Drug Delivery and Translational Research
https://doi.org/10.1007/s13346-022-01253-z

ORIGINAL ARTICLE

Machine learning in accelerating microsphere formulation development

Jiayin Deng[1] · Zhuyifan Ye[1] · Wenwen Zheng[2] · Jian Chen[3] · Haoshi Gao[1,4] · Zheng Wu[1] · Ging Chan[1,5] · Yongjun Wang[6] · Dongsheng Cao[7] · Yanqing Wang[3] · Simon Ming-Yuen Lee[1,5] · Defang Ouyang[1,5]

Accepted: 22 October 2022

Abstract

Microspheres have gained much attention from pharmaceutical and medical industry due to the excellent biodegradable and long controlled-release characteristics. However, the drug release behavior of microspheres is influenced by complicated formulation and manufacturing factors. The traditional formulation development of microspheres is intractable and inefficient by the experimentally trial-and-error methods. This research aims to build a prediction model to accelerate microspheres product development for small-molecule drugs by machine learning (ML) techniques. Two hundred eighty-six microsphere formulations with small-molecule drugs were collected from the publications and pharmaceutical company, including the dissolution temperature at both 37 ℃ and 45 ℃. After the comparison of fourteen ML approaches, the consensus model achieved accurate predictions for the validation set at 37 ℃ and 45 ℃ ($R^2 = 0.880$ vs. $R^2 = 0.958$), indicating the good performance to predict the in vitro drug release profiles at both 37 ℃ and 45 ℃. Meanwhile, the models revealed the feature importance of formulations, which offered meaningful insights to the microspheres development. Experiments of microsphere formulations further validated the accuracy of the consensus model. Furthermore, molecular dynamics (MD) simulation provided a microscopic view of the preparation process of microspheres. In conclusion, the prediction model of microsphere formulations for small-molecule drugs was successfully built with high accuracy, which is able to accelerate microspheres product development and promote the quality control of microspheres for the pharmaceutical industry.

Keywords Microspheres · Drug release · Machine learning · Molecular dynamics simulation

Introduction

Microsphere products have gained broad attention from the pharmaceutical and medical industry due to the advantage of biodegradability, biocompatibility, and tunability [1]. With the large amount of active pharmaceutical ingredients (APIs) encapsulated in biodegradable polymer matrix, such as polylactic acid (PLA) and poly (lactic-co-glycolic acid) (PLGA), microspheres can achieve extended release for weeks or even months, so as to dramatically reduce administration frequency [2, 3]. This extraordinary advantage significantly relieves patient suffering and effectively improves treatment adherence, especially

Jiayin Deng, Zhuyifan Ye, Wenwen Zheng, and Jian Chen equally contributed to this paper.

✉ Yanqing Wang
wangyanqing@livzon.cn

✉ Simon Ming-Yuen Lee
simonlee@um.edu.mo

✉ Defang Ouyang
defangouyang@um.edu.mo

[1] State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences (ICMS), University of Macau, Macau, China

[2] Department of Clinical Laboratory, The Sixth Affiliated Hospital of Sun Yat-Sen University, Guangzhou, China

[3] Zhuhai Livzon Microsphere Technology Co., Ltd, Zhuhai, China

[4] Institute of Applied Physics and Materials Engineering, University of Macau, Macau, China

[5] Faculty of Health Sciences, University of Macau, Macau, China

[6] Wuya College of Innovation, Shenyang Pharmaceutical University, Shenyang, China

[7] Xiangya School of Pharmaceutical Sciences, Central South University, Changsha, China

Published online: 01 December 2022

for some non-alternative demands, e.g., psychiatric medication, diabetes, and cancer [4, 5]. However, since the advent of microspheres manufacture technology 40 years ago, only several microsphere formulations have been successfully developed from bench to bedside (Table 1). Their application is hindered by the complicated formulation and preparation process.

The microspheres depend on numerous formulation and process factors, including API and carrier material properties, solvent, temperature, and stirring speed et al. [6, 7]. The release mechanism of microspheres includes drug diffusion, polymer degradation, and erosion. For small hydrophobic drugs, they can diffuse through the polymer matrix and also be released by erosion, whereas the peptide drugs mainly diffuse through the pores and the diffusion rate is determined by the degree of erosion [8]. In order to obtain the desired release profiles, numerous formulation combinations should be considered. Moreover, the formulation and process parameters need significant optimization from the laboratory-scale test to large-scale manufacture. Currently, the in vitro release testing at 37 ℃ applied in microsphere formulation screening may take up to several months. In order to reduce release time, the FDA regulatory science program [9] and EUFEPS workshop [10] have mentioned the elevated temperature method, which significantly benefits the preliminarily formulation space searching. However, given the high-dimensional space of microsphere formulation to explore, the traditional R&D process on the basis of trial-and-error experimental approaches could take years to go, which is still doomed to be laborious, material and time-consuming. Moreover, even some commercial products are still not the best fit for purpose, further far from the optimal setting [11]. Thus, the pharmaceutical industry requires a more efficient approach to accelerate the microsphere formulation development.

