaded liposomes *via* microfluidic systems. They found TFR had the most significant impact on liposome size, followed by lipid concentration and FRR.

Besides ANN, the LSBoost algorithm also demonstrated significant advantages in predicting the relationships between liposomal formulations, manufacturing processes, and their physicochemical parameters^{110,111}. For example, Eslami's team¹¹¹ developed an LSBoost algorithm to investigate the effects of lipid molar ratios, sonication time, aqueous phase pH, particle size, and PDI on the EE of curcumin-liposome. The algorithm predicted that the maximum EE of curcumin could be achieved with an HSPC: DPPG: Chol: DSPE-mPEG2000 M ratio of 55:5:35:5, a sonication time of 30 min, a pH range of 7.2–8, a particle size of 130 nm, and a PDI range of 0.09–0.13. Furthermore, XGBoost¹¹² and Lasso regression¹¹³ models demonstrated reliable prediction of liposome formation and size throughout a broad design space during microfluidic production. To construct a general predictive model for liposomal formulation development, Ouyang's group¹¹⁴ compiled a dataset of 665 formulations, including various lipid components and their ratios, along with corresponding particle size, PDI, zeta potential, and EE. Five algorithms were employed to identify the critical factors in liposome formulation design. Comparative analysis demonstrated that the LightGBM model exhibited the lowest MAE and the highest R^2 value in predicting size and zeta potential, whereas RF demonstrated superior predictive performance in terms of PDI and EE. Feature importance analysis revealed that drug molecules exhibiting a logS range of –3 to –6, molecular complexity between 500 and 1000, and XLogP3 values ≥ 2 were preferred to be prepared as liposomes with high EE¹¹⁴. To facilitate global research, their group established FormulationAI, an efficient platform for predicting liposome formulations (<https://formulationai.computpharm.org/>)¹¹⁵.

3.4. Solid lipid nanoparticles (SLNs) and nanostructured lipid carriers (NLCs)

Lipid nanoparticles represent a broad category of NDDS composed of lipids. Beyond the LNPs mentioned above for mRNA/siRNA delivery, they can be classified as SLNs and NLCs based on the physical state of the lipids¹¹⁶. SLNs consist of a monolayer shell of surfactant and a solid lipid core at physiological temperature¹¹⁷. The highly ordered crystalline structure of the lipid core in SLNs imposes stringent requirements on the payload drugs. To simulate drug loading in SLNs, Hathout and coworkers¹¹⁸ developed a GP-based supervised ML algorithm to predict DL in tripalmitin-based SLNs by correlating drug descriptors with their molecular docking binding energies. To optimize the formulation and process of SLNs, multiple ML methods, including ANN, DT, Lasso, Ridge, SVM, KNN, GP, MLP, RF, Bagging, and gradient boosting, have been employed^{119,120}. For instance, Öztürk et al.¹²⁰ developed 12 ML algorithms to explore the influence of homogenization time and lipid concentration, and surfactant concentration on SLN particle size. Their study demonstrated that the DT model outperformed other algorithms in predicting particle size under different conditions. It was found that homogenization time strongly influenced particle size, with longer durations resulting in smaller particles and a moderate decrease in PDI. Lipid concentration had a weaker effect, showing a slight positive correlation with particle size.

Despite the numerous advantages of SLNs, such as enhanced oral bioavailability and biocompatibility, the highly ordered crystalline structure of the core solid lipid limits payload efficiency and might lead to drug leakage during storage¹²¹. To address these issues, the more advanced NLCs have been developed. As the second generation of SLNs, NLCs could improve loading capability by mixing solid lipids with liquid lipids to form an amorphous solid lipid matrix, thereby maintaining drug stability during storage^{97,122}. This underscores the critical importance of lipid selection in formulation design. In a pioneering study, Basso et al.¹¹⁶ employed an integrated ML approach to evaluate two novel glycerol-based lipids for NLC optimization. They first used hierarchical clustering and PCA to assess structural similarity, finding that both lipids restored NLC's negative zeta potential with enhanced biocompatibility. Further ML analysis (*i.e.*, partial least squares regression and neural networks) of

multiple formulation factors on NLC quality attributes (size, PDI, zeta potential, cytotoxicity, drug loading) led to successful NLC optimization¹¹⁶. Besides, to tackle the major limitation of ANN concerning the requirement for extensive training data, researchers have integrated ANN with other methodologies. Hybrids such as ANN combined with Box-Behnken design and variable weight desirability functions¹²³, or ANN integrated with neuro-fuzzy logic and GA^{124,125}, demonstrate great potential. These methods effectively delineated the knowledge and design spaces for NLCs, thereby enabling more systematic formulation development.

Due to the structural and functional similarities between SLNs and NLCs, researchers have increasingly leveraged ML algorithms to concurrently optimize their formulation and manufacturing process^{126–128}. For instance, Bao and colleagues¹²⁸ combined ML methods with high-throughput and miniaturized automation to speed up formulation development for cannabidiol-loaded SLNs and NLCs. In their process, automated lab equipment was first used to prepare a small subset of formulations, generating data to train eight different ML regression algorithms. They found that RF, DT, and XGBoost models were the best predictors for particle size, DL, and EE, respectively. These selected models then generated a synthetic dataset ten times larger than the initial dataset to predict optimal formulations. Impressively, these predicted formulations greatly enhanced drug solubility (up to 3000-fold), prevented degradation, and improved oral bioavailability. Collectively, ML algorithms are increasingly crucial for optimizing SLNs and NLCs. By analyzing the impact of complex factors (*e.g.*, drug properties, process parameters, and lipid composition) on NDDS characteristics (*e.g.*, particle size, EE, DL), researchers can more effectively optimize these systems, overcome their inherent limitations, and significantly enhance drug delivery performance.

3.5. Nanoemulsions

Nanoemulsions are colloidal dispersions (20–200 nm droplets) of oil and water stabilized by surfactants and co-emulsifiers, appearing transparent to translucent due to weak light scattering¹²⁹. Unlike conventional emulsions, nanoemulsions exhibit both thermodynamic and kinetic stability, preventing flocculation and phase separation through steric hindrance. Their high specific surface area and permeation enhancement make them ideal for improving hydrophobic drug bioavailability^{130,131}. Nanoemulsions preparation is broadly categorized into high-energy and low-energy methods. High-energy methods, like ultrasonic emulsification and high-pressure homogenization, dominate large-scale production^{132,133}. ANN is the most widely adopted method in the development of nanoemulsions^{130,134–137}. Among these studies, most ANN models have found that surfactant concentration is the primary factor determining the size of nanoemulsions, regardless of whether ultrasonic emulsification¹³⁴ or high-pressure homogenization^{130,135} was used. This could be because surfactants adsorb at the water/oil interface, requiring more surfactant for smaller droplets due to their larger specific surface area. Interestingly, Samson et al.¹³⁶ identified xanthan gum as the key factor influencing particle size in virgin coconut oil nanoemulsion systems using a GA-enhanced ANN model. In addition to investigating the effect of formulation factors on particle size, Ngan et al.¹³⁷ investigated the impact of process parameters such as sonication amplitude, duration, and homogenization speed on the droplet size and viscosity of nanoemulsions. To conduct this analysis, they trained an ANN model using four distinct algorithms (*i.e.*,

incremental back propagation (IBP), batch back propagation (BBP), quick propagation (QP), and Levenberg-Marquardt (LM)). The results demonstrated that ultrasonic cavitation outperformed high-shear homogenization in regulating droplet size (*via* sonication time) and viscosity (*via* amplitude modulation). Additionally, the predictive capabilities of each learning algorithm ranked in the following order: LM > IBP, BBP > QP.

The clinical translation of nanoemulsions faces challenges such as limited shelf-life stability and cytotoxicity risks associated with surfactants¹³⁸. To address these issues, Amani's team¹³⁹ employed ANN to analyze four formulation variables (*e.g.*, the concentration of NP, oil, surfactant, and alcohol) influencing the short-term stability of superparamagnetic iron oxide-loaded nanoemulsions. The ANN model revealed that all variables exhibited a negative correlation with stability beyond their optimal thresholds. Maximum stability was achieved at specific ratios: 15.8% v/v Span 80/Tween 80, 3.6% v/v almond oil, 0.3 mg/15 cm³ NP concentration, and 2.3% v/v ethanol. Building on this work, they utilized an ANN to further investigate formulation factors governing both stability and cytotoxicity in a novel nanoemulsion system. The results revealed that the most critical component for stability was the surfactant concentration, while cytotoxicity was mainly influenced by the oil concentration¹³⁹.

3.6. Nanosuspensions

Nanosuspensions are colloidal drug dispersions in liquid media, offering advantages such as improved dissolution/solubility, scalable production, and enhanced stability¹⁴⁰. They are generally prepared using top-down (*e.g.*, media milling, high-pressure homogenization) or bottom-up (*e.g.*, antisolvent precipitation) methods¹⁴¹. Wet media milling is one of the most commonly used top-down methods¹⁴², and it reduces drug particle size through collisions. However, prolonged milling often raises temperature, posing challenges for thermolabile drugs. To address this, Bilgili's group^{143,144} developed a semi-theoretical lumped parameter model enhanced with power law and KNN models to predict the temperature evolution of the solution during the preparation of nanosuspensions, providing insights for controlling heat and optimizing the process for thermolabile drugs. In contrast, the bottom-up antisolvent precipitation method creates a supersaturated solution by mixing the drug solution with anti-solvent¹⁴⁵. Abdelbary and colleagues¹⁴⁶ utilized this method with ultrasonication to prepare carvedilol nanosuspensions. By combining ANN with GA, they optimized the nanosuspensions based on particle size, PDI, and zeta potential, finding Pluronic F127 to be the optimal stabilizer. Similarly, ANN models were also extensively used to optimize acetaminophen and iodine (¹²⁷I) nanosuspensions produced in microchannel reactors^{147–149}. For example, Amani's team¹⁴⁷ employed ANN to study acetaminophen nanosuspensions. They investigated how surfactant concentration, solvent/anti-solvent flow rates, and temperature affected sedimentation time and PDI. The results showed that anti-solvent flow rate and temperature positively influenced precipitation time, while solvent flow rate had an inverse effect. Surfactant concentration was identified as the most critical factor for stability. A deeper ANN-based analysis by the team showed that optimal PDI was achieved with higher anti-solvent flow rates and temperature, coupled with minimal solvent flow rate¹⁴⁸.

Besides optimizing formulation and process parameters, ANN enables real-time quality monitoring during continuous manufacturing processes. A notable example is the work by

Davidopoulou and colleagues¹⁵⁰. They integrated ANN with digital twin technology to predict and optimize the production of nanosuspensions. Briefly, the ANN framework was employed for data sampling, model deployment, and dynamic simulation of milling processes, allowing precise estimation of particle size distribution at any stage of grinding. By combining ANN with digital twin technology, real-time manufacturing data can be used to dynamically adjust process parameters. This collaborative approach allows for proactive control and prediction of both process progression and the quality control of the final product¹⁵⁰.

