Advanced Drug Delivery Reviews 202 (2023) 115108

Contents lists available at ScienceDirect

Advanced Drug Delivery Reviews

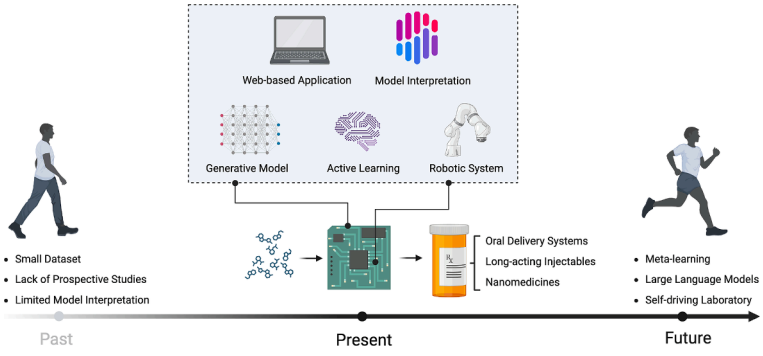
journal homepage: www.elsevier.com/locate/adr

Revolutionizing drug formulation development: The increasing impact of machine learning

Zeqing Bao [a], Jack Bufton [a], Riley J. Hickman [b,c,d], Alán Aspuru-Guzik [b,c,d,e,f,g,h,i], Pauric Bannigan [a,*,1], Christine Allen [a,f,i,*,1]

[a] Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON M5S 3M2, Canada
[b] Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada
[c] Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada
[d] Vector Institute for Artificial Intelligence, Toronto, ON M5S 1M1, Canada
[e] Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), Toronto, ON M5S 1M1, Canada
[f] Department of Chemical Engineering & Applied Chemistry, University of Toronto, Toronto, ON M5S 3E5, Canada
[g] Department of Materials Science & Engineering, University of Toronto, Toronto, ON M5S 3E4, Canada
[h] CIFAR Artificial Intelligence Research Chair, Vector Institute, Toronto, ON M5S 1M1, Canada
[i] Acceleration Consortium, Toronto, ON M5S 3H6, Canada

GRAPHICAL ABSTRACT



Abbreviations: ML, Machine learning; ASD, Amorphous solid dispersion; SEDDS, Self-emulsifying drug delivery system; RF, Random forest; ACC, Accuracy of classification models (reflected by the fraction of correct predictions); RMSE, Root mean squared error; MAE, Mean absolute error; MAPE, Mean absolute percentage error; SVR, Support vector regressor; NN, Neural network; PLS, Partial least squares; SMOTE, Synthetic minority oversampling technique; ENN, Edited nearest neighbour; MLR, Multi-linear regression; kNN, k-Nearest neighbors; SVM, Support vector machine; LAI, Long-acting injectable; LightGBM, Light gradient-boosted machine; SHAP, SHapley Additive exPlanations; PLGA, Poly(lactide-*co*-glycolide); MP, Microparticle; KRR, Kernel ridge regression; DT, Decision tree; LNP, Lipid nanoparticle; EPR, Enhanced permeability and retention; GAN, Generative adversarial network; QD, Quantum dot; OOD, Out-of-distribution; AL, Active learning; BO, Bayesian optimization; GP, Gaussian processes; SDL, Self-driving laboratory; MAP, Materials acceleration platform; PPH, Polymer-protein hybrid; GOx, Glucose oxidase; Lip, Lipase; HRP, Horseradish peroxidase; EA, Enzymatic activity; PET-RAFT, Photoinduced electron/energy transfer-reversible addition-fragmentation chain transfer; LLM, Large language model; AutoML, automated machine learning.

* Corresponding authors at: Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON M5S 3M2, Canada.
E-mail addresses: pauric.bannigan@utoronto.ca (P. Bannigan), cj.allen@utoronto.ca (C. Allen).
[1] These authors jointly supervised this work.

https://doi.org/10.1016/j.addr.2023.115108
Received 4 May 2023; Received in revised form 24 September 2023; Accepted 25 September 2023
Available online 27 September 2023

ARTICLE INFO

Keywords:
Machine learning
Active learning
Drug delivery
Drug development

ABSTRACT

Over the past few years, the adoption of machine learning (ML) techniques has rapidly expanded across many fields of research including formulation science. At the same time, the use of lipid nanoparticles to enable the successful delivery of mRNA vaccines in the recent COVID-19 pandemic demonstrated the impact of formulation science. Yet, the design of advanced pharmaceutical formulations is non-trivial and primarily relies on costly and time-consuming wet-lab experimentation. In 2021, our group published a review article focused on the use of ML as a means to accelerate drug formulation development. Since then, the field has witnessed significant growth and progress, reflected by an increasing number of studies published in this area. This updated review summarizes the current state of ML directed drug formulation development, introduces advanced ML techniques that have been implemented in formulation design and shares the progress on making self-driving laboratories a reality. Furthermore, this review highlights several future applications of ML yet to be fully exploited to advance drug formulation research and development.

1. Introduction

In recent years, there has been significant interest in the application of machine learning (ML), which is a branch of artificial intelligence, to accelerate development in the pharmaceutical sciences [1,2]. ML involves the development of algorithms to analyze data and identify statistical patterns or relationships. These relationships can be leveraged to make predictions about new data, derive insights for decision making, and discover underlying structures or characteristics in complex datasets. The process of learning from the data involves using optimization techniques to adjust the algorithm's parameters and improve its accuracy. Although both ML and statistics center around data analysis, they differ in their primary objective. Statistics traditionally focuses on hypothesis testing, confidence intervals, and model interpretation, while ML emphasizes predictive modeling, optimization, and pattern recognition [3]. Thus, ML has a wide range of applications, including image and speech recognition, natural language processing, and predictive analytics [4,5]. The broad applicability of these tools has given rise to a surge in the adoption of applied ML methods across many industrial sectors [6–8] including drug formulation design and development [1,2].

Pharmaceutical formulation plays a critical role in the development of safe, effective, and stable medicines. Through the optimization of drug formulations, pharmaceutical scientists can confer safety and/or efficacy improvements to therapeutic agents. These improvements can be the difference between clinical success and failure. For instance, it has recently been estimated that up to 90% of new therapeutics fail in clinical trials (from 2010 to 2017) due to poor efficacy, unacceptable toxicity, and/or poor drug-like properties [9]. However, optimized drug formulations can improve efficacy, reduce toxicity, and improve druggability. There are many examples of pharmaceutical formulations that have been designed to overcome such challenges. For example, Vyxeos® is a liposomal formulation, encapsulating a synergistic drug combination of daunorubicin and cytarabine, that results in an improvement in clinical efficacy relative to the free drug combination [10].

Optimizing drug formulations is essential for the development of safe and effective medicines, as it can significantly impact clinical success. However, the design and development of advanced pharmaceutical products is a complex process that requires significant time, resources, and expertise. This complexity arises from numerous factors, including the need to consider various parameters related to the drug, excipients, and manufacturing conditions within a high-dimensional design space. As a result, experimental evaluation of all parameter combinations is prohibitive. ML has the potential to enable pharmaceutical scientists to map the relationship between the composition and performance of advanced drug formulations to enhance *a priori* formulation design. Ultimately, ML tools may help navigate high-dimensional design spaces in search of drug formulations with targeted properties in a time and resource efficient manner.

In 2021, our group published a review article that examined efforts to integrate ML into drug formulation development [1]. Specifically, the article introduced fundamental ML principles (e.g., different ML models and the concept of cross-validation) to readers by summarizing the applications of ML in pharmaceutical research dating back to the 1990s. A Web of Science™ search shows the number of research articles that include drug formulation development and aspects of ML has continued to grow (Fig. 1). Of course, in recent years the term "machine learning" has gained prominence in research terminology, replacing specific modeling techniques such as linear regression and principal component analysis. This shift reflects the growing recognition of the broader field of ML as a powerful tool for data analysis and prediction. For instance, linear regression and principal component analysis are now commonly thought of as introductory ML techniques for supervised and unsupervised learning, respectively. While linear regression and principal component analysis are fundamental ML techniques, it is important to note that ML encompasses a vast array of methods and algorithms beyond these introductory techniques. More advanced approaches, such as random forest, support vector machine, neural network, kernel ridge regression, and deep learning, offer more complex and sophisticated modeling capabilities, each suited for different types of data and problem domains. It is these more advanced approaches that are typically incorporated into more recent studies to design advanced drug delivery systems. As shown in Fig. 1, there has been a surge in the use of ML, with 50% of all papers from the past two decades published over the past two years.

The current article builds on the previous article, with a summary of the latest studies that employ more advanced ML techniques to guide and accelerate development of a broader spectrum of drug formulations. In addition, this review outlines exciting future directions for the field. For a high-level introduction to ML methods and a stepwise summary of deploying ML pipelines in formulation development, we recommend the previous article [1]. For those with a more focused interest in a specific delivery strategy or dosage form, we recommend publications introducing ML in the context of solid dosage forms [11], hot-melt extrusion [12], nanomedicines [13], and 3D printing [14].

2. Recent applications of machine learning in advanced formulation development

2.1. Advanced oral delivery systems

Oral formulations are the most used medicines, accounting for 62% of all drug products currently on the US market [15]. This preference for oral products, in particular oral solid dosage forms, is generally attributed to their cost effectiveness, stability, ease of administration, and patient compliance [16]. Advanced oral delivery systems are designed to enhance the properties of therapeutic agents that are challenging to deliver via the gastrointestinal tract (i.e., due to poor solubility or intestinal permeability), or to provide extended release rates compared to conventional oral delivery systems (e.g., immediate-release tablets)

[17]. Recent research efforts in this space have investigated the use of ML to accelerate the design of various types of advanced oral delivery systems, including amorphous solid dispersions (ASD, [18–20]), 3D printed formulations [21–23]), and self-emulsifying drug delivery systems (SEDDS, [24,25]). For a more in-depth review of ML for oral dosage form development, the reader can refer to the recent article by Jiang et al [11].

2.1.1. Amorphous solid dispersions

ASDs offer a solution to the challenges associated with poorly water-soluble drugs by combining the drug with an excipient, typically a polymer, to create a solid mixture lacking an ordered crystalline structure. ASDs provide several advantages over the corresponding crystalline form of the drug, including increased apparent solubility, faster dissolution rates, and improved bioavailability [26,27]. More than 20 ASD products have been approved by the US FDA since 2000 [28]. The physicochemical nature of the excipients and drugs included in these formulations, as well as their relative proportions, can affect properties such as physical stability and dissolution rate of the ASD, and apparent solubility of the drug [29]. Consequently, the selection of suitable excipients has become a focus of many studies incorporating ML. For instance, Gao et al. compiled a dataset of 674 ASDs from the published literature [18]. The dataset was initially described by 374 input features, including ASD composition (e.g., polymer and drug types), *in vitro* release conditions (e.g., media), release sampling time points (i.e., 5 to 60 min), and drug/polymer molecular descriptors (e.g., drug melting point and polymer topological properties). The number of input features was reduced to 40 by removing irrelevant features via recursive feature elimination [30]. Using this refined representation, two random forest (RF) models were trained. The first, a classifier, was trained to distinguish between two types of drug release profiles from ASDs, "spring-and-parachute" and "maintain supersaturation" which accounted for 11% and 89% of the dataset, respectively. The second, a regressor, was trained only on the "maintain supersaturation" systems (i.e., 599 ASDs) to predict *in vitro* release. Both models were developed on 80% of the data (training set) using five-fold cross-validation and then evaluated using a test set (20%). The RF classifier had an accuracy of 0.95 (ACC, i.e., the fraction of correct predictions) and an F1 score of 0.97. The RF regressor had an accuracy of 0.781 R^2 and a root mean square error (RMSE) of 14.4. The definitions, ranges, and equations of these and other commonly used ML metrics are summarized in Table 1.