Recently, machine learning (ML) techniques have been integrated in different aspects of our life, such as automated driving, medical diagnosis, and drug discovery and development [12]. The unique advantage of ML is to explore the implicit knowledge and make predictions for complex issues, which could reduce abundant experiment work, and efficiently promote development process [12]. In addition to drug discovery, ML has also been introduced in pharmaceutical formulation development, such as the prediction of dissolution profile and physical stability of solid dispersions [13, 14], size and polydispersity index of drug nanocrystals [15], lipid nanoparticle for mRNA vaccine [16], self-emulsifying drug delivery systems [17], and binary cyclodextrin complexes [18]. ML approaches were also applied in microsphere formulation. Szlękand et al. built artificial neural network (ANN) models with 68 formulation data to fit the relationship between formulation and protein drug dissolution [19, 20]. Rodrigues de Azevedo et al. analyzed the impact of physicochemical factors on the initial drug release from PLGA(-PEG) systems by using partial least squares regression and decision tree (DT) on 152 experimental data [21]. To investigate potential combinations of polymer and drug, Bannigan et al. applied 4 ANN-structures on 181 formulation data of both long-acting injectable implants and microspheres [22]. These examples showed the feasibility of ML in

Table 1 Summary of drug-loaded microsphere products approved by the U.S. Food and Drug Administration (FDA)

Drug product	Active ingredient	Route of administration	Time range	Indications	Approval date
Decapeptyl	Triptorelin acetate	Intramuscular injection	1, 3 months	Prostate cancer; endometriosis; uterine fibroids	1986
Lurpon/Enantone	Leuprolide acetate	Intramuscular injection	1, 3, 4, 6 months	Prostate cancer; breast cancer	1995
Risperdal Consta	Risperidone	Intramuscular injection	2 weeks	Schizophrenia; bipolar I disorder	1997
Sandotatin LAR	Octreotide acetate	Subcutaneous injection	4 weeks	Severe diarrhea associated with metastatic carcinoid tumors or VIP-secreting tumors	1998
Nutropin	Somatropin	Intramuscular injection	1 month	Growth failure	1999
Trelstar	Triptorelin pamoate	Intramuscular injection	4 weeks	Advanced prostate cancer	2000
Arestin	Minocycline	Peridontal injection	1 week	Periodontitis	2001
Vivitrol	Naltrexone	Intramuscular injection	2 weeks	Alcohol dependence	2006
Bydureon	Exenatide	Intramuscular injection	1 week	Type 2 diabetes	2012
Signifor LAR	Pasireotide pamoate	Intramuscular injection	1 month	Acromegaly	2014
Zilretta	Triamcinolone acetonide	Intrarticular injection	3 months	Osteoarthritis pain of the knee	2017
Bydureon BCise	Exenatide	Intramuscular injection	1 week	Type 2 diabetes	2017
Triptodur kit	Triptorelin pamoate	Intramuscular injection	24 weeks	Central precocious puberty	2017

microsphere formulations. However, previous studies only did the preliminarily trial with limited data from the laboratory scale. A practical model for the industrial scale is still under urgent demand.

This research aimed to accelerate microspheres product development for the pharmaceutical industry by ML techniques. Two hundred eighty-six microsphere formulations with small-molecule drugs were collected from both publications and the pharmaceutical company, covering the dissolution temperature at 37 ℃ and 45 ℃. Fourteen ML algorithms were compared for predicting the in vitro drug release profiles and subsequently validated by the experiments. Furthermore, molecular dynamic simulation was applied for investigating molecular mechanism of microsphere formation.

Methods

Dataset construction

Two hundred eighty-six PLGA microsphere formulations for small-molecule drugs were obtained from publications (32 formulations) and Livzon Microsphere Ltd. (254 formulations). This dataset included 12 small-molecule drugs and 3182 release time points. Publications were collected from the SCOPUS database, where PLGA microspheres for small-molecule drugs mentioned were prepared by solvent evaporation/extraction methods. The words or phrases, "microspheres", "microparticle", "PLGA", "poly (lactic-co-glycolic acid)", and "sustained release" were used as keywords for searching articles in the database. Since release profiles at high temperature correlated well with those at 37 °C and were in large amounts, both the data of 37 °C and 45 °C were included in this dataset. Information about drug release behavior was collected and described as 35 features, which comprehensively considered the formulation parameters, processing conditions, microspheres characteristics, and in vitro dissolution conditions. Eleven physicochemical properties of the drugs were used as input features, which included molecular weight, XLogP3, hydrogen bond donor count, hydrogen bond acceptor count, rotatable bond count, topological polar surface area, heavy atom count, complexity, melting point, logS, and logP. The input features were summarized in Table 2, and cumulative drug release was the predictive target.

Before establishing prediction models on the dataset, data pre-processing was performed. Drugs were characterized as 11 physicochemical properties derived from the Pubchem database, and where logP and logS not given were calculated by ALOGPS [23]. For categorical features, each label was assigned as an integer. Extremely uneven distributed features were processed by taking the logarithm of the features to base 10. Missing values were dealt with either the statistical mean or the mode based on pharmaceutical knowledge. For neural networks, support vector machine (SVM), and k-nearest neighbors (kNN), the features were first standardized before being fed into the models.

Dataset splitting strategy

In general, the whole dataset was evenly split into the training set (70%) and the validation set (30%) by the group random sampling method. The training set was employed for training models and the validation set was used for tuning hyperparameters. Since a drug dissolution profile consists of multiple time points and corresponds to one formulation, all release points of the same profile would share the same formulation information. Here, all release points were firstly grouped into formulations and numbered by formulation index. Secondly, the dataset was then divided into training set and validation set by formulation index. The group-based split approach guaranteed that the whole release profile was divided into one subset and additionally, the similar training set and validation set distribution were obtained (Fig. 1).