3.7. Lipid-based hybrid NDDS

Lipid-based hybrid NDDS refer to nanocarriers engineered by combining lipid components with other materials, such as polymers, dendrimers, or inorganic nanoparticles, to leverage the unique properties of each constituent for enhanced drug delivery performance¹⁵¹. Due to the significant potential of ML algorithms in NDDS formulation development, their application extends beyond the aforementioned lipid-based nanocarriers to hybrid lipid systems. Among various ML methods, ANN is the most widely applied. It has been utilized for the formulation optimization of posaconazole-loaded phospholipid-based nanoformulations¹⁵², verapamil hydrochloride-loaded polymer-lipid NPs¹⁵³, and dendrimer-lipid nanocarriers¹⁵⁴. An illustrative example is the work by Henser-Brownhill et al.¹⁵⁴, who developed an ANN model for predicting mRNA delivery efficiency of dendrimer-lipid nanocarriers and RF algorithms for predicting size, PDI and zeta potential, respectively. Through the synergistic integration of these ML techniques with high-throughput screening, they created a platform capable of *in silico* evaluating millions of candidate formulations, which drastically accelerated the optimization of dendrimer lipid nanocarriers for nucleic acid delivery. Similarly, by combining the XGBoost model with high-throughput screening, Mirkin's team¹⁵⁵ achieved a tenfold reduction in the number of samples requiring testing. This allowed for a more thorough exploration of the nanomedicine-design space during the development of spherical nucleic acids.

3.8. Self-emulsifying drug delivery systems (SMEDDS)

SMEDDS are transparent/translucent mixtures of oil, water, surfactants, and co-surfactants that spontaneously form stable and homogeneous systems without mechanical energy¹⁵⁶⁻¹⁵⁸. They can improve the oral bioavailability of hydrophobic drugs by increasing solubility¹⁵⁹. To date, over a dozen FDA-approved products exist (*e.g.*, Gengraf®, Norvir®, Depakene®)¹⁶⁰, but their development is typically slow and expensive, requiring extensive screening and optimization.

Theoretical modeling, particularly using ML algorithms, provides a powerful solution to advance the SMEDDS development. By predicting solubility, particle size, and enabling phase diagram prediction, these computational methods drastically reduce the need for extensive experimental screening and optimize the formulation process. For instance, Li et al.¹⁶¹ developed a QSPR model using MLR and ANN algorithms. This model correlates the molecular structures of oils, surfactants, cosurfactants, and 2-arylpropionic acid with drug solubility in SEDDS by encoding electronic, geometric, topological, and quantum-chemical descriptors. The resulting QSPR model allows for rapid prediction of 2-arylpropionic acid solubility in SMEDDS and quick identification of optimal formulations. Beyond solubility prediction, Parikh

and coworkers¹⁶² constructed predictive models for the particle size of iloperidone self-emulsifying pre-concentrates (the initial physical state of SMEDDS) using I-optimal design and ANN methods. Comparative analysis demonstrated the superior predictive capability of ANN over the I-optimal design. Subsequent sensitivity analysis further revealed that oil composition exerted the strongest influence on microemulsion droplet size. Moreover, the critical technique for formulating SMEDDS involves mapping the ternary phase diagram of oil, surfactant, and co-surfactant to identify the self-emulsifying region. To enhance predictive accuracy for SMEDDS formulation properties, Ouyang's group¹⁶⁰ developed a pseudo-ternary phase diagram prediction model using an RF algorithm trained on a dataset of 4495 SMEDDS formulations. They then selected meloxicam as a model drug and experimentally constructed a pseudo-ternary phase diagram for meloxicam-SMEDDS. The results indicated that the constructed prediction model performed well, achieving an accuracy of 89.51%. Furthermore, molecular dynamics simulations were employed to investigate molecular interactions between excipients and the model drug, unveiling their diffusion behaviors in aqueous environments and elucidating the mechanistic role of co-surfactants. This integrated computational approach provides novel insights into rational SMEDDS formulation design. Building on this framework, they also developed a FormulationAI platform to guide the formulation design of SMEDDS¹¹⁵.

3.9. Niosomes

Niosomes are closed bilayer vesicles self-assembled from non-ionic surfactants and cholesterol in water, capable of encapsulating both hydrophilic and hydrophobic drugs like liposomes¹⁶³. Unlike liposomes, they do not require expensive phospholipids, making them more cost-effective, readily available, less prone to oxidation, and potentially lower in toxicity. Cholesterol incorporation enhances their robustness and stability. These attributes make niosomes versatile carriers for drugs and cosmetic actives^{164,165}.

Given the widespread use of niosomes over the past few decades, increasing studies leverage ML algorithms to analyze historical data, establishing correlations between formulation parameters and physicochemical properties to guide their optimization¹⁶⁶⁻¹⁶⁸. For example, Ghaemmaghami's group¹⁶⁶ developed an ML framework to predict the EE of niosomes. By collecting data from studies published over the past decade, they trained three ML models (*i.e.*, DNN, linear regression, polynomial regression) using four cost functions (MSE, MAE, MAE-M, Gaussian). The DNN model with MAE-M demonstrated the best predictive performance and identified HLB as the most influential parameter for EE. Experimental validation using seven doxycycline hyclate-loaded niosomes further confirmed the predictive accuracy of the model. Similarly, comparative analysis by Fahmy et al.¹⁶⁷ evaluated four regression algorithms (*i.e.*, categorical boosting (CatBoost), linear regression, SVR, and ANN) for EE prediction. Among them, CatBoost achieved optimal performance ($R^2 = 0.98$, $MSE < 10^{-4}$). The subsequent feature importance analysis revealed that the drug-to-lipid ratio is the primary factor influencing EE, followed by the lipid-to-surfactant molar ratio. Expanding beyond EE prediction, Shahiwal et al.¹⁶⁸ established an ANN model to predict particle size based on eleven niosome formulations and process variables, corroborating findings that drug-to-lipid and cholesterol-to-surfactant ratios are also critical determinants for size.

3.10. Nanocrystals

Nanocrystals are carrier-free submicron colloidal dispersions composed of pure drug active ingredients and stabilizers, exhibiting exceptionally high drug loading capacities. Their core advantage lies in the enhancement of solubility and bioavailability of poorly soluble drugs through nanoscale particle sizes. Additionally, nanocrystals offer stability, improved drug release, and scalable production³. They are typically produced *via* top-down (e.g., milling, homogenization) or bottom-up methods (e.g., antisolvent precipitation, spray drying)^{169,170}. Since hydrophobic drug dissolution rate and solubility inversely depend on size, controlling particle size and PDI of nanocrystals is crucial¹⁷¹. Building on this understanding, Ouyang's group¹⁷² systematically collected 910 datasets on nanocrystal size and 341 datasets on PDI. To construct regression models for predicting particle size and PDI of the nanocrystals prepared by high-pressure homogenization, ball wet milling, or antisolvent precipitation, eight ML algorithms were employed, including RF, DT, LightGBM, MLR, partial least squares, KNN, SVM, and DNN. By comparing the MAE value of each algorithm, LightGBM was determined to exhibit superior predictive performance. Feature importance analysis revealed that cycle index, milling time, and stabilizer concentration are critical factors for the preparation of nanocrystals by high-pressure homogenization, ball wet milling, and antisolvent precipitation, respectively. To further validate the generalization and predictive accuracy of the constructed ML model, six novel drug nanocrystals absent in the training datasets were prepared. Experimental measurements of their particle size and PDI demonstrated strong concordance with the values predicted by the models, confirming their robustness in practical applications.

4. ML-aided prediction of critical quality attributes in NDDS formulation development

The safety and efficacy of NDDS are highly dependent on their critical quality attributes. Even minor deviations in these parameters can significantly alter pharmacokinetic profiles or even lead to potential safety risks¹⁷³. In the context of ML empowered NDDS formulation development mentioned in Section 3, certain critical quality attributes emerge as particularly critical metrics, including particle size and distribution, morphological structure, surface charge, EE and DL, drug release behavior, and stability, as summarized below.

4.1. Prediction and characterization of particle size and distribution via ML

The particle size of NDDS not only influences critical parameters such as drug loading capacity and release kinetics¹⁷⁴, but also profoundly impacts their pharmacokinetics, biodistribution, clearance pathways¹⁷⁵, and even delivery mechanisms¹⁷⁶. Additionally, the size distribution reflects the stability and uniformity of NDDS. Consequently, precise control over both particle size and distribution is essential. Recognizing this, numerous ML algorithms have been applied in predicting particle size and distribution during formulation design and optimization of diverse NDDS. Among various ML methods, ANN has the widest application, being used in almost all the development of NDDS covered in this review^{46,47,53,57-59,65,67-69,73,74,102-105,109,130,134-137,148-150,152,162}. For scenarios with limited data or where

experimental design integration is crucial, hybrid approaches combining ANN with techniques like DOE^{72,95}, GA, molecular descriptors¹⁰⁸, variable weight desirability functions¹²³, or neuro-fuzzy logic¹²⁴ have shown great advantages. Besides, other neural network-based algorithms (e.g., neuro-fuzzy logic⁵², GCN-FCNN⁵⁴, MLP⁵⁵) have also been proven effective in predicting particle size and distribution. Another significant category is ensemble algorithms. These methods improve accuracy and robustness by leveraging the combined predictions of multiple models, often also reducing the risk of overfitting, making them highly recommended choices. Within this category, several ensemble algorithms have shown particularly strong predictive performance, including XGBoost, LSBoost^{110,111}, lightGBM^{114,172}, BAG-SVR⁴⁴, ADA-KNN⁴⁵, Bagging⁷⁰, SVEM⁹⁶, and RF. It should be noted that RF is also a typical representative of the Bagging strategy in ensemble learning. Additionally, a range of other established ML algorithms have been successfully employed for this task, such as DT^{49,120}, SVR⁵⁰, and Lasso regression¹¹³.

Furthermore, ML algorithms have been applied in the accurate characterization of these parameters. Although particle size and distribution data of NDDS are typically obtained using the dynamic light scattering method, microscopic imaging techniques (e.g., transmission electron microscopy, atomic force microscopy) offer alternative means to visualize and analyze particles¹⁷⁷. Recently, neural network-based image processing algorithms have been used for determining NDDS sizes¹⁶⁻¹⁸ and dispersity¹⁸.