In addition to predicting drug release kinetics, ML has been employed to predict the size of ASD particles made by spray drying [20] as well as their physical stability during storage [19]. Schmitt et al. prepared 680 spray dried ASDs in-house and used the resulting dataset to train a panel of ML models. The best model was found to be an ensemble of support vector regressor (SVR), neural network (NN), and partial least squares (PLS). In the ensemble model, predictions were generated by averaging the outputs of the three models. The ensemble model exhibited a high degree of accuracy, as evidenced by a RMSE of 6 μm (particle sizes in the dataset ranged from 8 to 104 μm). To predict ASD physical stability, Lee et al. [19] developed a NN multiclass classifier trained with a smaller open access dataset [31]. The feedforward NN was comprised of two hidden layers with 512 and 128 neurons, respectively. For multiclass classification, the output layer implemented a Softmax function. The dataset which contained 123 formulations, with 103 used for model development by five-fold cross-validation and 20 for testing, was used to train ML models to predict ASD stability as a binary outcome over three timeframes (i.e., < 3, 3 to 6, > 6, months) [19]. The dataset used in this study was described by the authors as imbalanced. Imbalanced datasets can be biased toward the largest class and lead to prediction errors. To address this, the authors employed oversampling (i.e., synthetic minority oversampling technique, SMOTE) by creating synthetic instances of the minority class. Additionally, to further enhance the dataset and improve model classification performance, undersampling (i.e., edited nearest neighbour, ENN) was employed to remove instances at the boundaries of class predictions. The NN developed with this hybrid sampling strategy (SMOTE + ENN) was found to

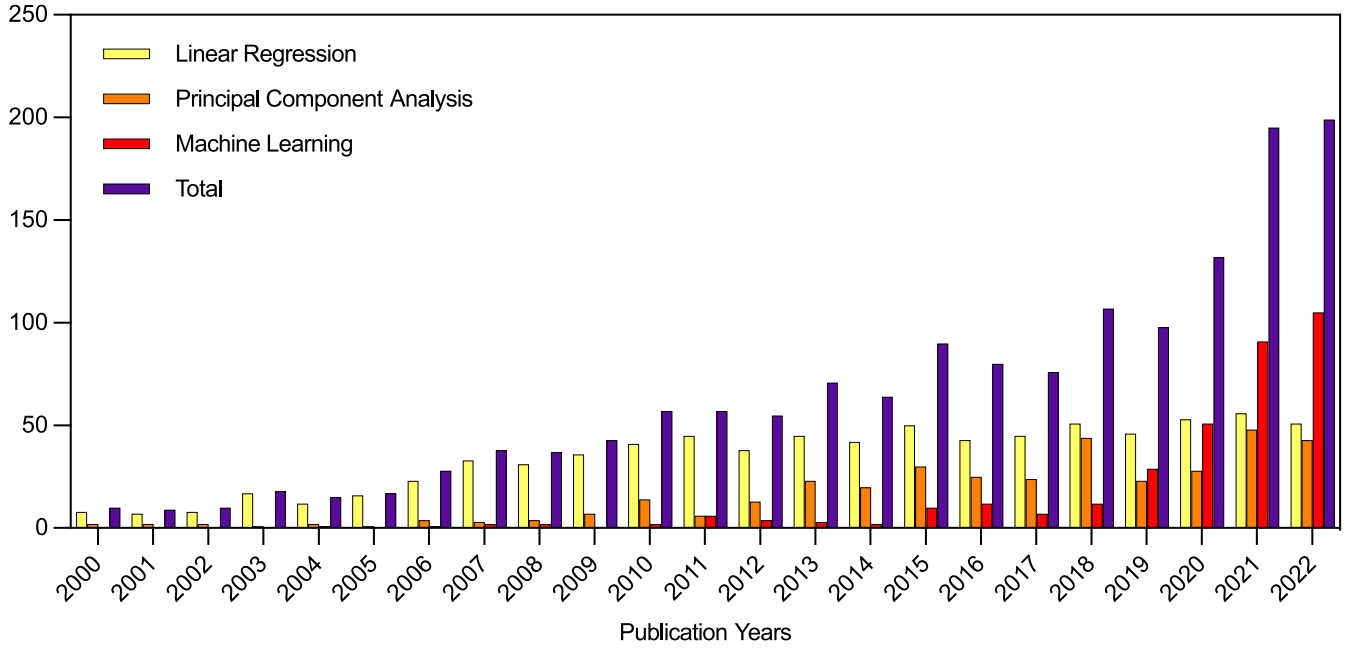


Fig. 1. Summary of the number of research articles focused on drug formulation and employing some aspect of machine learning. Sourced from Web of Science™ using keywords "linear regression" or "principal component analysis" or "machine learning" and "drug formulation" or "drug delivery" and applying the search to all fields (ALL). The resulting papers were limited to research articles published in the past two decades. The summation of publications across these categories is presented as the "Total". These searches can be reproduced with the following search queries: "ALL = ("linear regression" NOT "machine learning" NOT "principal component analysis") AND ALL = (drug formulation OR drug delivery)"; "ALL = ("principal component analysis" NOT "linear regression" NOT "machine learning") AND ALL = (drug formulation OR drug delivery)"; "ALL = ("machine learning" NOT "principal component analysis" NOT "linear regression") AND ALL = (drug formulation OR drug delivery)".

Table 1
A summary of commonly used metrics for machine learning model evaluation.

Metric	Definition	Range [worst, best]	Equation
Accuracy (ACC)	The fraction of correct predictions over all predictions.	[0, 1]	$ACC = \frac{\text{number of correct predictions}}{\text{number of all predictions}}$
F1 score	The harmonic mean of the model precision and recall.	[0, 1]	$F1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}}$
Coefficient of determination (R^2)	The proportion of the variance in the dependent variable that is explained by the model.	(-∞ to 1]	$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$
Root mean square error (RMSE)	The square root of the average of squared differences between actual and predicted values.	[0, ∞)	$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$
Mean absolute error (MAE)	The average of the absolute differences between actual and predicted values.	[0, ∞)	$MAE = \frac{\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert}{N}$
Mean absolute percentage error (MAPE)	The average of absolute percentage differences between actual and predicted values.	[0, ∞)	$MAPE = \frac{\sum_{i=1}^{N}\left\lvert\frac{y_i - \hat{y}_i}{y_i}\right\rvert}{N}$

be more accurate (ACC = 0.95; F1 score = 0.95) compared to the NN developed without sampling (ACC = 0.85; F1 score = 0.85) [32].

2.1.2. 3D printed oral formulations

In comparison to traditional processes used to prepare oral dosage forms (e.g., hot-melt extrusion, spray drying), 3D printing offers additional flexibility and advantages during small-scale production [33]. ML has been incorporated into manufacturing oral dosage forms using various 3D printing techniques [14]. For instance, ML has been used with fused deposition modeling [22,23,34,35], selective laser sintering [36], and inkjet printing [37,38].

Fused deposition modeling is commonly used to develop extended-release tablets [39]. Compared to conventional immediate-release, extended-release tablets can maintain therapeutic drug levels in the body over a longer period, reduce side effects, and improve patient compliance [40]. Elbadawi et al. [35], Castro et al. [22], and Ong et al. [23] recently reported the use of ML to accelerate the development of 3D printed extended-release tablets via fused deposition modeling (Fig. 2). The ML models were integrated into a web-based application called M3DISEEN [41], using the tablet composition as inputs to rapidly predict manufacturing parameters and *in vitro* release profiles.

M3DISEEN was initially developed by Elbadawi et al. through a study that included over 600 drug-loaded tablets that were extruded and 3D printed in-house [35]. The dataset was randomly split into training (75%) and test (25%) sets. A panel of ML models including multi-linear regression (MLR), k-nearest neighbors (kNN), support vector machine (SVM), RF, and NN with a feedforward structure, were developed for the regression and classification of manufacturing parameters. These included: printability (binary classification: yes or no), filament mechanical characteristics (multiclass classification: unextrudable, flexible, good, or brittle), extrusion and printing temperatures (both regression tasks). A five-fold cross-validation was applied to the training set to determine the optimal model hyperparameters. The optimal ML model for predicting individual parameters differed. For example, SVM achieved the highest ACC and F1 score predicting filament printability (0.76 for both), while the RF model was found to best predict printing temperature with a MAE of 8.4 °C and R^2 of 0.83.

To expand the scope of M3DISEEN, a follow-up study by Castro et al. focused on using a larger dataset collected from the literature (> 900 3D printed tablet formulations) [22]. The dataset was split into a training set (75%), a test set (25%), and 50-fold cross-validation was utilized on the training set. Castro et al. trained the same ML models as Elbadawi et al. and predicted the same manufacturing parameters with the addition of *in vitro* drug release (i.e., time to achieve 20%, 50%, and 80% release). Overall, the investigated models afforded relatively high accuracies for each of the different prediction tasks. The classification RF and the regression feedforward NN models were found to be optimal in predicting printability (ACC/F1 score = 0.93–0.94) and drug release (MAE = 24.29 min, R^2 = 0.86), respectively. Despite the promising performance of the models, the authors suggested that the dataset was imbalanced due to a lack of negative data (i.e., 85% of filaments were described as "good"). Thus, in a third study [23], the authors combined the literature-mined dataset [22] with their previously published experimental dataset [35]. The combined dataset was more balanced, with 66% of filaments described as "good". NN, SVM, and RF models were trained using this expanded dataset to predict manufacturing parameters (but not *in vitro* release). The optimal model was a classification RF, with a 0.84 ACC and a 0.80 F1 score when predicting printability and filament mechanical characteristics. The predictions for extrusion-and-printing temperatures had MAE values of 5.5 °C (R^2 = 0.91) and 6.0 °C (R^2 = 0.88), respectively.