Group 10-tenfold cross validation (CV) was applied for model performance evaluation, which ensured that the release points from the same formulation were not divided

Table 2 Summary of input features in the dataset

Categories	Input features
Formulation	Physicochemical properties of drug (molecular weight, XLogP3, hydrogen bond donor count, hydrogen bond acceptor count, rotatable bond count, topological polar surface area, heavy atom count, complexity, melting point, logS, logP), concentration of drug, PLGA and PVA, and type and volume of solvent.
Processing condition	Emulsification device, solvent elimination method, temperature during emulsification and solvent elimination process, and stirring speed.
Microspheres characteristics	Drug loading, encapsulation efficiency, particle size, and particle surface type.
In vitro dissolution condition	Dissolution temperature, type and pH of dissolution medium, surfactant in dissolution medium, dose, and time.

PLGA poly (lactic-co-glycolic acid), *PVA* polyvinyl alcohol

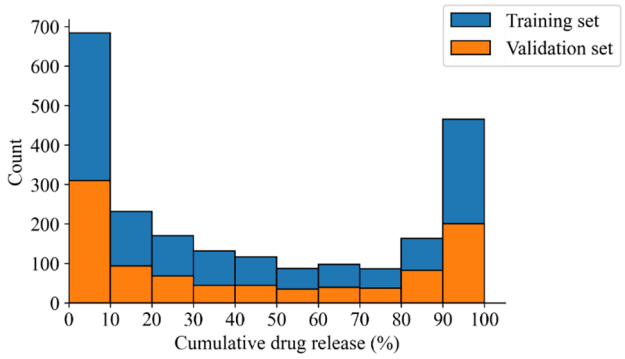

Fig. 1 Distribution of cumulative drug release in the training set and the validation set

into different subsets. In group 10-fold CV, the dataset was split into 10 subsets based on formulation, and for each fold, 9 subsets were served as the training data and the rest was for evaluating performance, this process iterated for 10 times.

Evaluation criteria

Mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2) were applied for evaluating the model performance. They are commonly used in ML, where MAE, MSE, and RMSE evaluate the distance between individual true and predicted values, and R^2 evaluates the correlation between them. The evaluation criteria were calculated by the following equations:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|,$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2,$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{n}},$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2},$$

where n was the number of samples, y_i was the real value for time point i, and $\hat{y}_i$ was the prediction, and $\bar{y}$ was the average value.

Furthermore, in experimental validation, to evaluate the model performance on the whole profile, we specified that predictions with an average prediction error no more than 10% for all time points are identified as successful. This was drawn from the requirement in pharmaceutics that a 10% error is considered acceptable [24–26]. Similar factor (f_2) was introduced to provide a pharmaceutical indication:

$$f_2 = 50 \times \log\left\{100 \times \left[1 + \frac{\sum_{i=1}^{n'}(y_i - \hat{y}_i)^2}{n'}\right]^{-0.5}\right\},$$

where n' was the number of time points in a curve, y_i was the experimental value for time point i, and $\hat{y}_i$ was the predicted value. According to the FDA recommendation, only the first time point after 85% dissolution is used in the f_2 calculation [27]. This is a more rigorous approach than averaging the variance of all time points, because the points after 85% dissolution are generally plateau with small variance and high similarity. Therefore, we only added the first time point after 85% release in the f_2 calculation for the good pharmaceutical evaluation.

Machine learning models establishment

Fourteen ML approaches were used to develop regression models for predicting the microspheres in vitro cumulative drug release, including XGBoost, random forest (RF), LightGBM, residual neural network (ResNet), deep neural network (DNN), DT, artificial neural network (ANN), artificial neural network-multilayer perceptron (ANN-MLP), kNN, SVM, ridge regression, partial least squares regression (PLS), lasso regression, and multiple linear regression (MLR). These algorithms covered conventional kernel-based and gradient-based ML approaches previously dominated in pharmaceutics, shallow models and deep learning, single DT and tree-based ensemble learning, and linear and non-linear models. The neural networks, the LightGBM model, the XGBoost model, and other ML models were established by the TensorFlow package (version 2.3.1) [28, 29], the LightGBM package (version 3.2.1), the XGBoost package (version 1.5.0), and the Scikit-Learn package (version 0.24.2) in Python, respectively.

Furthermore, the optimal hyperparameters were searched on the validation set by three searching methods, grid search, random search, and manual search. Specifically, DT, kNN, and PLS employed grid search, XGBoost, LightGBM, RF, and SVM utilized random search. Due to the large hyperparameter space of neural networks, grid search and random search may lead to overfitting or underfitting problems, for which manual search was conducted for neural networks. Lasso regression, ridge regression, and MLR had no hyperparameters. The hyperparameter

configurations of ML algorithms were shown in the Table 3, and the values were given in brackets.

Consensus modeling

In order to achieve better prediction capability and robustness than individual models, we have further constructed the consensus model. Individual models can result in different model uncertainty due to different hypothesis adopted. Consensus model is able to reduce the prediction error by integrating single models. Since the consensus model has combined different hypothesis, it can figure out the relationship between input features and output variable more comprehensively than one individual model [30]. By averaging the results of those well-performed models, the consensus model is expected to have good and also robust model performance.