4.2. Morphological structure characterization via ML

The morphological structural attributes of NDDS are fundamental determinants of their biological behavior, critically influencing interactions with biomolecules (e.g., proteins and cell membranes), drug release profiles, and intracellular trafficking pathways¹⁷⁸. Given this profound impact on performance, the morphological structure emerges as an output parameter in NDDS formulation development *via* computational methods¹⁹. Nanotechnologies yield diverse architectures, including solid nanoparticles, hollow nanocapsules, core-shell structures, or multilayer structures with common morphologies encompassing spherical, quasi-spherical, rod-like, or fibrous geometries¹. These morphological features are typically assessed using advanced imaging methodologies. Recently, the majority of studies have utilized deep learning approaches to analyze microscopic images of NDDS for the purpose of identifying their morphologies¹⁹⁻²³. Among them, many of these approaches rely on supervised learning, which requires substantial manual effort for data annotation. To circumvent these constraints, *K*-means clustering and Naïve Bayes classifier-based unsupervised ML algorithms have been engineered to autonomously extract intrinsic geometric properties of NDDS with minimal manual input, yielding substantially improved analytical reproducibility and throughput¹⁹.

4.3. Surface charge prediction via ML

The surface charge of NDDS is a critical determinant of their behavior in biological systems, profoundly influencing aggregation propensity, colloidal stability, cellular uptake, and biodistribution patterns^{179,180}. This makes it a particularly significant output parameter in ML-driven NDDS formulation development and quality control. Although surface charge itself is influenced by factors such as material composition and the dispersion medium,

measuring its zeta potential yields a crucial and quantifiable metric. Consequently, accurately predicting and controlling zeta potential is vital. ANN has become widely adopted for this purpose, successfully assessing zeta potential across diverse NDDS formulations, including polymethacrylate-based NPs⁵³, chitosan-based NPs^{65,67}, agar nanospheres⁷¹, timolol maleate-loaded liposome¹⁰⁵, and mRNA LNPs⁹⁵. Beyond conventional ANN frameworks, ensemble algorithms (such as lightGBM¹¹⁴, XGBoost^{43,93,94}, RF, BAG-PR⁴⁴, and SBNRR⁴⁵) are often preferred because they can integrate the advantages of different models, typically providing more accurate and robust predictions, making them a worthy option for prioritization.

4.4. Drug encapsulation efficiency (EE) and drug loading efficiency (DL) predictions via ML

EE enhances therapeutic stability under physiological conditions and enables controlled release kinetics. Optimal DL further ensures clinical efficacy aligned with dosing regimens¹⁸¹. EE and DL are closely related to the formulation composition and preparation process of NDDS. Alongside size and PDI, EE and DL are the most common output parameters for NDDS optimization^{50,51,72,105}. To streamline the design and optimization of diverse NDDS, ML algorithms have been widely used to predict EE and DL. This approach mirrors the application of ML in predicting NDDS size and PDI, where ANN^{46,58,68,73,100,101,103-105} and their hybrid models^{71,72,95,96,123} were predominantly employed. Ensemble algorithms, such as RF, XGBoost^{93,94}, LSBoost¹¹¹, CatBoost¹⁶⁷, and SVEM⁹⁶, also demonstrated superior predictive performance. Moreover, the study also found that GP^{48,51}, linear regression⁵⁰, and DNN¹⁶⁶ have been proven effective for EE and DL predictions.

Besides optimizing formulation and process parameters^{100,101,103,104,111}, ML methodologies demonstrate unique capabilities in screening drug candidates with superior encapsulation potential for various NDDS via AI-driven QSPR models, such as liposomes^{14,182} and polymeric micelles^{15,183,184}. Furthermore, due to the complexity of directly measuring EE and DL, a GP-based framework was established to predict EE and DL indirectly by generating predictive maps based on particle size data. Such maps facilitate the rapid identification of optimized PLGA-NPs, thereby significantly reducing reliance on trial-and-error experimentation⁵¹. Moreover, it has been reported that GP-based supervised ML algorithms have also been applied to predict DL of any drug in the studied tripalmitin-based SLNs by correlating drug descriptors with their molecular docking binding energies¹¹⁸.

4.5. Drug release behavior prediction via ML

In vitro drug release behavior serves as a critical quality control indicator for NDDS and provides valuable insights into their potential *in vivo* performance. This performance can profoundly influence absorption, *in vivo* safety and efficacy profiles, and even the long-term stability of the formulation^{185,186}. To better understand the underlying release mechanisms, several mathematical models have been developed to describe drug release behavior, such as zero-order, first-order, Hixson-Crowell, Higuchi, Weibull, Gompertz, and Korsmeyer-Peppas models¹⁸⁷. However, these empirical models are primarily useful for post hoc analysis of experimental data and lack the predictive power necessary for guiding formulation design.

To gain deeper insights into drug release dynamics and enhance predictive capabilities, researchers are increasingly adopting ML empowered approaches¹⁸⁸. In the context of PLNs, various algorithms have been utilized to forecast *in vitro* release behavior, such as ANN^{56,189}, GA-ANN^{71,190}, GP¹⁹¹, and LightGBM¹⁹⁰⁻¹⁹². Among these research studies, Allen's group¹⁹² conducted a systematic investigation comparing the performance of ten ML algorithms for predicting the release kinetics of polymer-based long-acting injectables. The results showed that tree-based models (*e.g.*, DT, RF, XGBoost, natural gradient boosting, LightGBM) generally achieved higher prediction accuracy than linear (*e.g.*, partial least squares, Lasso, MLR), instance-based (*e.g.*, KNN), and deep learning models (*e.g.*, neural network). Specifically, LightGBM demonstrated the highest predictive accuracy¹⁹⁰⁻¹⁹². Nevertheless, the limited quantity and heterogeneity of available data often constrain the predictive accuracy of these ML models. To address this, Bannigan et al.¹⁹² developed a larger dataset containing 181 drug release profiles with 3783 individual fractional release measurements for 43 unique drug-polymer combinations. Their analysis also confirmed LightGBM as optimal for predicting drug release from PLGA, PLA, or PCL-based polymeric systems. Furthermore, beyond direct prediction, ML algorithms (*e.g.*, neural network, gradient boosting regression, DT, extra Tree, ANN) can be used to estimate key parameters like diffusion coefficients in erosion models¹⁹³, drug concentration within the polymeric matrix at various locations¹⁹⁴⁻¹⁹⁶, and partition coefficients¹⁹⁷, enabling indirect drug release prediction from PLNs.

Beyond predicting drug release from polymeric NPs, some ML methods have been applied to predict the *in vitro* release behavior of liposomes, including deep learning¹⁹⁸, ANN^{199,200}, and RF²⁰¹. These models can forecast drug release profiles not only under simulated physiological conditions but also under varying pH environments¹⁹⁸ and different ultrasound frequencies^{200,201}. Furthermore, predictive models have been successfully constructed for diverse carriers such as lipidic mesophasic drug carriers²⁰², lecithin/chitosan NP²⁰³, and PEGylated KIT-6 NP²⁰⁴, utilizing techniques like XGBoost, DNN, and non-linear generalized ANN, respectively.

4.6. Stability predication via ML

Nanoparticle aggregation poses a significant challenge to the stability of NDDS. ML algorithms have been successfully applied to classify the aggregation states of these systems. For instance, the SVM model was proven effective in distinguishing the stable and unstable states of liposomes based on PDI data¹⁰⁹. ANN has been employed to predict the stability of nanoemulsions through transparency measurements^{139,205}. Beyond aggregation, drug leakage during storage is another critical destabilization mechanism in NDDS, significantly impacting both shelf-life and the integrity of the drug upon administration. To tackle this issue, a QSPR model established by iterative stochastic elimination and leave-one-out validation schemes can specifically target drug leakage from liposomes during storage²⁰⁶.

Taken together, ML significantly aids scientists in accurately predicting key characteristics of nanomedicines, thereby accelerating the development of safer and more effective drugs. The optimal ML algorithm selection depends on the specific task and data availability, with ensemble algorithms like LightGBM, XGBoost, and RF excelling in high-accuracy predictions for highly non-linear problems with sufficient data, while hybrid

ANN models incorporating DOE or GA are more suitable for limited data or integrated optimization. Although deep learning offers potential for image analysis, it requires extensive labeling, making unsupervised methods more practical for rapid analysis. Moreover, specialized algorithms like GP have unique advantages in predicting EE/DL through indirect parameters or screening drugs with high encapsulation potential, making them suitable for specific scenarios.

5. Conclusions, challenges and perspectives

The rapid advancement of AI, particularly ML techniques, is profoundly transforming research paradigms across multiple disciplines. In the development of NDDS, ML technology demonstrates substantial potential through its robust data analytics and predictive capabilities. Unlike the traditional, experience-dependent, and laborious trial-and-error development process, ML can efficiently analyze vast datasets. It not only aids in the design, optimization, and characterization of NDDS, but also shows potential for assisting in quality control, thereby remarkably accelerating the NDDS research pipeline. However, despite these significant advantages, the application of ML in NDDS development still faces several critical challenges.

The primary challenges hindering the application of ML in NDDS development are data insufficiency, fragmentation, and multimodality. The development of NDDS is highly complex, requiring substantial time and cost to obtain huge amounts of high-quality experimental data²⁰⁷. Besides, inconsistencies in technical standards and characterization methods, coupled with differing research focuses across disciplines, contribute to data fragmentation and imbalance. Taking NDDS size and distribution analysis as an example, some labs use the dynamic light scattering method, while others use transmission electron microscopy. However, regardless of the method used, the data are difficult to compare directly, which makes it hard to establish a unified ML model across studies. Apart from that, the inherent complexity of NDDS development, encompassing formulation design alongside systematic investigation of physicochemical properties, drug release kinetics, *in vitro* and *in vivo* efficacy, and safety profiles, generates multi-modal data. These data-related issues frequently lead to ML models suffering from overfitting or underfitting, resulting in a lack of robustness and generalizability²⁸.

Moreover, most ML models act as “black boxes”, lacking interpretability, making it hard to understand the internal logic and basis for their predictions²⁰⁸. This might hinder regulatory approval and erode trust among healthcare professionals and patients²⁰⁹. For instance, consider an ML model that suggests a particular polymer-drug ratio for a PLN formulation based on its training data. Without interpretability features, researchers cannot determine the rationale behind this recommendation. This absence of transparency makes it hard to convince regulators of the model’s reliability and complicates troubleshooting if the formulation subsequently runs into problems.

Additionally, ethical and safety concerns arise, including potential algorithmic bias, data privacy, and the potential safety and efficacy risks of ML-designed nanomedicines in clinical settings^{210,211}. For example, if historical experimental data used to train an ML model primarily comes from male subjects, the model might exhibit bias when predicting the efficacy or toxicity of a nanomedicine in female subjects, potentially leading to inappropriate clinical decisions.

Despite these hurdles, promising trends are shaping the future of ML in NDDS formulation development. Addressing the data bottleneck, researchers are increasingly extracting data from publications, utilizing open databases, and leveraging high-throughput computing and experimentation to expand data availability. Concurrently, the establishment of open nanomaterial databases and sharing platforms represents a significant advancement, enabling researchers to access and utilize standardized, high-quality data more readily²¹². Building on this enhanced data foundation, the integration of multi-omics data (*e.g.*, genomics, proteomics, metabolomics) holds immense promise for personalized nanomedicine²¹³.