2.2. Long-acting injectable systems

Long-acting injectables (LAIs) are formulations designed to provide sustained drug release following administration. The reduced fluctuation in drug levels and administration frequency afforded by LAIs make them well-suited for the treatment of chronic diseases such as osteoarthritis [42], cancer [43], HIV [44], and mental illness [45]. Excluding on-body devices, LAIs currently in clinical use can be categorized into three primary groups: polymeric systems, micro and nanocrystal drug suspensions, and oil-based formulations [46]. Given these formulations release drugs over an extended period, *in vitro* release assays are critical characterization steps during preclinical development. The timeframe for drug release from LAI products ranges from one week (e.g., Bydureon®) to six months (e.g., Trelstar®). As such, their extended *in vitro* drug release assays restrict iterative improvement in LAI design. To expedite the development cycle, researchers have investigated ML models to predict *in vitro* drug release *a priori* [47,48]. For example, our group recently trained 11 ML models on a dataset mined from the literature (i.e., 181 drug release profiles) to predict in vitro release of drugs from polymeric LAIs [47] (Fig. 3). The data was initially described by 17 features that included select physicochemical properties of the drugs, polymers, and LAIs, as well as experimental conditions, and three initial drug release values up to one day (T = 0.25, T = 0.5, and T = 1.0). These features were then used to train zero-shot and few-shot ML models. Zero-shot ML models were trained solely based on physicochemical properties and experimental conditions (14 features) without requiring early release data points. In contrast, few-shot ML models utilized early release data as additional input features (17 features in total). Both zero-shot and few-shot ML models were trained and assessed using a nested cross-validation approach, with the inner loop tuning hyperparameters and the outer loop evaluating model generalizability. Few-shot models were found to outperform zero-shot ones, as evaluated by the MAE on the outer loop. Amongst the few-shot models, LightGBM was optimal (i.e., lowest outer loop MAE of 0.114) and was further refined to remove highly correlated features. The complexity of ML

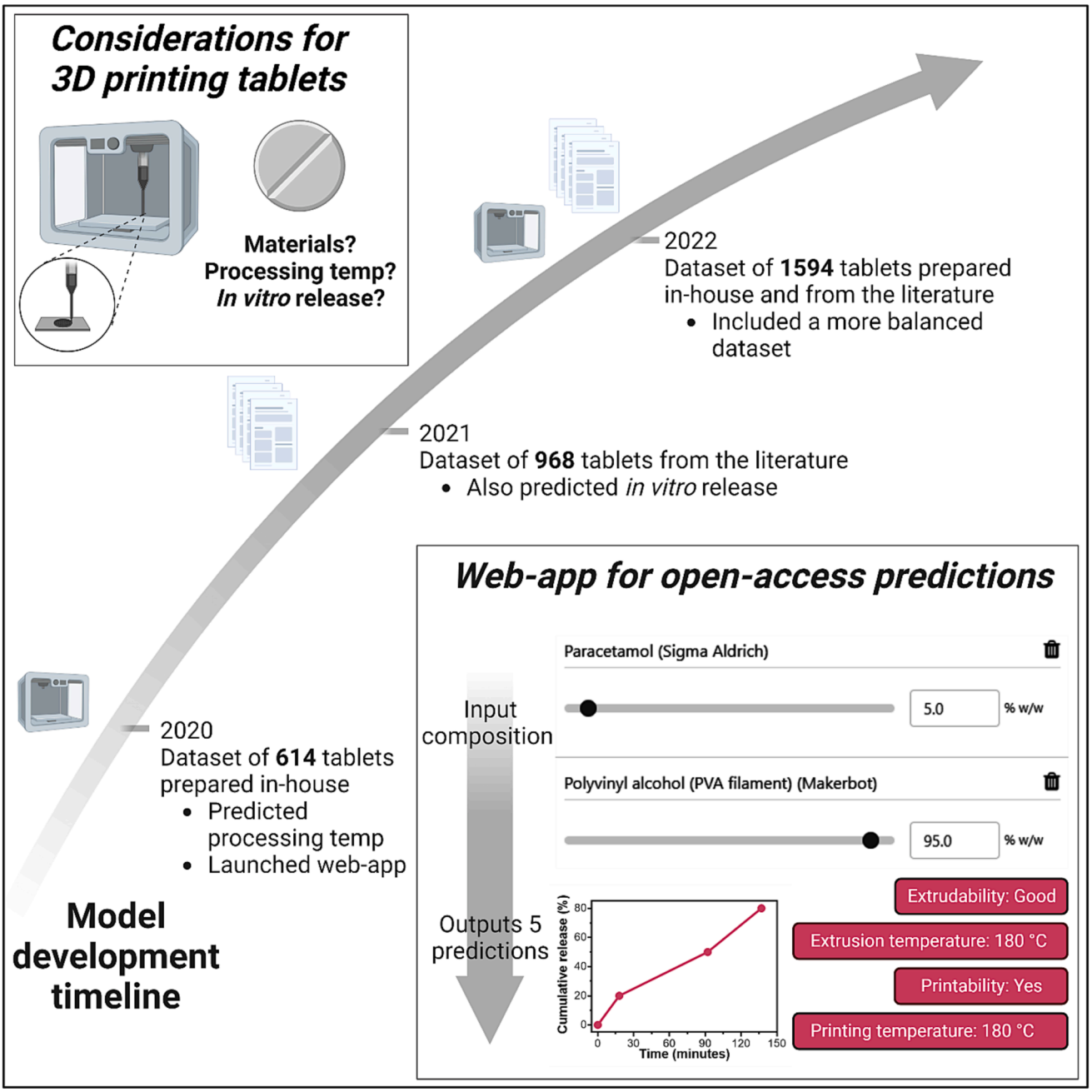


Fig. 2. Summary of the studies conducted by Elbadawi et al. [35], Castro et al. [22], and Ong et al. [23] on the development of models to predict the processing parameters and properties of 3D printed tablets. The web application M3DISEEN [41] requests the composition of a 3D printed material as input and yields five predictions including the extrudability, extrusion temperature, printability, printing temperature, and *in vitro* release profile.

models often makes them challenging to interpret in terms of how individual features impact overall prediction. To overcome this limitation, tools like SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations have been developed, each employing unique strategies to elucidate model insight [49]. In this study, SHAP was used to interpret the LightGBM model. Collective SHAP values indicated that time and drug release at T = 1.0 were the most influential input features for predicting fractional drug release. As time varies for a specific LAI release profile, drug release at T = 1.0 serves as a proxy for initial drug release rates. Feature SHAP values were further analyzed using unsupervised clustering methods (T-distributed stochastic neighbor embedding and principal component analysis). This revealed interesting correlations between T = 1.0 and other feature values. In particular, it enabled the identification of attributes that strongly contributed to fast and slow drug release from LAIs. These insights were applied to design two drug-loaded poly(lactide-co-glycolide) (PLGA)-based microparticle (MP) formulations not previously seen by the model. Specifically, a fast-release drug-loaded PLGA MP (i.e., salicylic acid + low molecular weight PLGA) and a slow-release system (i.e., olaparib + high molecular weight PLGA) were prepared and characterized in-house. Overall, good agreement was found between the predicted and experimental release profiles. The codes, dataset, and ML models developed through this study have been made openly available online [50].

In addition to predicting drug release, ML has been used to model other properties of LAIs including the average size of MPs [51,52] and drug loading levels [53]. To predict the size of MPs manufactured by electrospraying, Wang et al. trained ML models using data on 248

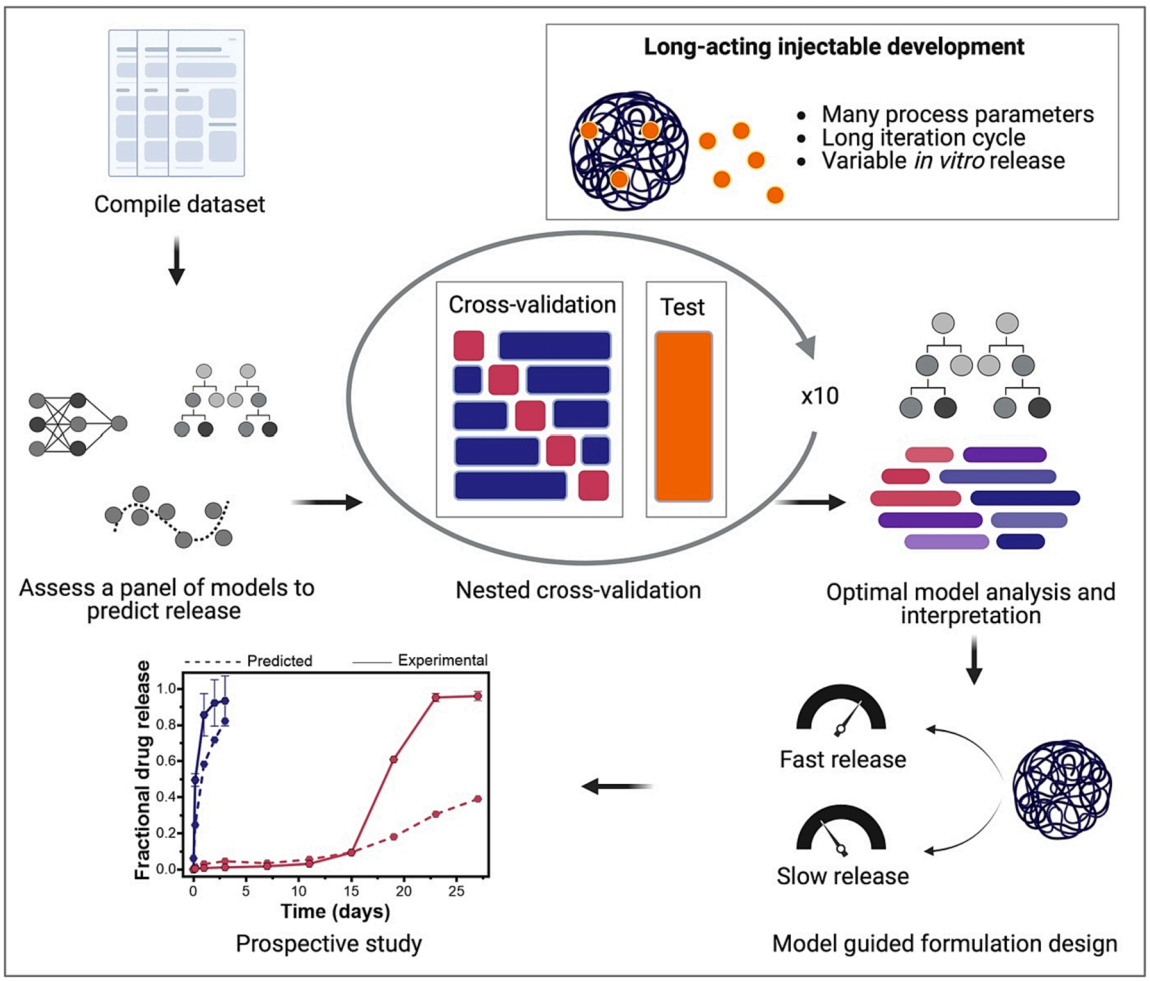


Fig. 3. Illustration of the workflow used by Bannigan et al. termed "data-driven LAI development" [47]. A total of 181 *in vitro* release profiles were collected from published literature and 17 input features were used to describe each experiment and predict fractional drug release. A panel of machine learning models were evaluated using a nested cross-validation approach. The optimal model, which had the lowest mean absolute error, was selected, further refined and subsequently analyzed to obtain insight into feature impact on the overall prediction. Two microparticle formulations were then designed prospectively using the developed model, with one formulation designed to provide fast drug release and another to provide slow release.

distinct PLGA formulations compiled from the literature [52]. Using this dataset, seven ML models (i.e., SVR, kernel ridge regression (KRR), kNN, feedforward NN (with 1–2 hidden layers and 4–20 neurons per layer), RF, XGBoost, and LightGBM) were trained and evaluated through a five-fold cross-validation. The XGBoost model was the optimal model showing the best accuracy (RMSE = 3.91 μm or MAPE = 50%). The generalizability of the model was then validated experimentally using 13 new MP formulations prepared in-house. Additionally, drug loading of MPs has been predicted with ML [53]. Specifically, approximately 100 MP formulations were prepared in-house using a hybrid extrusion-printing emulsification technique [53]. The resulting dataset was described with four manufacturing parameters (i.e., drug amount, printing speed, printing pressure, and nozzle size) as inputs and drug loading as the output. The data was split into training (85%) and test (15%) sets. The training set was used for model construction and hyperparameter optimization via a ten-fold cross-validation, and the test set was used for model evaluation. The authors concluded the decision tree (DT) outperformed other investigated models (RF, kNN, XGBoost, and LightGBM) with a MAE = 1.9 μg/mg and R^2 = 0.93. The developed models were further validated by experimental studies with the DT showing a high accuracy (MAE = 2.6 μg/mg, R^2 = 0.85).