Here, after comparing the performance of single models, those with comparably good performance were selected as sub-models to ensure the consensus model possesses high predictive capability. All sub-models were treated as equal contribution, in which their predictions were averaged as the output of consensus model.

Experimental validation

To further validate the accuracy of prediction models, 20 external microsphere formulations were prepared to further validate the accuracy of ML models. Risperidone, developed as the first microspheres product for small-molecule drug, was selected as the model drug for microspheres preparation by emulsion solvent evaporation method. The combinations of formulation and process parameters were listed in Table S1 in supplementary material, and the preparation process and characterization method were described as follows.

Materials

Risperidone was purchased from Jiangsu Nhwa Pharmaceutical Co., Ltd. (Jiangsu, China). PLGA, with a lactide/glycolide ratio of 75:25, was provided by Evonik Industries (Shanghai, China). Polyvinyl alcohol (PVA) was purchased from Mitsubishi Chemical (Tokyo, Japan). Dichloromethane (DCM) was obtained from Sino Pharm (Beijing, China). Polysorbate 20 was provided by Nanjing

Table 3 Hyperparameter configurations of machine learning approaches

Machine learning algorithm	Hyperparameter configurations
XGBoost	The learning rate (0.00788); the number of trees (1188); the subsample ratio (0.817); the subsample ratio of columns (0.869); the maximum depth of the trees (4)
RF	The number of the maximum features (24); the number of the trees (555); the maximum depth of the trees (23); the minimum number of the samples used to split (10); the minimum number of samples in a child leaf (5)
LightGBM	The learning rate (0.00315); the number of trees (1104); the subsample ratio (0.302); the subsample ratio of columns (0.853); maximum tree leaves for base learners (50)
ResNet	The number of hidden layers (20); the number of neurons in each of the hidden layers (1024, 256, 128, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64, 64); the learning rate (0.01); the optimization algorithm (Adam with β1, β2 of 0.9, 0.999); the learning rate decay (0.01); the epoch size (151); the l2 regularization coefficient lambda (0.16); the batch size (500)
DNN	The number of hidden layers (10); the number of neurons in each of the hidden layers (1024, 256, 128, 64, 64, 64, 64, 64, 64, 64); the learning rate (0.001); the optimization algorithm (Adam with β1, β2 of 0.9, 0.999); the learning rate decay (0.02); the epoch size (120); the l2 regularization coefficient lambda (0.2); the batch size (500)
DT	The maximum depth of the tree (15); the minimum number of the samples used to split (14); the minimum number of samples in a leaf (6)
kNN	The number of neighbors (4); weight function used in prediction (the standard Euclidean distance with uniform weights)
SVM	The penalty parameter C (426.9); the γ (0.14)
ANN	The number of hidden layers (4); the number of neurons in each of the hidden layers (1024, 256, 128, 64); the learning rate (0.001); the optimization algorithm (Adam with β1, β2 of 0.9, 0.999); the learning rate decay (0); the epoch size (600); the l2 regularization coefficient lambda (0); the batch size (500)
ANN-MLP	The number of hidden layers (1); the number of neurons in each of the hidden layers (1024); the learning rate (0.003); the optimization algorithm (Adam with β1, β2 of 0.9, 0.999); the learning rate decay (0); the epoch size (200); the l2 regularization coefficient lambda (0); the batch size (500)
PLS	The number of components (13)
Lasso regression	\
Ridge regression	\
MLR	\

";" was used for separating different hyperparameters; "," was for different components of a hyperparameter; "\" meant no hyperparameter

Well Pharmaceutical Group Co., Ltd. (Nanjing, China). The other chemicals or solvents were of reagent or analytical grade.

Microspheres preparation

Risperidone-loaded microspheres were prepared by the emulsion solvent evaporation method as previously described [31]. Briefly, risperidone and PLGA were dissolved completely in DCM. The organic solution was emulsified in aqueous solution containing PVA by stirring using a high-shear emulsifier. Emulsion was then stirred at room temperature to evaporate DCM and allowed microspheres formation. Solid microspheres were harvested by microporous membrane. Collected microspheres were then washed and freeze-dried. The batch size was ranged from 1.8 to 24 g.

$$\text{Entrapment Efficiency}(\%) = \frac{\text{Weight of encapsulated drug}}{\text{Weight of drug used in preparation}} \times 100\%$$

Scanning electron microscope (SEM)

Scanning electron microscope (Pro, Phenom World, Eindhoven, The Netherlands) was employed to investigate the microscopic morphology of microspheres. Risperidone-loaded microspheres were placed and mounted on the sample table. A thin gold layer was coated on the microspheres by ion sputtering instrument (LJ-16, Yulong Technology, Beijing, China). Conductive microspheres were then observed by SEM.

Size and size distribution

After uniform mixing, 30–60 mg microspheres were sampling. The samples were dispersed in 5 ml ultrapure water by sonicating. The size and size distribution of microspheres were estimated by a laser size analyzer (MS2000, Malvern Panalytical, Malvern, UK). Particle size was described by D_{10}, D_{50}, and D_{90}, and size distribution was evaluated by SPAN. SPAN is defined as $(D_{90} - D_{10})/D_{50}$, where D_{10}, D_{50}, and D_{90} were the size that 10%, 50%, and 90% of microspheres were less than.