To tackle algorithmic challenges, selecting methods appropriate for the specific datasets is crucial. Techniques suitable for small sample sizes, such as data augmentation, can be advantageous⁴⁹. Careful comparison of different ML algorithms’ predictive performance, coupled with advanced preprocessing⁴⁵, algorithmic optimization^{71,190}, and rigorous validation techniques like *K*-fold cross-validation, leave-one-out, and hold-out validation, can mitigate the risk of overfitting, especially on small datasets^{214,215}. Besides, hybrid approaches that combine ML algorithms with established techniques like DOE⁹⁵, combinatorial chemistry⁸⁵, or high-throughput screening^{83,154} also offer potential benefits.

To enhance the interpretability of ML algorithms, the focus on Explainable AI (XAI) is increasing²¹⁶. This shift involves moving beyond basic tools like SHAP and Local Interpretable Model-agnostic Explanations towards more sophisticated methods that offer deeper insights into model decision-making^{217,218}. Such enhanced transparency is vital not only for scientific understanding but also for gaining trust from regulatory bodies and clinicians²⁰⁹. Furthermore, addressing data privacy concerns, particularly with sensitive NDDS data, federated learning provides a valuable solution. It allows ML models to be trained across institutions without sharing raw patient data, facilitating large-scale collaborations while preserving data security²¹⁹. Other techniques, including multi-task learning, active learning, data synthesis, transfer learning, and one-shot learning, also contribute by improving the utilization of constrained or sensitive data while maintaining privacy and boosting model utility²²⁰.

Furthermore, collaborations between academia, industry, and regulatory bodies are also essential to address ethical and privacy concerns. Joint workshops, pilot projects, and the development of shared ethical guidelines can facilitate dialogue and cooperation^{33,221}. For instance, regulatory agencies like the NMPA or FDA could collaborate with academic researchers and industry partners to establish standardized evaluation protocols for ML-designed NDDS development. Crucially, the path from ML-optimized formulation to successful clinical translation is fraught with its own set of challenges. Key hurdles include scaling from lab to clinical manufacturing, maintaining quality control, navigating regulations requiring safety/efficacy data, and the fact that predictive models must be validated by *in vivo* tests and human trials to confirm performance and safety. Overcoming these requires interdisciplinary teamwork and a careful, validation-driven process. Close collaboration between computational scientists, experimental biologists, clinicians, and regulatory experts is vital to bridge the gap between promising ML predictions and successful clinical outcomes.

In summary, ML shows great promise for NDDS development. Due to potential hurdles like data acquisition, interpretability, ethics, and regulation, proactive approaches such as active data

integration, algorithm optimization, explainable AI, and cross-disciplinary and institutional collaboration are highly encouraged to accelerate formulation development and advance clinical translation of nanomedicine.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (No. 82104402), the Regulatory Science Research Project of the Guangdong-Hong Kong-Macao Greater Bay Area Center for Drug Evaluation and Inspection of National Medical Products Administration (No. GBA-JGKX-2507, China), Traditional Chinese Medicine Research Fund of Guangdong Provincial Bureau (No. 202405061017183700, China), and the Natural Science Foundation of Guangdong Province (No. 2023A1515010720, China).

Author contributions

Chunyan Shen and Mengyan Zhang conducted the literature review, data collection and wrote the manuscript. Meiting Lu, Errong Chang and Ziting Gao participated part of the literature review and data collection. Weikang Ban performed the drawing and part of the literature review. Qiang Liu revised the manuscript. Zhong Zuo and Cuiping Jiang designed the research and revised the manuscript. All of the authors have read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- Jayasinghe MK, Lee CY, Tran TT, Tan R, Chew SM, Yeo BZJ, et al. The role of *in silico* research in developing nanoparticle-based therapeutics. *Front Digit Health* 2022;**4**:838590.
- Afzal O, Altamimi AS, Nadeem MS, Alzarea SI, Almalki WH, Tariq A, et al. Nanoparticles in drug delivery: from history to therapeutic applications. *Nanomaterials* 2022;**12**:4494.
- Jia YC, Jiang YX, He YL, Zhang WT, Zou JH, Magar KT, et al. Approved nanomedicine against diseases. *Pharmaceutics* 2023;**15**:774.
- Zaslavsky J, Bannigan P, Allen C. Re-envisioning the design of nanomedicines: harnessing automation and artificial intelligence. *Expert Opin Drug Deliv* 2023;**20**:241–57.
- Das KP, J C. Nanoparticles and convergence of artificial intelligence for targeted drug delivery for cancer therapy: current progress and challenges. *Front Med Technol* 2023;**4**:1067144.
- Noorain Srivastava V, Parveen B, Parveen R. Artificial intelligence in drug formulation and development: applications and future prospects. *Curr Drug Metabol* 2023;**24**:622–34.
- Sun LM, Liu HM, Ye YQ, Lei Y, Islam R, Tan SM, et al. Smart nanoparticles for cancer therapy. *Signal Transduct Targeted Ther* 2023;**8**:418.
- Kolluri S, Lin JC, Liu R, Zhang YW, Zhang WW. Machine learning and artificial intelligence in pharmaceutical research and development: a review. *AAPS J* 2022;**24**:19.
- Gormley AJ. Machine learning in drug delivery. *J Control Release* 2024;**373**:23–30.
- Fan KJ, Cheng LJ, Li L. Artificial intelligence and machine learning methods in predicting anti-cancer drug combination effects. *Briefings Bioinf* 2021;**22**:bbab271.
- Wang W, Ye ZYF, Gao HL, Ouyang DF. Computational pharmaceutics—A new paradigm of drug delivery. *J Control Release* 2021;**338**:119–36.
- Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. *Adv Drug Deliv Rev* 2021;**175**:113806.
- Hamilton S, Kingston BR. Applying artificial intelligence and computational modeling to nanomedicine. *Curr Opin Biotechnol* 2024;**85**:103043.
- Cern A, Golbraikh A, Sedykh A, Tropsha A, Barenholz Y, Goldblum A. Quantitative structure–property relationship modeling of remote liposome loading of drugs. *J Control Release* 2012;**160**:147–57.
- Kehrein J, Bunker A, Luxenhofer R. POxload: machine learning estimates drug loadings of polymeric micelles. *Mol Pharm* 2024;**21**:3356–74.
- Eliasson H, Erni R. Improving nanoparticle size estimation from scanning transmission electron micrographs with a multislice surrogate model. *Nano Lett* 2025;**25**:2474–9.
- Zelenka C, Kamp M, Strohm K, Kadoura A, Johnny J, Koch R, et al. Automated classification of nanoparticles with various ultrastructures and sizes via deep learning. *Ultramicroscopy* 2023;**246**:113685.
- Ilett M, Wills J, Rees P, Sharma S, Mickelthwaite S, Brown A, et al. Application of automated electron microscopy imaging and machine learning to characterise and quantify nanoparticle dispersion in aqueous media. *J Microsc* 2019;**279**:177–84.
- Wang XZ, Li J, Ha HD, Dahl JC, Ondry JC, Moreno-Hernandez I, et al. AutoDetect-mNP: an unsupervised machine learning algorithm for automated analysis of transmission electron microscope images of metal nanoparticles. *JACS Au* 2021;**1**:316–27.
- Sun ZJ, Shi J, Wang J, Jiang MQ, Wang Z, Bai XP, et al. A deep learning-based framework for automatic analysis of the nanoparticle morphology in SEM/TEM images. *Nanoscale* 2022;**14**:10761–72.
- Oktay AB, Gurses A. Automatic detection, localization and segmentation of nano-particles with deep learning in microscopy images. *Micron* 2019;**120**:113–9.
- Colliard-Granero A, Jitsev J, Eikerling MH, Malek K, Eslamibidgoli MJ. UTILE-Gen: automated image analysis in nanoscience using synthetic dataset generator and deep learning. *ACS Nanosci Au* 2023;**3**:10.
- Hussain MI, Rafique MA, Jung WG, Kim BJ, Jeon M. Segmentation and morphology computation of a spiky nanoparticle using the hourglass neural network. *ACS Omega* 2023;**8**:17834–40.
- Matalqah S, Lafi Z, Mhaidat Q, Asha N, Yousef Asha S. Applications of machine learning in liposomal formulation and development. *Pharmaceut Dev Technol* 2025;**30**:126–36.
- Dorsey PJ, Lau CL, Chang TC, Doerschuk PC, D'Addio SM. Review of machine learning for lipid nanoparticle formulation and process development. *J Pharm Sci* 2024;**113**:3413–33.
- Tan P, Chen XT, Zhang H, Wei Q, Luo K. Artificial intelligence aids in development of nanomedicines for cancer management. *Semin Cancer Biol* 2023;**89**:61–75.
- Tao HC, Wu TY, Aldeghi M, Wu TC, Aspuru-Guzik A, Kumacheva E. Nanoparticle synthesis assisted by machine learning. *Nat Rev Mater* 2021;**6**:701–16.
- Mozafari N, Mozafari N, Dehshahri A, Azadi A. Knowledge gaps in generating cell-based drug delivery systems and a possible meeting with artificial intelligence. *Mol Pharm* 2023;**20**:3757–78.
- Singh AV, Ansari MHD, Rosenkranz D, Maharjan RS, Krieger FL, Gandhi K, et al. Artificial intelligence and machine learning in computational nanotoxicology: unlocking and empowering nanomedicine. *Adv Healthcare Mater* 2020;**9**:1901862.
- Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci* 2021;**2**:160.
- Sarker IH, Kayes A, Badsha S, Alqahtani H, Watters P, Ng A. Cybersecurity data science: an overview from machine learning perspective. *J Big Data* 2020;**7**:1–29.