2.3. Nanomedicines

Nanomedicines refer to nanotechnologies for use in a broad range of medical applications including prevention, monitoring, and treatment of disease. The recent success of lipid-based nanoparticles for delivery of mRNA and siRNA (i.e., Comirnaty™, and Spikevax™, Onpattro®) has significantly increased interest in nanomedicines. The development of nanomedicines has been an active area of research for decades, with early successes dating back to the 1990 s and the FDA approval of Doxil® (i.e., liposomal doxorubicin hydrochloride). Nanomedicines are relatively complex systems with their development posing significant challenges due to limitations in time, resources, and a historical reliance on Edisonian intuition. Overcoming these challenges requires a shift towards systematic and rational design strategies that leverage advanced computational tools, high-throughput screening, and data-driven approaches. By integrating interdisciplinary expertise and leveraging the power of emerging technologies, such as ML, it is possible to accelerate the development of nanomedicines, improve their properties, and perhaps even enhance their performance including therapeutic outcomes. In recent years, ML has been integrated into the design and evaluation of nanomedicines including to predict delivery efficiency [54–58], intratumoral biodistribution [59], drug release kinetics [60], cellular toxicity [54], stability [61], and particle size *a priori* [62,63].

For example, Gao et al. employed ML to predict the knockdown

efficiency of siRNA-loaded lipid nanoparticles (LNPs) [56]. Two datasets were collected from the literature and used separately to train two LightGBM models for the prediction of *in vitro* (129 formulations) and *in vivo* (301 formulations) siRNA gene knockdown efficiency. The knockdown efficiency was binarized using 50% as the threshold (i.e., > 50% knockdown efficiency was considered effective and *vice versa*). Both *in vitro* and *in vivo* models showed an ACC of around 0.8 for the classification tasks. The *in vitro* model was then validated experimentally using six siRNA-loaded LNP formulations prepared in-house. Cell transfection studies showed that all siRNA-loaded LNPs were effective, and the model was able to predict performance for five of the six formulations (i.e., ACC = 0.833) [56]. This study provides an intriguing exploration of the potential application of ML in predicting the performance of siRNA-loaded LNPs. To further enhance the robustness of the model, it would be beneficial to evaluate its ability to also predict negative experimental outcomes. For practical applications, the model's ability to accurately predict ineffective formulations is as important as its ability to predict effective ones.

ML methods have also been employed to develop nanomedicines targeted for oncology applications [59]. These applications classically rely on the passive tumor uptake of nanomedicines driven by the enhanced permeability and retention (EPR) effect [64]. The EPR effect, described in detail elsewhere [65,66], is a heterogenous phenomenon purported to result in the accumulation of nanomedicines at solid tumor sites due to leaky blood vessels and poor lymphatic drainage. The heterogeneity of EPR can partially be attributed to varying tumor microenvironments, including blood vessel density, cell density, and stromal components [67]. Through use of tissue section images as input features, advanced ML techniques can be employed to describe these complex microenvironments. For example, Tang et al. [59] proposed a generative adversarial network (GAN) to predict the intratumoral distribution of PEGylated CdSe/ZnS quantum dots (QDs) based on tumor microenvironment. GANs consist of two key components: a generator and a discriminator. The generator is responsible for producing synthetic data that is not present in the training dataset, while the discriminator's role is to differentiate between real data from the training dataset and the synthetic data generated by the generator. These two components engage in an adversarial training process, continually improving their performance. The ultimate objective is for the generator to create data that is so convincing that it becomes indistinguishable from the real training data, demonstrating its ability to generate highly authentic data. In this study, the authors trained the GAN model on a dataset collected from an in-house study that involved intravenous administration of PEGylated CdSe/ZnS QDs into six BALB/c mice bearing 4T1 breast cancer tumors. After 24 h, the tumors were sectioned, blood vessels and cell nuclei were stained (with CD31/Alex-Fluour-488 and DAPI, respectively) before imaging. The resulting images comprised three channels: CD31, DAPI, and QD, corresponding to spatial information on the blood vessels, cell nuclei, and QD biodistribution, respectively. The CD31 and DAPI channels accounted for tumor microenvironment and the QD channel showed the distribution of NPs in tumors. The images were decomposed into 512 × 512 pixel patches, resulting in 27,775 patches for training and 4,919 patches for validation. The GAN model was trained to predict QD intratumoral biodistribution (i.e., QD channel) using the tissue microenvironment (i.e., DAPI and CD31 channels). While the structure of and training strategies for GAN models are not the primary focus of this review, interested readers can refer to other publications for a more detailed introduction [68]. The developed GAN model accounted for 94% of the heterogeneity associated with the intratumoral biodistribution of the QDs (intraclass correlation = 0.94 on the validation set). While this model is likely limited to the prediction of the intratumoral distribution of PEGylated CdSe/ZnS QDs administered to BALB/c mice with 4T1 tumors, and by using inputs of stained tumor sections, it does demonstrate an interesting application of generative models in nanomedicines.

Up to this point, the studies discussed in this article have utilized supervised learning. Supervised learning relies on labeled datasets, using a specific set of input–output pairs to inform and guide the model's prediction patterns. Supervised learning applications in formulation development usually involve training a panel of models with a static dataset. Models which afford the best performance are then employed to rapidly screen novel formulations *in silico*. However, the flexibility of this approach is limited by the availability of high quality, open-access datasets, and their inherent biases [1]. Ultimately, supervised learning models tend not to extrapolate accurately to out-of-distribution (OOD) examples, especially when their uncertainty is not calibrated, and they are not generalizable. This issue arises because the models are trained on a specific distribution of data, making them prone to misinterpretation when encountering examples that deviate significantly from this distribution. However, OOD examples can be potentially interesting as they encompass formulations with undetermined properties, thereby presenting opportunities for true innovation.

Active learning (AL) can be employed to mitigate this limitation. While AL does not inherently overcome the problem of OOD prediction, the dynamic nature of AL facilitates rationale exploration of OOD instances. Compared to other similar approaches such as design of experiments and response surface methodology which require human decision-making [69], AL autonomously refines its understanding of the design space and then recommends the next instances to be evaluated. Bayesian optimization (BO) [70] is a prototypical AL strategy which combines a surrogate model (e.g., Gaussian processes, GPs), approximating the parameter-property relationship, with an acquisition function that determines which instances to investigate next. From a high level, GPs are a class of nonparametric models which apply normal distributions to infinite domains while retaining many attractive properties of normal distributions and estimate uncertainty. Currently, BO is considered the gold-standard approach to noisy black-box optimization [71]. Black-box optimization techniques are well-suited to drug formulation as the complex relationships between formulation manufacturing parameters and resulting formulation properties can be difficult to intuit explicitly.

AL workflows can be described as an iterative "build-test-learn-design" cycle [72]. When applied to formulation design, this process begins with the preparation ("build") and characterization ("test") of a small set of formulations within the design space, serving as the initial dataset. This initial dataset is used to train the AL model ("learn"), which then proposes formulation(s) expected to exhibit target properties ("design"). The synergy of AL strategies like BO with automated/robotic laboratory equipment constitutes a self-driving laboratory (SDL) or materials acceleration platform (MAP) [73–78]. Both SDLs and MAPs are considered potential next-generation technologies for scientific experimentation and have been recently prototyped for the accelerated design of chemical syntheses [79,80], catalysts [81], inorganic materials [82–85] and are beginning to see adoption in formulation development [55,61,86].

As an illustrative example, Tamasi et al. [61] applied BO and a robotic platform to accelerate the design of polymer-protein hybrids (PPHs) for three enzymes (glucose oxidase, Gox; lipase, Lip; and horseradish peroxidase, HRP) to maximize enzymatic activity (EA) following thermal stress (Fig. 4). Before conducting AL-guided experimentation, the authors constructed an initial dataset of 504 PPHs using expert intuition. PPHs were described using copolymer characteristics, such as monomer composition and degree of polymerization. The authors proceeded with batched BO (using a GP as a surrogate model), in which batches of 24 candidate copolymers were designed, built by automated photoinduced electron/energy transfer-reversible addition-fragmentation chain transfer (PET-RAFT) polymerization, and tested for EA following thermal stress. This "build-test-learn-design" loop was repeated for five batches, producing PPHs with enhanced thermostability for three distinct enzymes (46.2%, 31.5%, and 87.6% improvement in comparison to the initial seed batch for HRP, Gox, and Lip, respectively). SHAP analysis was employed to further quantify the

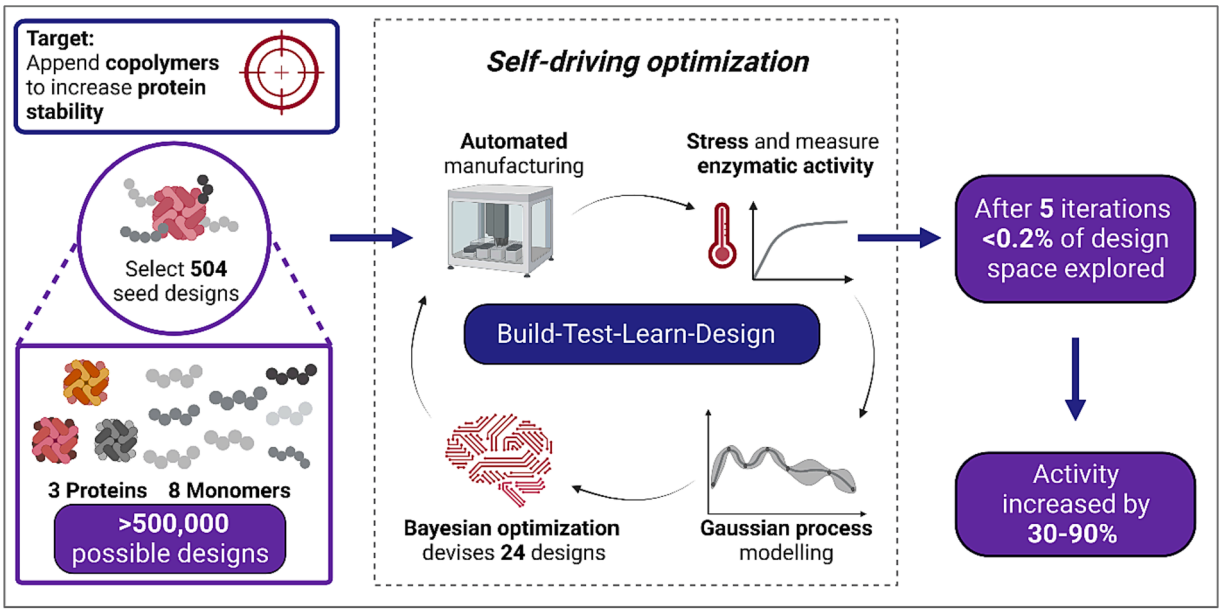


Fig. 4. Illustration of the study by Tamasi et al. [61]. First, three possible enzymes (glucose oxidase, horseradish peroxidase and lipase) and eight possible monomers (ionic, hydrophilic, hydrophobic) are selected to prepare polymer-protein hybrids (PPH). The design space for the PPH consisted of one protein and up to four distinct monomers with a total degree of polymerization of 50–200 in increments of 25. From the set of 545,622 possibilities, 504 seed designs were selected and manufactured using an automated liquid handling platform. The PPHs were thermally stressed, and retention of enzymatic activity was evaluated. Bayesian optimization was then used to design batches of 24 new PPHs which were synthesized in a Build-Test-Learn-Design cycle. This was repeated a total of five times and the optimized PPHs showed a 30 to 90% improvement in the enzymatic activity compared to the seed designs.