Entrapment efficiency and drug loading

A total of 20–40 mg of microspheres were sampling after uniform mixing, and the samples were dissolved in 20 ml ammonium acetate buffer-acetonitrile solution to extract risperidone completely. The aqueous solution was filtered with 0.22 μm microporous membrane, and then quantified by Agilent 1260 type HPLC equipped with Agilent G1314F 1260 VWD detector and Thermo Syncronis C18 chromatographic column (150 mm × 4.6 mm, 5 μm). The mobile phase formulated at 0.065 mol/L ammonium acetate buffer: acetonitrile (70:30) flowed at 1.0 mL/min and column temperature was 40 °C. Detection wavelength here was 275 nm. Drug loading referred to the ratio of the encapsulated drug in microspheres and microspheres weight. According to drug loading, entrapment efficiency was calculated as the ratio of content of drug encapsulated and drug content used in microspheres preparation.

$$\text{Drug Loading}(\%) = \frac{\text{Weight of encapsulated drug}}{\text{Weight of microspheres}} \times 100\%$$

In vitro release testing

In brief, 5 mg microspheres was dispersed in 50 ml HEPES buffer solution (pH 7.4) containing polysorbate 20 in Erlenmeyer flask at incubator. If not specified, the dissolution temperature was set to 45 ℃. At pre-determined time, 1 ml solution was sampled, and equal fresh medium was added. Sampling solution was filtered with 0.22 μm microporous membrane before quantified by HPLC. Concentration of risperidone at each sampling time was measured, and cumulative release was determined.

Molecular dynamics simulation

In this part, the formation of microspheres was molecular dynamically visualized and analyzed to investigate the microscopic interactions between drugs and excipients which the experiments cannot reach. Firstly, all three-dimensional molecules were constructed by Discovery Studio Visualizer. The relevant topology and partial charge were described by applying General AMBER force field (GAFF) [32] and AM1-BCC method [33]. In consistence with proportions in microspheres preparation, 21 risperidone molecules, 4 PLGA fragment with eight units, 1725 DCM molecules, and 20 PVA fragment with six units were used in initial model preparation. Since the preparation process involved the emulsification of oil phase in water phase, two initial models, oil phase model and water phase model, were built respectively. Risperidone, PLGA, DCM, and PVA molecules were randomly dispersed in two different vacuum

boxes by Packmol software [34]. Initial models were loaded in GAFF by using LEAP module. Specially, 10-Å thickness TIP3P water solution was added to initial structure of PVAs.

After the preparation, four simulations were performed by using AMBER18 software. Firstly, the oil phase model and water phase model were respectively conducted an 1 ns MD simulation to approximate the dispersion process of PLGAs and risperidone in DCMs, and PVAs in water solution. Subsequently, to simulate the oil-in-water emulsification process, the simulated oil phase model was immersed in simulated water phase model to construct O/W system by using visual molecular dynamics (VMD) software [35]. The constructed system was then performed 10 ns simulation. Finally, the DCMs of the last frame of the simulated O/W system extracted were removed. This initial structure constructed was subsequently conducted 180 ns MD simulation to simulate the solvent evaporation and microspheres solidification processes.

Each of the simulation procedures mentioned above included the following steps. Briefly, energy minimization was performed to remove the irrational contacts of atoms. Subsequently, the system was heated from 0 to 300 K in 40 ps and held for 160 ps. The coupling algorithm of Langevin thermostat [36] and Berendsen barostat [37] were used to hold the temperature at 300 K and pressure at 1 atm, respectively. The MD simulation was performed at 0.002 ps per step, and trajectory file was saved every 10 ps. The simulation process was visualized by VMD software and CPPTRAJ tool was used in subsequent analysis. Root-mean-square deviation (RMSD), radius of gyration (Rg), solvent-accessible surface area (SASA), and the number of hydrogen bonds were analyzed for evaluation.

Results

Data distribution in the dataset

In this study, a dataset for microspheres was collected from the publication and the pharmaceutical company, consisting of 12 small-molecule drugs, 286 microsphere formulations, and 3182 release time points, where both 37 ℃ and 45 ℃ dissolution temperatures were considered. Specifically, as shown in Fig. 2A, nearly 90% of data came from the pharmaceutical company. In terms of dissolution temperature, more than two thirds of the data were at 45 ℃ and less than one third were at 37 ℃ (Fig. 2B). Drug release testing at 37 °C typically required a release time of about 35 days, which was reduced to a maximum of 15 days at 45 °C to achieve 90% drug release (Fig. 2C). The elevated temperature method greatly reduced the release time and facilitated the accumulation of a substantial amount of data [10].

The data distribution of microspheres characteristics was shown in Fig. 3. For microspheres, particle size, drug loading, and encapsulation efficiency were three key quality attributes. Encapsulation efficiency and drug loading were important indicators for preparation process and production cost, especially significant for some expensive drug formulations [38]. In the dataset, over 76.9% of the microspheres exhibited a high encapsulation efficiency over 90% (Fig. 3A). Approximately 74.8% of the microspheres achieved a drug loading higher than 30% (Fig. 3B), which led to a reduction in the use of microspheres. Besides, particle size was an essential attribute for patient adherence. The large microparticles required injection with a large diameter needle, which would cause great pain during administration [39–41]. Figure 3C showed that the particle size ranged from 0 to 125 μm, where 88.8% of the microspheres exhibited a size less than 100 μm.