32. Dhoble S, Wu TH, Kenry. Decoding nanomaterial–biosystem interactions through machine learning. *Angew Chem Int Ed* 2024;**63**: e202318380.
33. Suriyaamporn P, Pamornpathomkul B, Patrojanasophon P, Ngawhirunpat T, Rojanarata T, Opanasopit P. The artificial intelligence-powered new era in pharmaceutical research and development: a review. *AAPS PharmSciTech* 2024;**25**:188.
34. Damiaty SA. Digital pharmaceutical sciences. *AAPS PharmSciTech* 2020;**21**:206.
35. He S, Leanse LG, Feng YF. Artificial intelligence and machine learning assisted drug delivery for effective treatment of infectious diseases. *Adv Drug Deliv Rev* 2021;**178**:113922.
36. Albadr MA, Tiun S, Ayob M, Al-Dhief F. Genetic algorithm based on natural selection theory for optimization problems. *Symmetry* 2020;**12**:1758.
37. Wang XL, Jin YC, Schmitt S, Olhofer M. Recent advances in Bayesian optimization. *ACM Comput Surv* 2023;**55**:1–36.
38. Idrees H, Zaidi SZJ, Sabir A, Khan RU, Zhang XL, Hassan SU. A review of biodegradable natural polymer-based nanoparticles for drug delivery applications. *Nanomaterials* 2020;**10**:1970.
39. Kamaly N, Yameen B, Wu J, Farokhzad OC. Degradable controlled-release polymers and polymeric nanoparticles: mechanisms of controlling drug release. *Chem Rev* 2016;**116**:2602–63.
40. Mir M, Ahmed N, Rehman AU. Recent applications of PLGA based nanostructures in drug delivery. *Colloids Surf, B* 2017;**159**:217–31.
41. Sarkar C, Kommineni N, Butreddy A, Kumar R, Bunekar N, Gugulothu K. PLGA nanoparticles in drug delivery. *Nano-engineering of Biomaterials* 2022:217–60.
42. Dwivedi P, Karumbaiah KM, Das R. Nano-size polymers via precipitation of polymer solutions. In: Fakirov S, editor. *Nano-size polymers preparation, properties. Applications*. Springer International Publishing; 2016. p. 251–82.
43. Seegobin N, Abdalla Y, Li G, Murdan S, Shorthouse D, Basit AW. Optimising the production of PLGA nanoparticles by combining design of experiment and machine learning. *Int J Pharm* 2024;**667**: 124905.
44. Huwaimel B, Alqarni S. Design of poly(lactic-co-glycolic acid) nanoparticles in drug delivery by artificial intelligence methods to find the conditions of nanoparticles synthesis. *Chemometr Intell Lab Syst* 2025;**258**:105335.
45. Almansour K, Alqahtani AS. Utilization of machine learning approach for production of optimized PLGA nanoparticles for drug delivery applications. *Sci Rep* 2025;**15**:8840.
46. Zabihi F, Xin N, Jia JF, Chen T, Zhao YP. High yield and high loading preparation of curcumin–PLGA nanoparticles using a modified supercritical antisolvent technique. *Ind Eng Chem Res* 2014;**53**:6569–74.
47. Yaghoobi N, Majidi RF, Faramarzi MA, Baharifar H, Amani A. Preparation, optimization and activity evaluation of PLGA/streptokinase nanoparticles using electrospray. *Adv Pharmaceut Bull* 2017;**7**:131.
48. Dong S, Yu HL, Poupart P, Ho EA. Gaussian processes modeling for the prediction of polymeric nanoparticle formulation design to enhance encapsulation efficiency and therapeutic efficacy. *Drug Delivery Transl Res* 2025;**15**:372–88.
49. Mihandoost S, Rezvantab S, M Pallares R, Schulz V, Kiessling F. A generative adversarial network approach to predict nanoparticle size in microfluidics. *ACS Biomater Sci Eng* 2025;**11**:268–79.
50. Rezvantab S, Mihandoost S, Rezaiee M. Machine learning assisted exploration of the influential parameters on the PLGA nanoparticles. *Sci Rep* 2024;**14**:1114.
51. Noorain L, Nguyen V, Kim HW, Nguyen LTB. A machine learning approach for PLGA nanoparticles in antiviral drug delivery. *Pharmaceutics* 2023;**15**:495.
52. Jara MO, Landin M, Morales JO. Screening of critical variables in fabricating polycaprolactone nanoparticles using neuro fuzzy logic. *Int J Pharm* 2021;**601**:120558.
53. Jara MO, Catalan-Figueroa J, Landin M, Morales JO. Finding key nanoprecipitation variables for achieving uniform polymeric nanoparticles using neurofuzzy logic technology. *Drug Delivery Transl Res* 2018;**8**:1797–806.
54. Kimmig J, Schuett T, Vollrath A, Zechel S, Schubert US. Prediction of nanoparticle sizes for arbitrary methacrylates using artificial neuronal networks. *Adv Sci* 2021;**8**:2102429.
55. Furtuna R, Curteanu S, Racles C. NSGA-II-RJG applied to multi-objective optimization of polymeric nanoparticles synthesis with silicone surfactants. *Open Chem* 2011;**9**:1080–95.
56. Shahsavari S, Shirmard LR, Amini M, Dokoosh FA. Application of artificial neural networks in the design and optimization of a nanoparticulate fingolimod delivery system based on biodegradable poly(3-hydroxybutyrate-co-3-hydroxyvalerate). *J Pharm Sci* 2017;**106**:176–82.
57. Youshia J, Ali ME, Lamprecht A. Artificial neural network based particle size prediction of polymeric nanoparticles. *Eur J Pharm Biopharm* 2017;**119**:333–42.
58. Shalaby KS, Soliman ME, Casettari L, Bonacucina G, Cespi M, Palmieri GF, et al. Determination of factors controlling the particle size and entrapment efficiency of nescapine in PEG/PLA nanoparticles using artificial neural networks. *Int J Nanomed* 2014;**9**: 4953–64.
59. Asadi H, Rostamizadeh K, Salari D, Hamidi M. Preparation of biodegradable nanoparticles of tri-block PLA–PEG–PLA copolymer and determination of factors controlling the particle size using artificial neural network. *J Microencapsulation* 2011;**28**:406–16.
60. Gong D, Ben-Akiva E, Singh A, Yamagata H, Est-Witte S, Shade JK, et al. Machine learning guided structure function predictions enable *in silico* nanoparticle screening for polymeric gene delivery. *Acta Biomater* 2022;**154**:349–58.
61. Potta T, Zhen Z, Grandhi TSP, Christensen MD, Ramos J, Breneman CM, et al. Discovery of antibiotics-derived polymers for gene delivery using combinatorial synthesis and cheminformatics modeling. *Biomaterials* 2014;**35**:1977–88.
62. Dalal RJ, Oviedo F, Leyden MC, Reineke TM. Polymer design via SHAP and Bayesian machine learning optimizes pDNA and CRISPR ribonucleoprotein delivery. *Chem Sci* 2024;**15**:7219–28.
63. Fazal T, Murtaza BN, Shah M, Iqbal S, Rehman MU, Jaber F, et al. Recent developments in natural biopolymer based drug delivery systems. *RSC Adv* 2023;**13**:23087–121.
64. Algharib SA, Dawood A, Zhou KX, Chen DM, Li C, Meng KY, et al. Preparation of chitosan nanoparticles by ionotropic gelation technique: effects of formulation parameters and *in vitro* characterization. *J Mol Struct* 2022;**1252**:132129.
65. Hashad RA, Ishak RA, Fahmy S, Mansour S, Geneidi AS. Chitosan-tripolyphosphate nanoparticles: optimization of formulation parameters for improving process yield at a novel pH using artificial neural networks. *Int J Biol Macromol* 2016;**86**:50–8.
66. Echeverri-Cuatas CE, Gartner C, Lapitsky Y. PEGylation and folate conjugation effects on the stability of chitosan-tripolyphosphate nanoparticles. *Int J Biol Macromol* 2020;**158**:1055–65.
67. Bozuyuk U, Dogan NO, Kizilel S. Deep insight into PEGylation of bioadhesive chitosan nanoparticles: sensitivity study for the key parameters through artificial neural network model. *ACS Appl Mater Interfaces* 2018;**10**:33945–55.
68. Baharifar H, Amani A. Size, loading efficiency, and cytotoxicity of albumin-loaded chitosan nanoparticles: an artificial neural networks study. *J Pharm Sci* 2017;**106**:411–7.
69. Esmailzadeh-Gharedaghi E, Faramarzi MA, Amini MA, Rouholamini NA, Rezayat SM, Amani A. Effects of processing parameters on particle size of ultrasound prepared chitosan nanoparticles: an artificial neural networks study. *Pharmaceut Dev Technol* 2012;**17**:638–47.
70. Wu HS, He JB, Cheng HR, Yang L, Park HJ, Li JL. Development and analysis of machine-learning guided flash nanoprecipitation (FNP) for continuous chitosan nanoparticles production. *Int J Biol Macromol* 2022;**222**:1229–37.
71. Zaki MR, Varshosaz J, Fathi M. Preparation of agar nanospheres: comparison of response surface and artificial neural network