contribution of the copolymer composition and degree of polymerization to EA predictions. The SHAP analysis showed that the contributions of copolymer features to EA were found to be enzyme-dependent. For example, a degree of polymerization ranging from 50–75 increased the EA for HRP PPHs but decreased the EA for Lip PPHs.

Overall, various ML techniques have been employed for formulation development in the last few years. A summary of the publications highlighted in this review are included in Table 2.

3. Outlook

Pharmaceutical formulation plays a key role in the drug development process. It is a critical step in the design and development of drug products that are safe, effective, stable, and suitable for administration to patients. For decades the process of engineering such drug products has involved continuous cycles of experimentation and evaluation to identify formulation compositions. While this approach has resulted in the development of many drug products, it remains inefficient and resource intensive. In the last few years, ML has been applied increasingly to accelerate the design of advanced drug delivery systems. Despite the advances that have been made, there remain several challenges hindering the widespread adoption of ML in pharmaceutical formulation development.

The first challenge is the scarcity of professionals who possess expertise in both pharmaceutical formulation development and ML. As a result, meaningful integration of ML into formulation development is currently reliant on collaboration between pharmaceutical and computer scientists. However, as the years go by it is expected that more and more individuals will possess both AI/ML knowledge and skills as well as domain specific expertise. Another challenge is the limited access to data, algorithms, or research findings for scientists to use freely, to modify, and to distribute. Although an increasing number of researchers are choosing to make their data and code open access (via public repositories such as Github and Zenodo), and some researchers have launched their own web-based applications to enable the use of trained ML models [41], there is still a reluctance to share many of the research outputs with the broader formulation community. As a result, other researchers may collect similar datasets and retrain models instead of building upon existing work. This will not only limit the potential impact of the research but also slow down the progress in the field as a whole. To fully realize the potential of ML-based formulation development, it is essential to embrace open science principles and share research outputs in a transparent and accessible manner. Finally, there is what is often referred to as selective reporting in the scientific literature which makes it difficult to reproduce studies and to curate reliable datasets from publications [87,88]. Beyond this, as highlighted by Lammers et al. [89], there is a lack of standardization in studies conducted to characterize formulations such as nanoparticles. The lack of open access datasets and selective reporting in the literature has driven many to pursue creation of their own datasets through integration of automation into their research [55,61,86].

In our previous review, we identified several advanced ML techniques, including generative ML models and AL/SDL, that were yet to be fully utilized by pharmaceutical scientists but have the potential to further expedite formulation design [1]. In recent years, we have seen pioneering formulation studies integrating these approaches. For example, as summarized previously generative models have been deployed in a proof of concept study for the characterization of the intratumoral biodistribution of PEGylated CdSe/ZnS dots [59]. This is a pivotal study that may encourage more pharmaceutical scientists to employ generative models, which have been used in other fields (e.g., chemistry and materials science) [90,91]. Moreover, other research groups have begun to deploy AL for formulation optimization [55,61]. The goal of using AL is to autonomously select the next batch of experiments to perform in such a way that the outcomes of these experiments will be closer to the targets. When AL is combined with robotic systems that can automate these selected experiments, the resulting platforms are called SDLs [55,61]. The development of such autonomous platforms has immense transformative potential especially if more clinically relevant assays can be deployed in SDLs. On this front, there is increasing regulatory acceptance of organ-on-a-chip technology, which can be integrated into SDLs. For example, the 2021 report “Advancing

Table 2
This table provides a summary of the studies highlighted in this review article that have incorporated machine learning techniques into drug formulation development. These studies are categorized based on the type of formulation, including advanced oral delivery systems, long-acting injectables, and nanomedicines.

Formulation		Therapeutic	Model Target	ML Algorithm	Reference
Advanced oral delivery systems	Amorphous solid dispersion	Small molecule	*In vitro* release	RF	[18]
	Amorphous solid dispersion	Small molecule	Physical stability	NN	[19]
	Amorphous solid dispersion	Small molecule	Particle size	PLS, Ridge, Lasso, SVR, KRR, XGBoost, NN, Ensemble: SVR, NN, PLS	[20]
	3D printed formulation	Small molecule	Manufacturing parameter and *in vitro* release	MLR, kNN, SVM, RF, NN	[22]
	3D printed formulation	Small molecule	Manufacturing parameters	NN, SVM, RF	[23]
	Self-emulsifying drug delivery system	Small molecule	Apparent degree of supersaturation	NN, PLS	[24]
	Self-emulsifying drug delivery system	Small molecule	Pseudo-ternary phase diagram	RF, kNN, DT, Naïve Bayes, SVM, LightGBM, XGBoost	[25]
	3D printed formulation	Small molecule	Manufacturing parameters	MLR, kNN, SVM, RF, NN	[35]
Long-acting injectable systems	Polymer microparticle	Small molecule	*In vitro* release	LightGBM, RF, NGB, XGBoost, DT, NN, kNN, SVR, Lasso, PLS, MLR	[47]
	Polymer microparticle	Small molecule	*In vitro* release	XGBoost, RF, LightGBM, ResNet, NN, DT, kNN, SVM, PLS, Lasso, Ridge, MLR	[48]
	Polymer microparticle	N/A	Droplet size	NN	[51]
	Polymer microparticle	N/A	Particle size	SVR, RF, NN, kRR, kNN, XGBoost, LightGBM	[52]
	Polymer microparticle	Small molecule	Drug loading	DT, RF, kNN, LightGBM, XGBoost	[53]
Nanomedicines	Polymer nanoparticle	Nucleic acid	Delivery efficiency and cytotoxicity	RF, XGBoost, DT, kNN, Relevance determination regression, Elastic net regression	[54]
	Polyplex	Nucleic acid	Delivery efficiency	Bayesian optimization	[55]
	Lipid nanoparticle	Nucleic acid	Delivery efficiency	LightGBM	[56]
	Inorganic/organic nanoparticle	N/A	Delivery efficiency	MLR, kNN, RF, Bagged model, GBM, SVM, NN	[57]
	Lipid nanoparticle	Nucleic acid	Delivery efficiency	LightGBM	[58]
	Quantum dot	N/A	Intratumoral biodistribution	GAN	[59]
	Polymer nanoparticle	Small molecule	*In vitro* release	NN	[60]
	Polymer-protein hybrid	Protein	Thermal stability	Bayesian optimization	[61]
	Polymer nanoparticle	N/A	Particle size	Neurofuzzy logic	[62]
	Liposome	N/A	Particle size	NN	[63]

Abbreviations: random forest (RF), neural network (NN), partial least squares (PLS), support vector regression (SVR), kernel ridge regression (KRR), extreme gradient boosted trees (XGBoost), multiple linear regression (MLR), k-nearest neighbor (kNN), support vector machines (SVM), light gradient boosting machine (LightGBM), residual neural network (ResNet), kernel ridge regression (kRR), gradient boosting model (GBM), generative adversarial network (GAN). N/A denotes that no therapeutic agents were incorporated into the formulations.

New Alternative Methodologies at FDA" presents a strategic roadmap for incorporating cutting-edge organ-on-a-chip technology into pre-clinical development workflows. Reinforcing this trend, the US FDA Modernization Act indicates that the regulatory body will begin to accept toxicity data generated from organ-on-a-chip technology in place of pre-clinical animal models [92]. These developments are crucial in efforts to reduce our reliance on animal studies, which pose ethical concerns and require considerable amounts of time and resources. Overall, the adoption of SDL workflows could dramatically expedite preclinical development.

In the coming years, the inclusion of ML is expected to play an increasingly important role in drug product development. With each passing year, advances in computational hardware and the increased availability of open access software packages increase ML accessibility [93]. Furthermore, a number of recently published papers aimed at pharmaceutical scientists also provide near-stepwise instructions to use ML approaches for drug formulation design [2,94–97]. The recent release of ChatGPT which was developed based on large language models (LLMs) has further reduced the barrier to entry for individuals interested in coding and training ML models. In addition, these LLMs also possess significant potential to rapidly collate and summarize relevant literature [98], which can streamline dataset construction for ML model training. However, building these LLMs requires substantial computational resources, expertise in ML, and access to a significant amount of training data. Therefore, developing LLMs from scratch may not be feasible for most individuals or small teams. Alternatively, investigators can leverage existing pre-trained models (e.g., GPT-3) which are readily available and can be fine-tuned for specific tasks with a smaller dataset. This approach allows researchers to benefit from the advancements made by companies like OpenAI while reducing the computational and data requirements for training such models. In parallel, automated ML (AutoML) has emerged in recent years [99]. These platforms (e.g., Google Cloud AutoML and Microsoft Azure) have increased accessibility to non-ML specialists by reducing the coding and data science skills required to deploy ML models. For example, using Google Cloud AutoML, two ophthalmology trainees with no coding background successfully built a deep learning model to classify the individual steps (i.e., phases) of cataract surgery from publicly available recordings of the procedure [100].

There also remain several exciting ML techniques that are yet to be fully explored for drug formulation development. For example, meta-learning is a promising approach to be considered. Meta-learning, often referred to as "learning to learn", is a subfield of ML where algorithms are designed to improve their knowledge or performance through experience gained from prior tasks [101,102]. This approach facilitates rapid adaptation to new tasks, leveraging previously learned concepts. In the context of drug formulation design, this could be a valuable strategy where the scarcity of training data poses a significant challenge for traditional ML approaches. SDLs, another platform requiring fewer data to initialize the model, still has significant unexploited potential.

Although there have been exciting advancements in SDLs [55,61,86], current platforms still require human intervention to assist with some procedures in the loop. However, with advancements in experimental automation and ML algorithms, we anticipate a reduction in human involvement and an increase in experimental throughput. As these new technologies continue to develop and mature, they will undoubtedly bring about transformative changes in the field of drug formulation design.