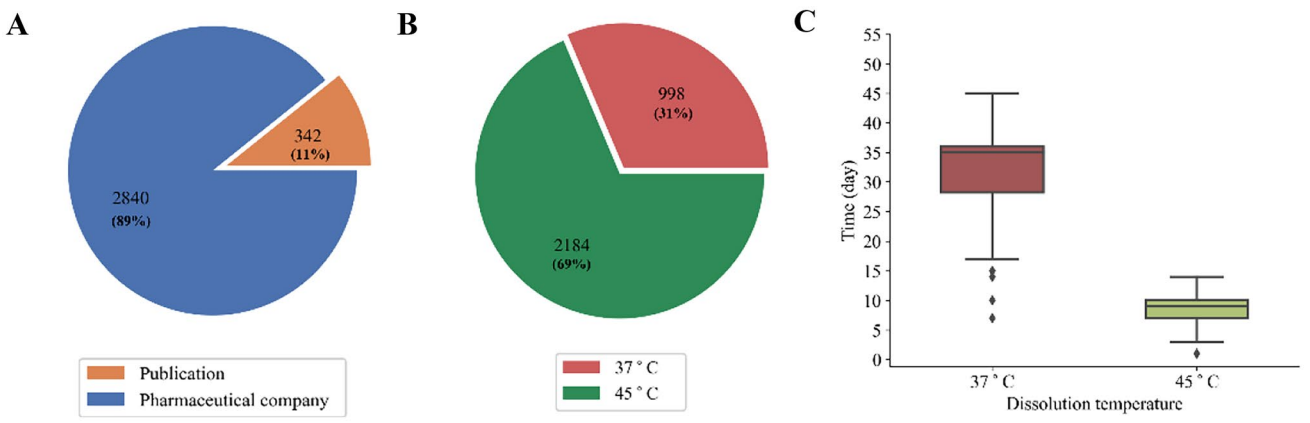

Fig. 2 Percentage of the data from the publication and pharmaceutical company **A**, percentage of the data at 37 °C and 45 °C **B**, distribution of the release time when reaching 90% cumulative drug release at 37 °C and 45 °C **C** in the dataset

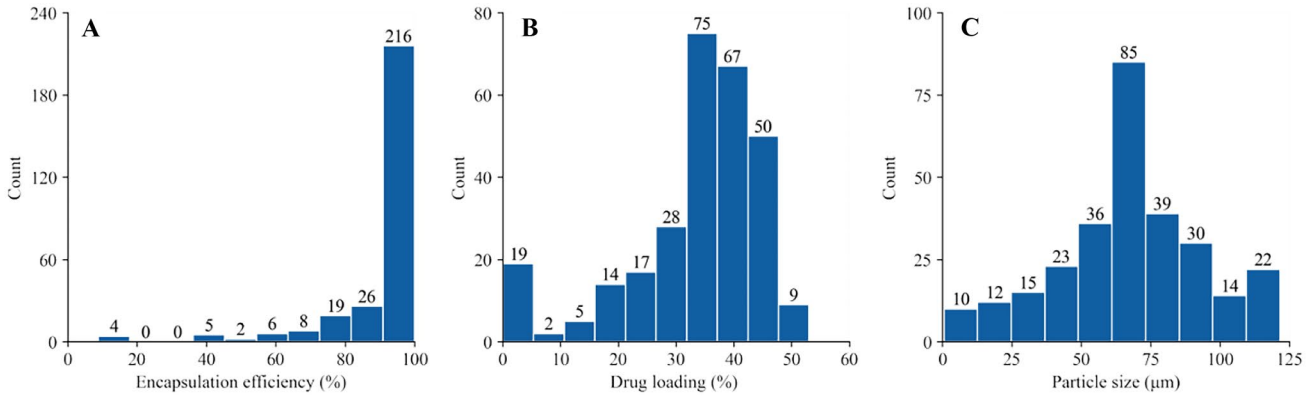

Fig. 3 The distribution of encapsulation efficiency **A**, drug loading **B**, and particle size **C** in the dataset

Comparison of various machine learning models

Fourteen ML algorithms were applied and compared to obtain the optimal models for predicting in vitro drug release, and a group 10-fold CV was used to evaluate the model performance. The model accuracy was then described by four metrics, MAE, MSE, RMSE, and R^2. The group 10-fold CV result in Table 4 showed that XGBoost obtained the lowest MSE, RMSE, and highest R^2 while RF presented the lowest MAE. Specifically, tree-based ensemble models, including XGBoost, RF, and LightGBM, obtained the best model performance and highest model robustness. The model performance of neural network architectures was slightly worse than that of tree-based ensemble models, but better than the other algorithms in the group 10-fold CV results. Particularly, ResNet showed better predictions than tree-based ensemble models on the validation set, as reflected by lower validation MAE, MSE, RMSE, and higher R^2, but was slightly worse in group 10-fold CV. This is because ResNet is sensitive to hyperparameters and was not optimized for each fold in CV, resulting in lower model performance. The linear regression algorithms obtained the largest prediction error, where MLR gave the highest MAE, MSE, RMSE, and the lowest R^2, indicating that linear approaches were not suitable for in vitro drug release prediction. kNN and SVM presented moderate performance, giving an R^2 of 0.863 and 0.833 in the group 10-fold CV, showing a slightly worse predictive ability than neural network-based algorithms, but better than linear approaches.