- modeling by a genetic algorithm approach. *Carbohydr Polym* 2015; **122**:314–20.
72. Chaurawal N, Quadir SS, Joshi G, Barkat MA, Alanezi AA, Raza K. Development of fucoidan/polyethyleneimine based sorafenib-loaded self-assembled nanoparticles with machine learning and DoE-ANN implementation: optimization, characterization, and *in-vitro* assessment for the anticancer drug delivery. *Int J Biol Macromol* 2024; **279**: 135123.
 73. Koletti AE, Tsarouchi E, Kapourani A, Kontogiannopoulos KN, Assimopoulou AN, Barmplexis P. Gelatin nanoparticles for NSAID systemic administration: quality by design and artificial neural networks implementation. *Int J Pharm* 2020; **578**:119118.
 74. Smeraldo A, Ponsiglione AM, Netti PA, Torino E. Artificial neural network modelling hydrophobicity for optimal design by microfluidics of polymer nanoparticles to apply in magnetic resonance imaging. *Acta Biomater* 2023; **171**:440–50.
 75. Imani S, Li XY, Chen KY, Maghsoudloo M, Jabbarzadeh KP, Hashemi M, et al. Computational biology and artificial intelligence in mRNA vaccine design for cancer immunotherapy. *Front Cell Infect Microbiol* 2025; **14**:1501010.
 76. Xu LF, Li C, Liao R, Xiao Q, Wang XR, Zhao Z, et al. From sequence to system: enhancing IVT mRNA vaccine effectiveness through cutting-edge technologies. *Mol Pharm* 2024; **22**:81–102.
 77. Soroudi S, Jaafari MR, Arabi L. Lipid nanoparticle (LNP) mediated mRNA delivery in cardiovascular diseases: advances in genome editing and CAR T cell therapy. *J. Controlled Release* 2024; **372**: 113–40.
 78. Kulkarni JA, Cullis PR, Van Der Meel R. Lipid nanoparticles enabling gene therapies: from concepts to clinical utility. *Nucleic Acid Ther* 2018; **28**:146–57.
 79. Moosavi SG, Rahiman N, Jaafari MR, Arabi L. Lipid nanoparticle (LNP) mediated mRNA delivery in neurodegenerative diseases. *J. Controlled Release* 2025; **381**:113641.
 80. Hassett KJ, Benenato KE, Jacquinet E, Lee A, Woods A, Yuzhakov O, et al. Optimization of lipid nanoparticles for intramuscular administration of mRNA vaccines. *Mol Ther Nucleic acids* 2019; **15**:1–11.
 81. Wang W, Chen KP, Jiang T, Wu YY, Wu Z, Ying H, et al. Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery. *Nat Commun* 2024; **15**:10804.
 82. Zhang YB, Sun CZ, Wang C, Jankovic KE, Dong YZ. Lipids and lipid derivatives for RNA delivery. *Chem Rev* 2021; **121**:12181–277.
 83. Li BW, Raji IO, Gordon ARG, Sun LZ, Raimondo TM, Oladimeji FA, et al. Accelerating ionizable lipid discovery for mRNA delivery using machine learning and combinatorial chemistry. *Nat Mater* 2024; **23**:1002–8.
 84. Witten J, Raji I, Manan RS, Beyer E, Bartlett S, Tang YH, et al. Artificial intelligence-guided design of lipid nanoparticles for pulmonary gene therapy. *Nat Biotechnol* 2024; **1**–10.
 85. Xu Y, Ma SH, Cui HT, Chen JA, Xu SF, Gong FL, et al. AGILE platform: a deep learning powered approach to accelerate LNP development for mRNA delivery. *Nat Commun* 2024; **15**:6305.
 86. Wu K, Yang XL, Wang ZX, Li N, Zhang JL, Liu LZ. Data-balanced transformer for accelerated ionizable lipid nanoparticles screening in mRNA delivery. *Briefings Bioinf* 2024; **25**:bbae186.
 87. Bae SH, Choi H, Lee J, Kang MH, Ahn SH, Lee YS, et al. Rational design of lipid nanoparticles for enhanced mRNA vaccine delivery via machine learning. *Small* 2025; **21**:2405618.
 88. Cheng L, Zhu YN, Ma JY, Aggarwal A, Toh WH, Shin C, et al. Machine learning elucidates design features of plasmid deoxyribonucleic acid lipid nanoparticles for cell type-preferential transfection. *ACS Nano* 2024; **18**:28735–47.
 89. Rafiei M, Shojaei A, Chau Y. Machine learning-assisted design of immunomodulatory lipid nanoparticles for delivery of mRNA to repolarize hyperactivated microglia. *Drug Deliv* 2025; **32**:2465909.
 90. Harrison PJ, Wieslander H, Sabirsh A, Karlsson J, Malmjö V, Hellander A, et al. Deep-learning models for lipid nanoparticle-based drug delivery. *Nanomed* 2021; **16**:1097–110.
 91. Wang W, Feng S, Ye ZYF, Gao HL, Lin JZ, Ouyang DF. Prediction of lipid nanoparticles for mRNA vaccines by the machine learning algorithm. *Acta Pharm Sin B* 2022; **12**:2950–62.
 92. Gao HS, Kan S, Ye ZYF, Feng YC, Jin L, Zhang XD, et al. Development of *in silico* methodology for siRNA lipid nanoparticle formulations. *Chem Eng J* 2022; **442**:136310.
 93. Sato S, Sano S, Muto H, Kubara K, Kondo K, Miyazaki T, et al. Understanding the manufacturing process of lipid nanoparticles for mRNA delivery using machine learning. *Chem Pharm Bull* 2024; **72**: 529–39.
 94. Abostait A, Abdelkarim M, Bao ZQ, Miyake Y, Tse WH, Di Ciano-Oliveir C, et al. Optimizing lipid nanoparticles for fetal gene delivery *in vitro*, *ex vivo*, and aided with machine learning. *J. Controlled Release* 2024; **376**:678–700.
 95. Maharjan R, Hada S, Lee JE, Han HK, Kim KH, Seo HJ, et al. Comparative study of lipid nanoparticle-based mRNA vaccine bioprocess with machine learning and combinatorial artificial neural network-design of experiment approach. *Int J Pharm* 2023; **640**: 123012.
 96. Maharjan R, Kim KH, Lee K, Han HK, Jeong SH. Machine learning-driven optimization of mRNA-lipid nanoparticle vaccine quality with XGBoost/Bayesian method and ensemble model approaches. *J Pharm Anal* 2024; **14**:100996.
 97. Abdulbaqi IM, Assi RA, Yaghamur A, Darwis Y, Mohtar N, Parumasivam T, et al. Pulmonary delivery of anticancer drugs via lipid-based nanocarriers for the treatment of lung cancer: an update. *Pharmaceuticals* 2021; **14**:725.
 98. Zahid AA, Chakraborty A, Luo W, Coyle A, Paul A. Tailoring the inherent properties of biobased nanoparticles for nanomedicine. *ACS Biomater Sci Eng* 2023; **9**:3972–86.
 99. Peng TX, Xu WH, Li QQ, Ding Y, Huang YZ. Pharmaceutical liposomal delivery—specific considerations of innovation and challenges. *Biomater Sci* 2023; **11**:62–75.
 100. Cardoso-Daodu IM, Ilomuanya MO, Amenaghawon AN, Azubuike CP. Artificial neural network for optimizing the formulation of curcumin-loaded liposomes from statistically designed experiments. *Prog Biomater* 2022; **11**:55–65.
 101. Hathout RM, Gad HA, Metwally AA. Gelatinized-core liposomes: toward a more robust carrier for hydrophilic molecules. *J Biomed Mater Res, Part A* 2017; **105**:3086–92.
 102. Rebollo R, Oyouf F, Corvis Y, El-Hammadi MM, Saubamea B, Andrieux K, et al. Microfluidic manufacturing of liposomes: development and optimization by design of experiment and machine learning. *ACS Appl Mater Interfaces* 2022; **14**:39736–45.
 103. Huang SM, Kuo CH, Chen CA, Liu YC, Shieh CJ. RSM and ANN modeling-based optimization approach for the development of ultrasound-assisted liposome encapsulation of piceid. *Ultrason Sonochem* 2017; **36**:112–22.
 104. Zhao F, Lu J, Jin X, Wang Z, Sun YH, Gao DD, et al. Comparison of response surface methodology and artificial neural network to optimize novel ophthalmic flexible nano-liposomes: characterization, evaluation, *in vivo* pharmacokinetics and molecular dynamics simulation. *Colloids Surf, B* 2018; **172**:288–97.
 105. León Blanco JM, González-R PL, Arroyo García CM, Cózar-Bernal MJ, Calle Suárez M, Canca Ortiz D, et al. Artificial neural networks as alternative tool for minimizing error predictions in manufacturing ultradeformable nanoliposome formulations. *Drug Dev Ind Pharm* 2018; **44**:135–43.
 106. Jaafar-Maalej C, Diab R, Andrieu V, Elaissari A, Fessi H. Ethanol injection method for hydrophilic and lipophilic drug-loaded liposome preparation. *J Liposome Res* 2010; **20**:228–43.
 107. Katritzky AR, Kuanar M, Slavov S, Hall CD, Karelson M, Kahn I, et al. Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction. *Chem Rev* 2010; **110**:5714–89.
 108. Sansare S, Duran T, Mohammadiarani H, Goyal M, Yenduri G, Costa A, et al. Artificial neural networks in tandem with molecular descriptors as predictive tools for continuous liposome manufacturing. *Int J Pharm* 2021; **603**:120713.