Perhaps one of the best indicators of progress in this field is the emergence of start-up companies (such as Persist AI and Nanite Inc.) that intend to leverage ML techniques to monetize drug product development as a service. These start-ups aim to capitalize on the potential of ML to accelerate drug development and reduce development costs. Besides these start-ups, there is a growing adoption of ML by well-established companies. For example, Simulations Plus Inc. has employed ML to build ADMET Predictor® as a platform to assist in the prediction of the pharmacokinetics of a drug. These ML models have the potential to predict physiological based pharmacokinetic (PBPK) model parameters, a concept more thoroughly detailed elsewhere [103,104]. These models can be particularly advantageous for extended-release dosage forms such as LAIs, where pharmacokinetic studies require significant time.

In the coming years, as ML directed drug formulation development becomes more mainstream, we anticipate that many more companies will emerge at this interface between computer and pharmaceutical sciences. By enabling the identification of more promising candidate formulations earlier in the development process, the use of ML in the pharmaceutical industry is poised to drive significant progress. This will result in the development of safer, more effective, and more accessible drug products for patients.

Funding sources

This research was undertaken thanks in part to funding provided to the University of Toronto's Acceleration Consortium from the Canada First Research Excellence Fund CFREF-2022-00042. NSERC Discovery grant (RGPIN-2022-04910) to C.A. A.A.G. acknowledges support from the Defense Advanced Research Projects Agency under the Accelerated Molecular Discovery Program under Cooperative Agreement No. HR00111920027 dated August 1, 2019. Z.B. acknowledges a CPO scholarship from the Centre for Pharmaceutical Oncology, Leslie Dan Faculty of Pharmacy, University of Toronto. J.B. acknowledges a Dean's scholarship from the Leslie Dan Faculty of Pharmacy. R.J.H. acknowledges NSERC for a postgraduate scholarship (PGSD3-534584-2019), as well as support from the Vector Institute. A.A.G. would like to thank Dr. Anders Frøseth for his support.

CRediT authorship contribution statement

Zeqing Bao: Investigation, Visualization, Writing – original draft. **Jack Bufton:** Investigation, Visualization, Writing – review & editing. **Riley J. Hickman:** Investigation, Writing – review & editing. **Alán Aspuru-Guzik:** Project administration, Writing – review & editing, Funding acquisition. **Pauric Bannigan:** Conceptualization, Project administration, Writing – review & editing, Supervision. **Christine Allen:** Conceptualization, Project administration, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

Z.B., R.J.H., A.A.-G., P.B., and C.A. are founding members of a new company, 15073383 Canada Inc. A.A.-G is a founding member of Kebotix, Inc.

Acknowledgment

Figures were created with BioRender.com and Prism.

References

[1] P. Bannigan, M. Aldeghi, Z. Bao, F. Häse, A. Aspuru-Guzik, C. Allen, Machine learning directed drug formulation development, Adv. Drug Deliv. Rev. 175 (2021), 113806, https://doi.org/10.1016/j.addr.2021.05.016.
[2] N. Wang, Y. Zhang, W. Wang, Z. Ye, H. Chen, G. Hu, D. Ouyang, How can machine learning and multiscale modeling benefit ocular drug development? Adv. Drug Deliv. Rev. 196 (2023), 114772 https://doi.org/10.1016/j.addr.2023.114772.
[3] D. Bzdok, N. Altman, M. Krzywinski, Statistics versus machine learning, Nat. Methods 15 (2018) 233–234, https://doi.org/10.1038/nmeth.4642.
[4] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, https://doi.org/10.1038/nature14539.
[5] A.R. Samanpour, A. Ruegenberg, R. Ahlers, The Future of Machine Learning and Predictive Analytics, in: C. Linnhoff-Popien, R. Schneider, M. Zaddach (Eds.), Digital Marketplaces Unleashed, Springer, Berlin, Heidelberg, 2018, pp. 297–309, https://doi.org/10.1007/978-3-662-49275-8_30.
[6] X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski, T.-Y.-J. Han, Explainable machine learning in materials science, npj Comput. Mater. 8 (2022) 1–19, https://doi.org/10.1038/s41524-022-00884-7.
[7] A. Torfi, R.A. Shirvani, Y. Keneshloo, N. Tavaf, E.A. Fox, Natural Language Processing Advancements By Deep Learning: A Survey, (2021). https://doi.org/10.48550/arXiv.2003.01200.
[8] S. Dara, S. Dhamercherla, S.S. Jadav, C.M. Babu, M.J. Ahsan, Machine Learning in Drug Discovery: A Review, Artif. Intell. Rev. 55 (2022) 1947–1999, https://doi.org/10.1007/s10462-021-10058-4.
[9] D. Sun, W. Gao, H. Hu, S. Zhou, Why 90% of clinical drug development fails and how to improve it? Acta Pharm. Sin. B 12 (2022) 3049–3062, https://doi.org/10.1016/j.apsb.2022.02.002.
[10] A.C. Anselmo, S. Mitragotri, Nanoparticles in the clinic: An update, Bioeng. Transl. Med. 4 (2019), e10143, https://doi.org/10.1002/btm2.10143.
[11] J. Jiang, X. Ma, D. Ouyang, R.O. Williams, Emerging Artificial Intelligence (AI) Technologies Used in the Development of Solid Dosage Forms, Pharmaceutics 14 (2022) 2257, https://doi.org/10.3390/pharmaceutics14112257.
[12] N. Munir, M. Nugent, D. Whitaker, M. McAfee, Machine Learning for Process Monitoring and Control of Hot-Melt Extrusion: Current State of the Art and Future Directions, Pharmaceutics 13 (2021) 1432, https://doi.org/10.3390/pharmaceutics13091432.
[13] N. Serov, V. Vinogradov, Artificial intelligence to bring nanomedicine to life, Adv. Drug Deliv. Rev. 184 (2022), 114194, https://doi.org/10.1016/j.addr.2022.114194.
[14] M. Elbadawi, L.E. McCoubrey, F.K.H. Gavins, J.J. Ong, A. Goyanes, S. Gaisford, A. W. Basit, Harnessing artificial intelligence for the next generation of 3D printed medicines, Adv. Drug Deliv. Rev. 175 (2021), 113805, https://doi.org/10.1016/j.addr.2021.05.015.
[15] H. Zhong, G. Chan, Y. Hu, H. Hu, D. Ouyang, A Comprehensive Map of FDA-Approved Pharmaceutical Products, Pharmaceutics. 10 (2018) 263, https://doi.org/10.3390/pharmaceutics10040263.
[16] A.C. Anselmo, Y. Gokarn, S. Mitragotri, Non-invasive delivery strategies for biologics, Nat. Rev. Drug Discov. 18 (2019) 19–40, https://doi.org/10.1038/nrd.2018.183.
[17] S. Salunke, F. O'Brien, D. Cheng Thiam Tan, D. Harris, M.-C. Math, T. Ariën, S. Klein, C. Timpe, Oral drug delivery strategies for development of poorly water soluble drugs in paediatric patient population, Adv. Drug Deliv. Rev. 190 (2022), 114507, https://doi.org/10.1016/j.addr.2022.114507.
[18] H. Gao, W. Wang, J. Dong, Z. Ye, D. Ouyang, An integrated computational methodology with data-driven machine learning, molecular modeling and PBPK modeling to accelerate solid dispersion formulation design, Eur. J. Pharm. Biopharm. 158 (2021) 336–346, https://doi.org/10.1016/j.ejpb.2020.12.001.
[19] H. Lee, J. Kim, S. Kim, J. Yoo, G.J. Choi, Y.-S. Jeong, Deep Learning-Based Prediction of Physical Stability considering Class Imbalance for Amorphous Solid Dispersions, J. Chem. 2022 (2022), e4148443, https://doi.org/10.1155/2022/4148443.
[20] J.M. Schmitt, J.M. Baumann, M.M. Morgen, Predicting Spray Dried Dispersion Particle Size Via Machine Learning Regression Methods, Pharm. Res. (2022), https://doi.org/10.1007/s11095-022-03370-3.
[21] A. Lu, J. Zhang, J. Jiang, Y. Zhang, B.R. Giri, V.R. Kulkarni, N.H. Aghda, J. Wang, M. Maniruzzaman, Novel 3D Printed Modular Tablets Containing Multiple Anti-Viral Drugs: a Case of High Precision Drop-on-Demand Drug Deposition, Pharm. Res. 39 (2022) 2905–2918, https://doi.org/10.1007/s11095-022-03378-9.
[22] B. Muñiz Castro, M. Elbadawi, J.J. Ong, T. Pollard, Z. Song, S. Gaisford, G. Pérez, A.W. Basit, P. Cabalar, A. Goyanes, Machine learning predicts 3D printing performance of over 900 drug delivery systems, J. Control. Release 337 (2021) 530–545, https://doi.org/10.1016/j.jconrel.2021.07.046.
[23] J.J. Ong, B.M. Castro, S. Gaisford, P. Cabalar, A.W. Basit, G. Pérez, A. Goyanes, Accelerating 3D printing of pharmaceutical products using machine learning, Int. J. Pharm. X 4 (2022), 100120, https://doi.org/10.1016/j.ijpx.2022.100120.
[24] H. Bennett-Lenane, J.P. O'Shea, J.D. Murray, A.-R. Ilie, R. Holm, M. Kuentz, B. T. Griffin, Artificial Neural Networks to Predict the Apparent Degree of Supersaturation in Supersaturated Lipid-Based Formulations: A Pilot Study,