As XGBoost, RF, LightGBM, and ResNet showed good results on this dataset, we further analyzed their predicted cumulative release. The predicted drug release was compared with the experimental values and was shown in the scatter plots (Fig. 4).

Overall, the models provided accurate predictions on this dataset, with most of the scatters falling around the diagonals. Specifically, on the validation set, the overall predictions for ResNet were accurate but a few points suffered from prediction bias. For the tree-based ensemble models, most of the samples were in the less than 10% error region, as shown in Fig. 4. The performance exhibit a little differences between 0–30% (the initial release period) and 30–80%, and between 30–80% and 80–100% (the complete release period). We divided the release profile into three release periods, and further analyzed the differences. Within the range 0–30%, XGBoost, RF, and LightGBM algorithms overestimated the drug release in some formulations, which might be caused by the complicated multi-modal non-Gaussian distribution existing in the data range. We analyzed the data points and formulations in the initial release period and found that 41 of the total 286 formulations had initial burst drug release. The initial burst drug release indicated over 30% fast drug dissolution in the first day, which resulted in a distinct release pattern from other formulations. Therefore, the complicated data distribution is a challenge for ML modeling and making robust predictions in the initial release period. Although the burst release of the "fail formulations" may cause the serum concentration much higher than target therapeutic window and potentially lead to toxicity [38, 42, 43], these "failed formulations" would still help to identify and distinguish the patterns of risky formulations in the modeling. On the other hand, the models sometimes may make underestimation at the complete release stage, which may be caused by the different release rates under 37 °C and 45 °C conditions.

Predicting in vitro drug release profiles by the consensus model

Since four individual models based on different assumptions all provided good predictions, averaging their predicted results could expect to obtain good model performance and further improve robustness. Here, all four individual models were treated with equal contribution and referred to as a

Table 4 Results for different machine learning models on the training set and the validation set and the group 10-fold cross-validation

	MAE			MSE			RMSE			R^2		
	Training set	**Validation set**	**Group 10-fold CV**	**Training set**	**Validation set**	**Group 10-fold CV**	**Training set**	**Validation set**	**Group 10-fold CV**	**Training set**	**Validation set**	**Group 10-fold CV**
Consensus model	3.6	5.6	\	34.6	**81.9**	\	5.9	**9.0**	\	0.973	**0.937**	\
XGBoost	4.2	6.6	6.9 ± 1.0	42.5	114.6	**115.7 ± 34.2**	6.5	10.7	**10.6 ± 1.6**	0.970	0.921	**0.918 ± 0.024**
RF	4.0	6.4	**6.6 ± 1.1**	43.8	107.4	118.6 ± 37.4	6.6	10.4	10.8 ± 1.7	0.969	0.926	0.916 ± 0.026
LightGBM	4.6	6.9	7.4 ± 1.0	42.6	108.2	120.7 ± 34.3	6.5	10.4	10.9 ± 1.6	0.970	0.926	0.915 ± 0.024
ResNet	4.0	**5.5**	7.0 ± 0.9	57.2	102.0	141.1 ± 56.7	7.6	10.1	11.7 ± 2.2	0.960	0.930	0.903 ± 0.036
DNN	4.3	6.5	7.6 ± 1.5	52.7	124.6	150.4 ± 43.3	7.3	11.2	12.1 ± 1.7	0.963	0.915	0.889 ± 0.050
ANN	6.8	7.2	9.5 ± 1.4	97.4	116.9	178.6 ± 70.3	9.9	10.8	13.1 ± 2.4	0.931	0.920	0.880 ± 0.041
DT	4.7	7.8	7.9 ± 1.1	66.1	183.9	179.3 ± 49.9	8.1	13.6	13.2 ± 1.9	0.953	0.874	0.873 ± 0.036
ANN-MLP	7.3	7.7	9.9 ± 1.1	112.5	128.8	183.8 ± 47.3	10.6	11.3	13.5 ± 1.7	0.920	0.912	0.868 ± 0.044
kNN	4.4	8.2	8.9 ± 1.2	57.6	187.7	194.4 ± 52.0	7.6	13.7	13.8 ± 1.8	0.959	0.871	0.863 ± 0.035
SVM	5.0	8.6	10.6 ± 1.1	69.8	168.6	238.0 ± 52.0	8.4	13.0	15.3 ± 1.8	0.951	0.884	0.833 ± 0.035
PLS	18.8	17.2	22.2 ± 1.2	532.1	462.9	660.6 ± 63.9	23.1	21.5	25.7 ± 1.2	0.623	0.682	0.534 ± 0.051
Ridge regression	18.7	17.4	22.5 ± 1.4	528.4	469.7	680.5 ± 85.7	23.0	21.7	26.0 ± 1.6	0.626	0.678	0.520 ± 0.063
Lasso regression	19.5	18.3	22.4 ± 1.3	578.8	527.6	689.4 ± 74.2	24.1	23.0	26.2 ± 1.4	0.590	0.638	0.513 ± 0.058
MLR	18.7	17.6	— —	526.3	496.4	— —	22.9	22.3	— —	0.627	0.659	— —