109. Di Francesco V, Boso DP, Moore TL, Schrefler BA, Decuzzi P. Machine learning instructed microfluidic synthesis of curcumin-loaded liposomes. *Biomed Microdevices* 2023;**25**:29.
110. Hoseini B, Jaafari MR, Golabpour A, Momtazi-Borojeni AA, Karimi M, Eslami S. Application of ensemble machine learning approach to assess the factors affecting size and polydispersity index of liposomal nanoparticles. *Sci Rep* 2023;**13**:18012.
111. Hoseini B, Jaafari MR, Golabpour A, Momtazi-Borojeni AA, Eslami S. Optimizing nanoliposomal formulations: assessing factors affecting entrapment efficiency of curcumin-loaded liposomes using machine learning. *Int J Pharm* 2023;**646**:12.
112. Eugster R, Orsi M, Buttitta G, Serafini N, Tiboni M, Casettari L, et al. Leveraging machine learning to streamline the development of liposomal drug delivery systems. *J Control Release* 2024;**376**: 1025–38.
113. Lindsay S, Tumolva O, Khamiakova T, Coppenolle H, Kovarik M, Shah S, et al. Can we simplify liposome manufacturing using a complex DoE approach?. *Pharmaceutics* 2024;**16**:1159.
114. Han R, Ye ZYF, Zhang YS, Cheng YX, Zheng Y, Ouyang DF. Predicting liposome formulations by the integrated machine learning and molecular modeling approaches. *Asian J Pharm Sci* 2023;**18**:100811.
115. Dong J, Wu Z, Xu HL, Ouyang DF. FormulationAI: a novel web-based platform for drug formulation design driven by artificial intelligence. *Briefings Bioinf* 2024;**25**:bbad419.
116. Basso J, Mendes M, Silva J, Cova T, Luque-Michel E, Jorge AF, et al. Sorting hidden patterns in nanoparticle performance for glioblastoma using machine learning algorithms. *Int J Pharm* 2021;**592**:120095.
117. Pink DL, Loruthai O, Ziolk RM, Terry AE, Barlow DJ, Lawrence MJ, et al. Interplay of lipid and surfactant: impact on nanoparticle structure. *J Colloid Interface Sci* 2021;**597**:278–88.
118. Hathout RM, Metwally AA. Towards better modelling of drug-loading in solid lipid nanoparticles: molecular dynamics, docking experiments and Gaussian processes machine learning. *Eur J Pharm Biopharm* 2016;**108**:262–8.
119. Ghorbannejad Nashli F, Aghajpour S, Farmoudeh A, Balef SSH, Torkamanian M, Razavi A, et al. Preparation and optimisation of solid lipid nanoparticles of rivaroxaban using artificial neural networks and response surface method. *J Microencapsulation* 2025;**42**: 70–82.
120. Öztürk AA, Gündüz AB, Ozisik O. Supervised machine learning algorithms for evaluation of solid lipid nanoparticles and particle size. *Comb Chem High Throughput Screening* 2018;**21**:693–9.
121. Vishwakarma N, Jain A, Sharma R, Mody N, Vyas S, Vyas SP. Lipid-based nanocarriers for lymphatic transportation. *AAPS PharmSciTech* 2019;**20**:83.
122. Mishra DK, Shandilya R, Mishra PK. Lipid based nanocarriers: a translational perspective. *Nanomedicine* 2018;**14**:2023–50.
123. Elkomy MH, Elmowafy M, Shalaby K, Azmy AF, Ahmad N, Zafar A, et al. Development and machine-learning optimization of mucoadhesive nanostructured lipid carriers loaded with fluconazole for treatment of oral candidiasis. *Drug Dev Ind Pharm* 2021;**47**: 246–58.
124. Rouco H, Diaz-Rodriguez P, Rama-Molinos S, Remuñán-López C, Landin M. Delimiting the knowledge space and the design space of nanostructured lipid carriers through Artificial Intelligence tools. *Int J Pharm* 2018;**553**:522–30.
125. Martinez-Borrajó R, Diaz-Rodriguez P, Landin M. Engineering mannose-functionalized nanostructured lipid carriers by sequential design using hybrid artificial intelligence tools. *Drug Delivery Transl Res* 2025;**15**:343–54.
126. Amasya G, Badilli U, Aksu B, Tarimci N. Quality by design case study 1: design of 5-fluorouracil loaded lipid nanoparticles by the W/O/W double emulsion—Solvent evaporation method. *Eur J Pharm Sci* 2016;**84**:92–102.
127. Dawoud MHS, Fayeze AM, Mohamed RA, Sweed NM. Optimization of nanovesicular carriers of a poorly soluble drug using factorial design methodology and artificial neural network by applying quality by design approach. *Pharmaceut Dev Technol* 2021;**26**:1035–50.
128. Bao ZQ, Yung F, Hickman RJ, Aspuru-Guzik A, Bannigan P, Allen C. Data-driven development of an oral lipid-based nanoparticle formulation of a hydrophobic drug. *Drug Delivery Transl Res* 2024;**14**:1872–87.
129. Kumar H, Kumar V. Ultrasound assisted synthesis of water-in-oil nanoemulsions: parametric optimization using hybrid ANN-GA approach. *Chem Eng Process* 2019;**144**:107649.
130. Samiun WS, Basri M, Masoumi HRF, Khairudin N. The prediction of the optimum compositions of a parenteral nanoemulsion system loaded with a low water solubility drug for the treatment of schizophrenia by artificial neural networks. *RSC Adv* 2016;**6**: 14068–76.
131. Abdulaal WH, Hosny KM, Alhakamy NA, Bakhaidar RB, Almuhananna Y, Sabei FY, et al. Fabrication, assessment, and optimization of alendronate sodium nanoemulsion-based injectable *in situ* gel formulation for management of osteoporosis. *Drug Deliv* 2023;**30**:2164094.
132. Kumar M, Bishnoi RS, Shukla AK, Jain CP. Techniques for formulation of nanoemulsion drug delivery system: a review. *Prev Nutr Food Sci* 2019;**24**:225–34.
133. Shi YB, Li HL, Li JC, Zhi DJ, Zhang XY, Liu H, et al. Development, optimization and evaluation of emodin loaded nanoemulsion prepared by ultrasonic emulsification. *J Drug Delivery Sci Technol* 2015;**27**:46–55.
134. Khani S, Abbasi S, Keyhanfar F, Amani A. Use of artificial neural networks for analysis of the factors affecting particle size in mebupidine nanoemulsion. *J Biomol Struct Dyn* 2019;**37**:3162–7.
135. Daoud GS, Basri H, Stanslas J, Masoumi HRF, Basri M. Predicting the optimum compositions of a parenteral nanoemulsion system loaded with azithromycin antibiotic utilizing the artificial neural network model. *RSC Adv* 2015;**5**:82654–65.
136. Samson S, Basri M, Fard Masoumi HR, Abdul Malek E, Abedi Karjiban R. An artificial neural network based analysis of factors controlling particle size in a virgin coconut oil-based nanoemulsion system containing copper peptide. *PLoS One* 2016;**11**:e0157737.
137. Ngan CL, Masoumi HRF, Basri M, Rahman MBA. Development of nano-colloidal system for fullerene by ultrasonic-assisted emulsification techniques based on artificial neural network. *Arabian J Chem* 2019;**12**:4162–70.
138. Seyedhassantehrani N, Karimi R, Tavoosidana G, Amani A. Concurrent study of stability and cytotoxicity of a novel nanoemulsion system—An artificial neural networks approach. *Pharmaceut Dev Technol* 2017;**22**:383–9.
139. Ahmadi Lakalayeh G, Faridi-Majidi R, Saber R, Partoazar A, Mehr SE, Amani A. Investigating the parameters affecting the stability of superparamagnetic iron oxide-loaded nanoemulsion using artificial neural networks. *AAPS PharmSciTech* 2012;**13**:1386–95.
140. Yadollahi R, Vasilev K, Simovic S. Nanosuspension technologies for delivery of poorly soluble drugs. *J Nanomater* 2015;**2015**:216375.
141. Jadhav SP, Singh SK, Chawra HS. Review on nanosuspension as a novel method for solubility and bioavailability enhancement of poorly soluble drugs. *Adv Pharmacol Pharm* 2023;**11**:117–30.
142. Bhakay A, Rahman M, Dave RN, Bilgili E. Bioavailability enhancement of poorly water-soluble drugs via nanocomposites: formulation—Processing aspects and challenges. *Pharmaceutics* 2018;**10**:86.
143. Guner G, Yilmaz D, Yao HF, Clancy DJ, Bilgili E. Predicting the temperature evolution during nanomilling of drug suspensions via a semi-theoretical lumped-parameter model. *Pharmaceutics* 2022;**14**: 2840.
144. Guner G, Elashri S, Mehaj M, Seetharaman N, Yao HF, Clancy DJ, et al. An enthalpy-balance model for timewise evolution of temperature during wet stirred media milling of drug suspensions. *Pharm Res* 2022;**39**:2065–82.
145. Mahesh KV, Singh SK, Gulati M. A comparative study of top-down and bottom-up approaches for the preparation of nanosuspensions of glipizide. *Powder Technol* 2014;**256**:436–49.
146. Abdelbary AA, Al-Mahallawi AM, Abdelrahim ME, Ali AM. Preparation, optimization, and *in vitro* simulated inhalation delivery

- of carvedilol nanoparticles loaded on a coarse carrier intended for pulmonary administration. *Int J Nanomed* 2015;**10**:6339–53.
147. Aghajani M, Shahverdi AR, Rezayat SM, Amini MA, Amani A. Preparation and optimization of acetaminophen nanosuspension through nanoprecipitation using microfluidic devices: an artificial neural networks study. *Pharmaceut Dev Technol* 2013;**18**:609–18.
 148. Aghajani M, Shahverdi AR, Amani A. The use of artificial neural networks for optimizing polydispersity index (PDI) in nanoprecipitation process of acetaminophen in microfluidic devices. *AAPS PharmSciTech* 2012;**13**:1293–301.
 149. Aghajani MH, Pashazadeh AM, Mostafavi SH, Abbasi S, Hajibagheri-Fard M-J, Assadi M, et al. Size control in the nanoprecipitation process of stable iodine (127 I) using microchannel reactor—Optimization by artificial neural networks. *AAPS PharmSciTech* 2015;**16**:1059–68.
 150. Davidopoulou C, Ouranidis A. Pharma 4.0-artificially intelligent digital twins for solidified nanosuspensions. *Pharmaceutics* 2022;**14**:2113.
 151. Buse J, El-Aneed A. Properties, engineering and applications of lipid-based nanoparticle drug-delivery systems: current research and advances. *Nanomed* 2010;**5**:1237–60.
 152. Bayat F, Dadashzadeh S, Aboofazeli R, Torshabi M, Baghi AH, Tamiji Z, et al. Oral delivery of posaconazole-loaded phospholipid-based nanoformulation: preparation and optimization using design of experiments, machine learning, and TOPSIS. *Int J Pharm* 2024;**653**:123879.
 153. Li YQ, Abbaspour MR, Grootendorst PV, Rauth AM, Wu XY. Optimization of controlled release nanoparticle formulation of verapamil hydrochloride using artificial neural networks with genetic algorithm and response surface methodology. *Eur J Pharm Biopharm* 2015;**94**:170–9.
 154. Henser-Brownhill T, Martin L, Samangouei P, Ladak A, Apostolidou M, Nagel B, et al. *In silico* screening accelerates nano-carrier design for efficient mRNA delivery. *Adv Sci* 2024;**11**:2401935.
 155. Yamankurt G, Berns EJ, Xue A, Lee A, Bagheri N, Mrksich M, et al. Exploration of the nanomedicine-design space with high-throughput screening and machine learning. *Nat Biomed Eng* 2019;**3**:318–27.
 156. Uttreja P, Karnik I, Adel Ali Youssef A, Narala N, Elkanayati RM, Baisa S, et al. Self-emulsifying drug delivery systems (SEDDS): transition from liquid to solid—A comprehensive review of formulation, characterization, applications, and future trends. *Pharmaceutics* 2025;**17**:63.
 157. Akula S, Gurram AK, Devireddy SR. Self-microemulsifying drug delivery systems: an attractive strategy for enhanced therapeutic profile. *Int Scholarly Res Not* 2014;**2014**:964051.
 158. Kyatanwar AU, Jadhav KR, Kadam VJ. Self micro-emulsifying drug delivery system (SMEDDS). *J Pharm Res* 2010;**3**:75–83.
 159. Cherniakov I, Domb AJ, Hoffman A. Self-nano-emulsifying drug delivery systems: an update of the biopharmaceutical aspects. *Expet Opin Drug Deliv* 2015;**12**:1121–33.
 160. Gao HS, Jia HY, Dong J, Yang XG, Li HF, Ouyang DF. Integrated *in silico* formulation design of self-emulsifying drug delivery systems. *Acta Pharm Sin B* 2021;**11**:3585–94.
 161. Li CW, Yang SY, He R, Tao WJ, Yin ZN. Development of quantitative structure-property relationship models for self-emulsifying drug delivery system of 2-aryl propionic acid NSAIDs. *J Nanomater* 2011;**2011**:206320.
 162. Parikh KJ, Sawant KK. Comparative study for optimization of pharmaceutical self-emulsifying pre-concentrate by design of experiment and artificial neural network. *AAPS PharmSciTech* 2018;**19**:3311–21.
 163. Yasamineh S, Yasamineh P, Kalajahi HG, Gholizadeh O, Yekanipour Z, Afkhami H, et al. A state-of-the-art review on the recent advances of niosomes as a targeted drug delivery system. *Int J Pharm* 2022;**624**:121878.
 164. Gao SQ, Sui Z, Jiang Q, Jiang YY. Functional evaluation of niosomes utilizing surfactants in nanomedicine applications. *Int J Nanomed* 2024;**19**:10283–305.
 165. Chen S, Hanning S, Falconer J, Locke M, Wen JY. Recent advances in non-ionic surfactant vesicles (niosomes): fabrication, characterization, pharmaceutical and cosmetic applications. *Eur J Pharm Biopharm* 2019;**144**:18–39.
 166. Kashani-Asadi-Jafari F, Aftab A, Ghaemmaghami S. A machine learning framework for predicting entrapment efficiency in niosomal particles. *Int J Pharm* 2022;**627**:122203.
 167. Fahmy OM, Eissa RA, Mohamed HH, Eissa NG, Elsabahy M. Machine learning algorithms for prediction of entrapment efficiency in nanomaterials. *Methods* 2023;**218**:133–40.
 168. Shahiwal A, Qawoogha SS, Faruqi N. Designing optimum drug delivery systems using machine learning approaches: a prototype study of niosomes. *AAPS PharmSciTech* 2023;**24**:94.
 169. Al-Kassas R, Bansal M, Shaw J. Nanosizing techniques for improving bioavailability of drugs. *J Control Release* 2017;**260**:202–12.
 170. Rathod PS, Narkhede MR, Dongare SL. A recent review on nanocrystal manufacturing techniques with pharmaceutical application. *Curr Nanomed* 2024;**14**:4–12.
 171. Müller RH, Gohla S, Keck CM. State of the art of nanocrystals—special features, production, nanotoxicology aspects and intracellular delivery. *Eur J Pharm Biopharm* 2011;**78**:1–9.
 172. He YH, Ye ZYF, Liu XY, Wei ZJ, Qiu F, Li HF, et al. Can machine learning predict drug nanocrystals?. *J Control Release* 2020;**322**:274–85.
 173. Rama B, Ribeiro AJ. Role of nanotechnology in the prolonged release of drugs by the subcutaneous route. *Expet Opin Drug Deliv* 2023;**20**:559–77.
 174. Chen WL, Palazzo A, Hennink WE, Kok RJ. Effect of particle size on drug loading and release kinetics of gefitinib-loaded PLGA microspheres. *Mol Pharm* 2017;**14**:459–67.
 175. Meng XP, Li XL. Size limit and energy analysis of nanoparticles during wrapping process by membrane. *Nanomaterials* 2018;**8**:899.
 176. Gao XL, Dong JJ, Zhang XR. The effect of nanoparticle size on endocytosis dynamics depends on membrane—nanoparticle interaction. *Mol Simul* 2015;**41**:531–7.
 177. Gupta P, Rai N, Verma A, Gautam V. Microscopy based methods for characterization, drug delivery, and understanding the dynamics of nanoparticles. *Med Res Rev* 2024;**44**:138–68.
 178. Yue J, Feliciano TJ, Li WL, Lee A, Odom TW. Gold nanoparticle size and shape effects on cellular uptake and intracellular distribution of siRNA nanoconstructs. *Bioconjug Chem* 2017;**28**:1791–800.
 179. Chen C, Yaari Z, Apfelbaum E, Grodzinski P, Shamay Y, Heller DA. Merging data curation and machine learning to improve nanomedicines. *Adv Drug Deliv Rev* 2022;**183**:114172.
 180. Honary S, Zahir F. Effect of zeta potential on the properties of nano-drug delivery systems—A review (part 1). *Trop J Pharmaceut Res* 2013;**12**:255–64.
 181. Palanikumar L, Choi ES, Oh JY, Park SA, Choi H, Kim K, et al. Importance of encapsulation stability of nanocarriers with high drug loading capacity for increasing *in vivo* therapeutic efficacy. *Biomacromolecules* 2018;**19**:3030–9.
 182. Cern A, Barenholz Y, Tropsha A, Goldblum A. Computer-aided design of liposomal drugs: *in silico* prediction and experimental validation of drug candidates for liposomal remote loading. *J Control Release* 2014;**173**:125–31.
 183. Wu WS, Zhang CY, Lin WJ, Chen Q, Guo XD, Qian Y, et al. Quantitative structure—property relationship (QSPR) modeling of drug-loaded polymeric micelles via genetic function approximation. *PLoS One* 2015;**10**:e0119575.
 184. Alves VM, Hwang D, Muratov E, Sokolsky-Papkov M, Varlamova E, Vinod N, et al. Cheminformatics-driven discovery of polymeric micelle formulations for poorly soluble drugs. *Sci Adv* 2019;**5**:eaav9784.
 185. Grassi M, Lamberti G, Cascone S, Grassi G. Mathematical modeling of simultaneous drug release and *in vivo* absorption. *Int J Pharm* 2011;**418**:130–41.
 186. Solomon D, Gupta N, Mulla NS, Shukla S, Guerrero YA, Gupta V. Role of *in vitro* release methods in liposomal formulation