Pharmaceutics 13 (2021) 1398, https://doi.org/10.3390/pharmaceutics13091398.
[25] H. Gao, H. Jia, J. Dong, X. Yang, H. Li, D. Ouyang, Integrated in silico formulation design of self-emulsifying drug delivery systems, Acta Pharm. Sin. B 11 (2021) 3585–3594, https://doi.org/10.1016/j.apsb.2021.04.017.
[26] T. Vasconcelos, S. Marques, J. das Neves, B. Sarmento, Amorphous solid dispersions: Rational selection of a manufacturing process, Adv. Drug Deliv. Rev. 100 (2016) 85–101, https://doi.org/10.1016/j.addr.2016.01.012.
[27] S.A. Thompson, R.O. Williams, Specific mechanical energy – An essential parameter in the processing of amorphous solid dispersions, Adv. Drug Deliv. Rev. 173 (2021) 374–393, https://doi.org/10.1016/j.addr.2021.03.006.
[28] A. Butreddy, Hydroxypropyl methylcellulose acetate succinate as an exceptional polymer for amorphous solid dispersion formulations: A review from bench to clinic, Eur. J. Pharm. Biopharm. 177 (2022) 289–307, https://doi.org/10.1016/j.ejpb.2022.07.010.
[29] R. Iyer, V. Petrovska Jovanovska, K. Berginc, M. Jaklič, F. Fabiani, C. Harlacher, T. Huzjak, M.V. Sanchez-Felix, Amorphous Solid Dispersions (ASDs): The Influence of Material Properties, Manufacturing Processes and Analytical Technologies in Drug Product Development, Pharmaceutics. 13 (2021) 1682. https://doi.org/10.3390/pharmaceutics13101682.
[30] P.M. Granitto, C. Furlanello, F. Biasioli, F. Gasperi, Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products, Chemom. Intel. Lab. Syst. 83 (2006) 83–90, https://doi.org/10.1016/j.chemolab.2006.01.007.
[31] Y. Yang, DeepPharm, (2018). https://github.com/yylonly/DeepPharm-InVitro (accessed March 24, 2023).
[32] R. Han, H. Xiong, Z. Ye, Y. Yang, T. Huang, Q. Jing, J. Lu, H. Pan, F. Ren, D. Ouyang, Predicting physical stability of solid dispersions by machine learning techniques, J. Control. Release 311–312 (2019) 16–25, https://doi.org/10.1016/j.jconrel.2019.08.030.
[33] N. Gottschalk, M. Bogdahn, J. Quodbach, 3D printing of amorphous solid dispersions: A comparison of fused deposition modeling and drop-on-powder printing, Int. J. Pharm.: X 5 (2023), 100179, https://doi.org/10.1016/j.ijpx.2023.100179.
[34] M. Elbadawi, T. Gustaffson, S. Gaisford, A.W. Basit, 3D printing tablets: Predicting printability and drug dissolution from rheological data, Int. J. Pharm. 590 (2020), 119868, https://doi.org/10.1016/j.ijpharm.2020.119868.
[35] M. Elbadawi, B. Muñiz Castro, F.K.H. Gavins, J.J. Ong, S. Gaisford, G. Pérez, A. W. Basit, P. Cabalar, A. Goyanes, M3DISEEN: A novel machine learning approach for predicting the 3D printability of medicines, Int. J. Pharm. 590 (2020), 119837, https://doi.org/10.1016/j.ijpharm.2020.119837.
[36] Y. Abdalla, M. Elbadawi, M. Ji, M. Alkahtani, A. Awad, M. Orlu, S. Gaisford, A. W. Basit, Machine learning using multi-modal data predicts the production of selective laser sintered 3D printed drug products, Int. J. Pharm. 633 (2023), 122628, https://doi.org/10.1016/j.ijpharm.2023.122628.
[37] C.S. O'Reilly, M. Elbadawi, N. Desai, S. Gaisford, A.W. Basit, M. Orlu, Machine Learning and Machine Vision Accelerate 3D Printed Orodispersible Film Development, Pharmaceutics 13 (2021) 2187, https://doi.org/10.3390/pharmaceutics13122187.
[38] P. Carou-Senra, J.J. Ong, B.M. Castro, I. Seoane-Viaño, L. Rodríguez-Pombo, P. Cabalar, C. Alvarez-Lorenzo, A.W. Basit, G. Pérez, A. Goyanes, Predicting pharmaceutical inkjet printing outcomes using machine learning, Int. J. Pharm.: X. 5 (2023), 100181, https://doi.org/10.1016/j.ijpx.2023.100181.
[39] J. Skowyra, K. Pietrzak, M.A. Alhnan, Fabrication of extended-release patient-tailored prednisolone tablets via fused deposition modelling (FDM) 3D printing, Eur. J. Pharm. Sci. 68 (2015) 11–17, https://doi.org/10.1016/j.ejps.2014.11.009.
[40] S.K. Patel, M. Khoder, M. Peak, M.A. Alhnan, Controlling drug release with additive manufacturing-based solutions, Adv. Drug Deliv. Rev. 174 (2021) 369–386, https://doi.org/10.1016/j.addr.2021.04.020.
[41] M3DISEEN, (n.d.). https://m3diseen.com/ (accessed March 24, 2023).
[42] V.B. Kraus, P.G. Conaghan, H.A. Aazami, P. Mehra, A.J. Kivitz, J. Lufkin, J. Hauben, J.R. Johnson, N. Bodick, Synovial and systemic pharmacokinetics (PK) of triamcinolone acetonide (TA) following intra-articular (IA) injection of an extended-release microsphere-based formulation (FX006) or standard crystalline suspension in patients with knee osteoarthritis (OA), Osteoarthr. Cartil. 26 (2018) 34–42, https://doi.org/10.1016/j.joca.2017.10.003.
[43] S.R. Abulateefeh, Long-acting injectable PLGA/PLA depots for leuprolide acetate: successful translation from bench to clinic, Drug Deliv. Transl. Res. 13 (2023) 520–530, https://doi.org/10.1007/s13346-022-01228-0.
[44] N.G. Kotla, A. Pandey, Y. Vijaya Kumar, F. Ramazani, A. Fisch, Polyester-based long acting injectables: Advancements in molecular dynamics simulation and technological insights, Drug Discov. Today 28 (2023), 103463, https://doi.org/10.1016/j.drudis.2022.103463.
[45] S. Miyamoto, W. Wolfgang Fleischhacker, The Use of Long-Acting Injectable Antipsychotics in Schizophrenia, Curr Treat Options, Psychiatry 4 (2017) 117–126, https://doi.org/10.1007/s40501-017-0115-z.
[46] C.I. Nkanga, A. Fisch, M. Rad-Malekshahi, M.D. Romic, B. Kittel, T. Ullrich, J. Wang, R.W.M. Krause, S. Adler, T. Lammers, W.E. Hennink, F. Ramazani, Clinically established biodegradable long acting injectables: An industry perspective, Adv. Drug Deliv. Rev. 167 (2020) 19–46, https://doi.org/10.1016/j.addr.2020.11.008.
[47] P. Bannigan, Z. Bao, R.J. Hickman, M. Aldeghi, F. Häse, A. Aspuru-Guzik, C. Allen, Machine learning models to accelerate the design of polymeric long-acting injectables, Nat. Commun. 14 (2023) 35, https://doi.org/10.1038/s41467-022-35343-w.
[48] J. Deng, Z. Ye, W. Zheng, J. Chen, H. Gao, Z. Wu, G. Chan, Y. Wang, D. Cao, Y. Wang, S.-M.-Y. Lee, D. Ouyang, Machine learning in accelerating microsphere formulation development, Drug Deliv. Transl. Res. 13 (2023) 966–982, https://doi.org/10.1007/s13346-022-01253-z.
[49] A. Bhattacharya, Applied Machine Learning Explainability Techniques: Make ML models explainable and trustworthy for practical applications using LIME, SHAP, and more, Packt Publishing Ltd (2022).
[50] long-acting-injectables, (2023). https://github.com/aspuru-guzik-group/long-acting-injectables (accessed March 27, 2023).
[51] S.A. Damiati, S. Damiati, Microfluidic Synthesis of Indomethacin-Loaded PLGA Microparticles Optimized by Machine Learning, accessed February 6, 2023, Front. Mol. Biosci. 8 (2021), https://www.frontiersin.org/articles/10.3389/fmolb.2021.677547.
[52] F. Wang, M. Elbadawi, S.L. Tsilova, S. Gaisford, A.W. Basit, M. Parhizkar, Machine learning predicts electrospray particle size, Mater. Des. 219 (2022), 110735, https://doi.org/10.1016/j.matdes.2022.110735.
[53] J. Wang, N. Heshmati Aghda, J. Jiang, A. Mridula Habib, D. Ouyang, M. Maniruzzaman, 3D bioprinted microparticles: Optimizing loading efficiency using advanced DoE technique and machine learning modeling, Int. J. Pharm. 628 (2022), 122302, https://doi.org/10.1016/j.ijpharm.2022.122302.
[54] D. Gong, E. Ben-Akiva, A. Singh, H. Yamagata, S. Est-Witte, J.K. Shade, N.A. Trayanova, J.J. Green, Machine learning guided structure function predictions enable in silico nanoparticle screening for polymeric gene delivery, Acta Biomater. (2022) S1742-7061(22)00644–4. https://doi.org/10.1016/j.actbio.2022.09.072.
[55] R. Dalal, M. Leyden, F. Oviedo, T. Reineke, Polymer Design via SHAP and Bayesian Machine Learning Optimizes pDNA and CRISPR Ribonucleoprotein Deliver, (2022). https://doi.org/10.21203/rs.3.rs-1785891/v1.
[56] H. Gao, S. Kan, Z. Ye, Y. Feng, L. Jin, X. Zhang, J. Deng, G. Chan, Y. Hu, Y. Wang, D. Cao, Y. Ji, M. Liang, H. Li, D. Ouyang, Development of in silico methodology for siRNA lipid nanoparticle formulations, Chem. Eng. J. 442 (2022), 136310, https://doi.org/10.1016/j.cej.2022.136310.
[57] Z. Lin, W.-C. Chou, Y.-H. Cheng, C. He, N.A. Monteiro-Riviere, J.E. Riviere, Predicting Nanoparticle Delivery to Tumors Using Machine Learning and Artificial Intelligence Approaches, Int. J. Nanomed. 17 (2022) 1365–1379, https://doi.org/10.2147/IJN.S344208.
[58] W. Wang, S. Feng, Z. Ye, H. Gao, J. Lin, D. Ouyang, Prediction of lipid nanoparticles for mRNA vaccines by the machine learning algorithm, Acta Pharm. Sin. B 12 (2022) 2950–2962, https://doi.org/10.1016/j.apsb.2021.11.021.
[59] Y. Tang, J. Zhang, D. He, W. Miao, W. Liu, Y. Li, G. Lu, F. Wu, S. Wang, GANDA: A deep generative adversarial network conditionally generates intratumoral nanoparticles distribution pixels-to-pixels, J. Control. Release 336 (2021) 336–343, https://doi.org/10.1016/j.jconrel.2021.06.039.
[60] R.T. Stiepel, E.S. Pena, S.A. Ehrenzeller, M.D. Gallovic, L.M. Lifshits, C.J. Genito, E.M. Bachelder, K.M. Ainslie, A predictive mechanistic model of drug release from surface eroding polymeric nanoparticles, J. Control. Release 351 (2022) 883–895, https://doi.org/10.1016/j.jconrel.2022.09.067.
[61] M.J. Tamasi, R.A. Patel, C.H. Borca, S. Kosuri, H. Mugnier, R. Upadhya, N. S. Murthy, M.A. Webb, A.J. Gormley, Machine Learning on a Robotic Platform for the Design of Polymer-Protein Hybrids, Adv. Mater. 34 (2022) 2201809, https://doi.org/10.1002/adma.202201809.
[62] M.O. Jara, M. Landin, J.O. Morales, Screening of critical variables in fabricating polycaprolactone nanoparticles using Neuro Fuzzy Logic, Int. J. Pharm. 601 (2021), 120558, https://doi.org/10.1016/j.ijpharm.2021.120558.
[63] R. Rebollo, F. Oyoun, Y. Corvis, M.M. El-Hammadi, B. Saubamea, K. Andrieux, N. Mignet, K. Alhareth, Microfluidic Manufacturing of Liposomes: Development and Optimization by Design of Experiment and Machine Learning, ACS Appl. Mater. Interfaces 14 (2022) 39736–39745, https://doi.org/10.1021/acsami.2c06627.
[64] H. Maeda, K. Tsukigawa, J. Fang, A Retrospective 30 Years After Discovery of the Enhanced Permeability and Retention Effect of Solid Tumors: Next-Generation Chemotherapeutics and Photodynamic Therapy—Problems, Solutions, and Prospects, Microcirculation 23 (2016) 173–182, https://doi.org/10.1111/micc.12228.
[65] R. Sun, J. Xiang, Q. Zhou, Y. Piao, J. Tang, S. Shao, Z. Zhou, Y.H. Bae, Y. Shen, The tumor EPR effect for cancer drug delivery: Current status, limitations, and alternatives, Adv. Drug Deliv. Rev. 191 (2022), 114614, https://doi.org/10.1016/j.addr.2022.114614.
[66] Y. Zi, K. Yang, J. He, Z. Wu, J. Liu, W. Zhang, Strategies to enhance drug delivery to solid tumors by harnessing the EPR effects and alternative targeting mechanisms, Adv. Drug Deliv. Rev. 188 (2022), 114449, https://doi.org/10.1016/j.addr.2022.114449.
[67] L. Miao, C.M. Lin, L. Huang, Stromal barriers and strategies for the delivery of nanomedicine to desmoplastic tumors, J. Control. Release 219 (2015) 192–204, https://doi.org/10.1016/j.jconrel.2015.08.017.
[68] S. Karthika, M. Durgadevi, Generative Adversarial Network (GAN): a general review on different variants of GAN and applications, in: In: 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 1–8, https://doi.org/10.1109/ICCES51350.2021.9489160.
[69] S.N. Politis, P. Colombo, G. Colombo, D.M. Rekkas, Design of experiments (DoE) in pharmaceutical development, Drug Dev. Ind. Pharm. 43 (2017) 889–901, https://doi.org/10.1080/03639045.2017.1291672.
[70] J. Močkus, On bayesian methods for seeking the extremum, in: G.I. Marchuk (Ed.), Optimization Techniques IFIP Technical Conference Novosibirsk, Springer, Berlin, Heidelberg, 1975, pp. 400–404, https://doi.org/10.1007/3-540-07165-2_55.