- development: challenges and regulatory perspective. *AAPS J* 2017; **19**:1669–81.
187. Askarizadeh M, Esfandiari N, Honarvar B, Sajadian SA, Azdarpour A. Kinetic modeling to explain the release of medicine from drug delivery systems. *ChemBioEng Rev* 2023; **10**:1006–49.
 188. Aghajani S, Amiria H, Esfandiyari-Manesh M, Ebrahimnejad P, Jeelani H, Henschel A, et al. Utilizing machine learning for predicting drug release from polymeric drug delivery systems. *Comput Biol Med* 2025; **188**:109756.
 189. Siafaka PI, Barmbalexis P, Bikiaris DN. Novel electrospun nanofibrous matrices prepared from poly(lactic acid)/poly(butylene adipate) blends for controlled release formulations of an anti-rheumatoid agent. *Eur J Pharm Sci* 2016; **88**:12–25.
 190. Baghaei B, Saeb MR, Jafari SH, Khonakdar HA, Rezaee B, Goodarzi V, et al. Modeling and closed-loop control of particle size and initial burst of PLGA biodegradable nanoparticles for targeted drug delivery. *J Appl Polym Sci* 2017; **134**:45145.
 191. Sun Y, Qin SH, Li YL, Hasan N, Li YV, Liu JG. Machine learning integrated with *in vitro* experiments for study of drug release from PLGA nanoparticles. *Sci Rep* 2025; **15**:4218.
 192. Bannigan P, Bao ZQ, Hickman RJ, Aldeghi M, Hse F, Aspuru-Guzik A, et al. Machine learning models to accelerate the design of polymeric long-acting injectables. *Nat Commun* 2023; **14**:35.
 193. Stiepel RT, Pena ES, Ehrenzeller SA, Gallovic MD, Lifshits LM, Genito CJ, et al. A predictive mechanistic model of drug release from surface eroding polymeric nanoparticles. *J Control Release* 2022; **351**:883–95.
 194. Alshahrani SM, Alotaibi HF, Begum MY. Computational analysis of controlled drug release from porous polymeric carrier with the aid of mass transfer and artificial intelligence modeling. *Sci Rep* 2024; **14**:28422.
 195. Alshahrani SM, Alotaibi HF, Alqarni M. Modeling and validation of drug release kinetics using hybrid method for prediction of drug efficiency and novel formulations. *Front Chem* 2024; **12**:1395359.
 196. Alqarni M, Harthi SMA, Alzubaidi MA, Alqarni AA, Shukr BS, Shawli HT. Model development using hybrid method for prediction of drug release from biomaterial matrix. *Chemometr Intell Lab Syst* 2024; **253**:9.
 197. Pramanik A, Garg S. Prediction of the partition coefficients using QSPR modeling and simulation of paclitaxel release from the diffusion-controlled drug delivery devices. *Drug Delivery Transl Res* 2018; **8**:1300–12.
 198. Hussein GA, Sabouni R, Puzyrev V, Ghommam M. Deep learning for the accurate prediction of triggered drug delivery. *IEEE Trans NanoBioscience* 2025; **24**:102–12.
 199. Movahedi F, Ebrahimi Shahmabadi H, Alavi SE, Koohi Moftakhari Esfahani M. Release modeling and comparison of nano-archaeosomal, nanoliposomal and pegylated nanoliposomal carriers for paclitaxel. *Tumor Biol* 2014; **35**:8665–72.
 200. Moussa HG, Hussein GA, Abel-Jabbar N, Ahmad SE. Use of model predictive control and artificial neural networks to optimize the ultrasonic release of a model drug from liposomes. *IEEE Trans NanoBioscience* 2017; **16**:149–56.
 201. Shomope I, Percival KM, Abdel Jabbar NM, Hussein GA. Predicting calcein release from ultrasound-targeted liposomes: a comparative analysis of random forest and support vector machine. *Technol Cancer Res Treat* 2024; **23**:15330338241296725.
 202. Gazzi R, Gelli R, Eugster R, Bühr C, Schürch S, Mezzenga R, et al. Transitioning the production of lipidic mesophase-based delivery systems from lab-scale to robust industrial manufacturing following a risk-based quality by design approach augmented by artificial intelligence. *J Colloid Interface Sci* 2025; **678**:595–607.
 203. Dawoud MHS, Mannaa IS, Abdel-Daim A, Sweed NM. Integrating artificial intelligence with quality by design in the formulation of lecithin/chitosan nanoparticles of a poorly water-soluble drug. *AAPS PharmSciTech* 2023; **24**:169.
 204. Davoudizadeh T, Sajjadi SM, Ma'mani L. Exhaustive investigation of drug delivery systems to achieve optimal condition of drug release using non-linear generalized artificial neural network method: feedback from the loading step of drug. *J Iran Chem Soc* 2018; **15**:1999–2006.
 205. Seyedhassantehrani N, Karimi R, Tavoosidana G, Amani A. Concurrent study of stability and cytotoxicity of a novel nanoemulsion system—An artificial neural networks approach. *Pharmaceut Dev Technol* 2017; **22**:1–7.
 206. Cern A, Marcus D, Tropsha A, Barenholz Y, Goldblum A. New drug candidates for liposomal delivery identified by computer modeling of liposomes' remote loading and leakage. *J Control Release* 2017; **252**:18–27.
 207. Xu PC, Ji XB, Li MJ, Lu WC. Small data machine learning in materials science. *npj Comput Mater* 2023; **9**:42.
 208. Gupta NV, Murugan MD, Shanbhag SJ. AI/ML-driven nanocarriers for the management of neurodegeneration. In: Koduru TS, Osmani RAM, Singh E, Dutta S, editors. *The neurodegeneration revolution*. Elsevier; 2025. p. 361–72.
 209. Kiseleva A, Kotzinos D, De Hert P. Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations. *Front Artif Intell* 2022; **5**:879603.
 210. Floridi L, Cows J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds Mach* 2018; **28**:689–707.
 211. 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.
 212. Jena GK, Patra CN, Jammula S, Rana R, Chand S. Artificial intelligence and machine learning implemented drug delivery systems: a paradigm shift in the pharmaceutical industry. *Journal of Bio-X Research* 2024; **7**:16.
 213. Rosenblum D, Peer D. Omics-based nanomedicine: the future of personalized oncology. *Cancer Lett* 2014; **352**:126–36.
 214. He JB, Fan XT. Evaluating the performance of the K-fold cross-validation approach for model selection in growth mixture modeling. *Struct Equ Model* 2019; **26**:66–79.
 215. Zhang J, Wang S. A fast leave-one-out cross-validation for SVM-like family. *Neural Comput Appl* 2016; **27**:1717–30.
 216. Ambreen S, Umar M, Noor A, Jain H, Ali R. Advanced AI and ML frameworks for transforming drug discovery and optimization: with innovative insights in polypharmacology, drug repurposing, combination therapy and nanomedicine. *Eur J Med Chem* 2025; **284**:117164.
 217. Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M. Explainable AI for retinoblastoma diagnosis: interpreting deep learning models with LIME and SHAP. *Diagnostics* 2023; **13**:1932.
 218. Gholap AD, Uddin MJ, Faiyazuddin M, Omri A, Gowri S, Khalid M. Advances in artificial intelligence in drug delivery and development: a comprehensive review. *Comput Biol Med* 2024; **178**:108702.
 219. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* 2020; **10**:12598.
 220. Gangwal A, Ansari A, Ahmad I, Azad AK, Sulaiman WMAW. Current strategies to address data scarcity in artificial intelligence-based drug discovery: a comprehensive review. *Comput Biol Med* 2024; **179**:108734.
 221. Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility?. *Front Surg* 2022; **9**:862322.