[71] B. Letham, B. Karrer, G. Ottoni, E. Bakshy, Constrained Bayesian Optimization with Noisy Experiments, (2018). https://doi.org/10.48550/arXiv.1706.07094.
[72] B. Settles, Active Learning, Springer International Publishing, Cham (2012), https://doi.org/10.1007/978-3-031-01560-1.
[73] C.W. Coley, N.S. Eyke, K.F. Jensen, Autonomous Discovery in the Chemical Sciences Part I: Progress, Angew. Chem. Int. Ed. 59 (2020) 22858–22893, https://doi.org/10.1002/anie.201909987.
[74] C.W. Coley, N.S. Eyke, K.F. Jensen, Autonomous Discovery in the Chemical Sciences Part II: Outlook, Angew. Chem. Int. Ed. 59 (2020) 23414–23436, https://doi.org/10.1002/anie.201909989.
[75] F. Häse, L.M. Roch, A. Aspuru-Guzik, Next-Generation Experimentation with Self-Driving Laboratories, Trends in Chemistry. 1 (2019) 282–291, https://doi.org/10.1016/j.trechm.2019.02.007.
[76] M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C.T. Ser, J. Vestfrid, T.C. Wu, A. Aspuru-Guzik, Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab, Acc. Chem. Res. 55 (2022) 2454–2466, https://doi.org/10.1021/acs.accounts.2c00220.
[77] E. Stach, B. DeCost, A.G. Kusne, J. Hattrick-Simpers, K.A. Brown, K.G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, C.P. Gomes, J.M. Gregoire, A. Mehta, J. Montoya, E. Olivetti, C. Park, E. Rotenberg, S.K. Saikin, S. Smullin, V. Stanev, B. Maruyama, Autonomous experimentation systems for materials development: A community perspective, Matter. 4 (2021) 2702–2726, https://doi.org/10.1016/j.matt.2021.06.036.
[78] H.S. Stein, J.M. Gregoire, Progress and prospects for accelerating materials science with automated and autonomous workflows, Chem. Sci. 10 (2019) 9640–9649, https://doi.org/10.1039/C9SC03766G.
[79] B.J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J.I.M. Alvarado, J. M. Janey, R.P. Adams, A.G. Doyle, Bayesian reaction optimization as a tool for chemical synthesis, Nature 590 (2021) 89–96, https://doi.org/10.1038/s41586-021-03213-y.
[80] J.A.G. Torres, S.H. Lau, P. Anchuri, J.M. Stevens, J.E. Tabora, J. Li, A. Borovika, R.P. Adams, A.G. Doyle, A Multi-Objective Active Learning Platform and Web App for Reaction Optimization, J. Am. Chem. Soc. 144 (2022) 19999–20007, https://doi.org/10.1021/jacs.2c08592.
[81] B. Burger, P.M. Maffettone, V.V. Gusev, C.M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R.S. Sprick, A.I. Cooper, A mobile robotic chemist, Nature 583 (2020) 237–241, https://doi.org/10.1038/s41586-020-2442-2.
[82] A. Dave, J. Mitchell, K. Kandasamy, H. Wang, S. Burke, B. Paria, B. Póczos, J. Whitacre, V. Viswanathan, Autonomous Discovery of Battery Electrolytes with Robotic Experimentation and Machine Learning, Cell Reports Phys. Sci. 1 (2020), 100264, https://doi.org/10.1016/j.xcrp.2020.100264.
[83] B.P. MacLeod, F.G.L. Parlane, C.C. Rupnow, K.E. Dettelbach, M.S. Elliott, T.D. Morrissey, T.H. Haley, O. Proskurin, M.B. Rooney, N. Taherimakhsousi, D.J. Dvorak, H.N. Chiu, C.E.B. Waizenegger, K. Ocean, M. Mokhtari, C.P. Berlinguette, Advancing the Pareto front for thin-film materials using a self-driving laboratory, (2021). https://doi.org/10.48550/arXiv.2106.08899.
[84] H. Tao, T. Wu, S. Kheiri, M. Aldeghi, A. Aspuru-Guzik, E. Kumacheva, Self-Driving Platform for Metal Nanoparticle Synthesis: Combining Microfluidics and Machine Learning, Adv. Funct. Mater. 31 (2021) 2106725, https://doi.org/10.1002/adfm.202106725.
[85] K. Vaddi, H.T. Chiang, L.D. Pozzo, Autonomous retrosynthesis of gold nanoparticles *via* spectral shape matching, Digital, Discovery 1 (2022) 502–510, https://doi.org/10.1039/D2DD00025C.
[86] R.J. Hickman, P. Bannigan, Z. Bao, A. Aspuru-Guzik, C. Allen, Self-driving laboratories: A paradigm shift in nanomedicine development, Matter. (2023), https://doi.org/10.1016/j.matt.2023.02.007.
[87] M. Baker, 1,500 scientists lift the lid on reproducibility, Nature 533 (2016) 452–454, https://doi.org/10.1038/533452a.
[88] J. Zaslavsky, P. Bannigan, C. Allen, Re-envisioning the design of nanomedicines: harnessing automation and artificial intelligence, Expert Opin. Drug Deliv. 20 (2023) 241–257, https://doi.org/10.1080/17425247.2023.2167978.
[89] T. Lammers, G. Storm, Setting standards to promote progress in bio–nano science, Nat. Nanotechnol. 14 (2019) 626, https://doi.org/10.1038/s41565-019-0497-8.
[90] D.M. Anstine, O. Isayev, Generative Models as an Emerging Paradigm in the Chemical Sciences, J. Am. Chem. Soc. 145 (2023) 8736–8750, https://doi.org/10.1021/jacs.2c13467.
[91] D. Menon, R. Ranganathan, A Generative Approach to Materials Discovery, Design, and Optimization, ACS, Omega 7 (2022) 25958–25973, https://doi.org/10.1021/acsomega.2c03264.
[92] FDA no longer needs to require animal tests before human drug trials, (n.d.). https://www.science.org/content/article/fda-no-longer-needs-require-animal-tests-human-drug-trials (accessed April 30, 2023).
[93] A.H. Vo, T.R. Van Vleet, R.R. Gupta, M.J. Liguori, M.S. Rao, An Overview of Machine Learning and Big Data for Drug Toxicity Evaluation, Chem. Res. Toxicol. 33 (2020) 20–37, https://doi.org/10.1021/acs.chemrestox.9b00227.
[94] T.A. Meyer, C. Ramirez, M.J. Tamasi, A.J. Gormley, A User's Guide to Machine Learning for Polymeric Biomaterials, ACS Polym. Au. 3 (2023) 141–157, https://doi.org/10.1021/acspolymersau.2c00037.
[95] C. Chen, Z. Yaari, E. Apfelbaum, P. Grodzinski, Y. Shamay, D.A. Heller, Merging data curation and machine learning to improve nanomedicines, Adv. Drug Deliv. Rev. 183 (2022), 114172, https://doi.org/10.1016/j.addr.2022.114172.
[96] S. He, L.G. Leanse, Y. Feng, Artificial intelligence and machine learning assisted drug delivery for effective treatment of infectious diseases, Adv. Drug Deliv. Rev. 178 (2021), 113922, https://doi.org/10.1016/j.addr.2021.113922.
[97] W. Wang, Z. Ye, H. Gao, D. Ouyang, Computational pharmaceutics - A new paradigm of drug delivery, J. Control. Release 338 (2021) 119–136, https://doi.org/10.1016/j.jconrel.2021.08.030.
[98] R. Qureshi, D. Shaughnessy, K.A.R. Gill, K.A. Robinson, T. Li, E. Agai, Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? Syst. Rev. 12 (2023) 72, https://doi.org/10.1186/s13643-023-02243-z.
[99] C. O'Byrne, A. Abbas, E. Korot, P.A. Keane, Automated deep learning in ophthalmology: AI that can build AI, Curr. Opin. Ophthalmol. 32 (2021) 406, https://doi.org/10.1097/ICU.0000000000000779.
[100] S. Touma, F. Antaki, R. Duval, Development of a code-free machine learning model for the classification of cataract surgery phases, Sci. Rep. 12 (2022) 2398, https://doi.org/10.1038/s41598-022-06127-5.
[101] M. Huisman, J.N. van Rijn, A. Plaat, A survey of deep meta-learning, Artif. Intell. Rev. 54 (2021) 4483–4541, https://doi.org/10.1007/s10462-021-10004-4.
[102] J. Vanschoren, Meta-Learning: A Survey, (2018). https://doi.org/10.48550/arXiv.1810.03548.
[103] W.-C. Chou, Z. Lin, Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling, Toxicological Sciences : An Official Journal of the Society of, Toxicology 191 (2023) 1–14, https://doi.org/10.1093/toxsci/kfac101.
[104] A.V. Singh, M.H.D. Ansari, D. Rosenkranz, R.S. Maharjan, F.L. Kriegel, K. Gandhi, A. Kanase, R. Singh, P. Laux, A. Luch, Artificial Intelligence and Machine Learning in Computational Nanotoxicology: Unlocking and Empowering Nanomedicine, Adv. Healthc. Mater. 9 (2020) 1901862, https://doi.org/10.1002/adhm.201901862.