Bold represent the best performance; "\" meant the consensus model had no CV result; "— —" meant the MLR model showed a large prediction error on the group 10-fold cross-validation, in which the mean and standard deviation values of MAE, MSE, RMSE, R^2 were $1.35\times10^{10}\pm2.18\times10^{10}$, $1.49\times10^{22}\pm2.91\times10^{22}$, $6.43\times10^{10}\pm1.04\times10^{11}$, and $-1.03\times10^{19}\pm2.01\times10^{19}$, respectively

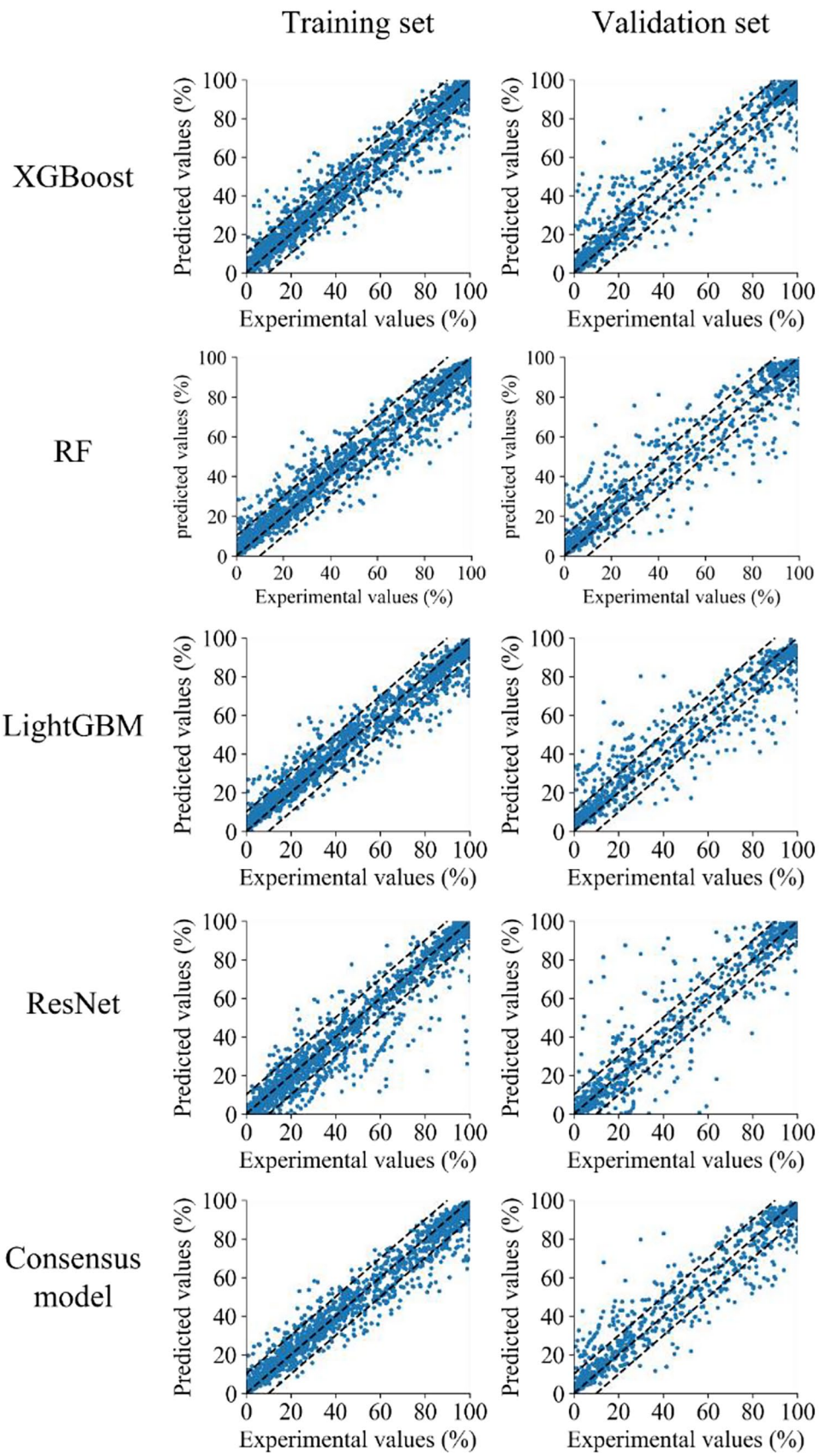

Fig. 4 Scatter plots of experimental values versus predicted values calculated by XGBoost, RF, LightGBM, ResNet, and the consensus model on the training set and the validation set

consensus model. As shown in the scatter plots (Fig. 4), the consensus model improved the predictions in both the initial release period and the complete release period. The consensus model reduced the prediction errors and showed smaller MAE values than that of the tree-based models in the 0–30% release period (MAE = 20.41 vs. 21.40, 21.41, 22.97), and in the 80–100% release period (MAE = 18.35 vs. 20.21, 20.31, 20.35) for the hard-to-predict samples. Table 4 also indicated that the consensus model achieved better performance on the validation set, suggesting an improvement in prediction accuracy. Additionally, the consensus model treated the prediction variance of these four models as prediction uncertainty, which would be small when individual models gave the similar predictions and presented large values when the disagreement appeared in the predictions. Overall, the consensus model may be a reliable solution for in vitro drug release prediction.

Feature importance for prediction model

ML can also provide feature importance of each factor during the model construction process, where the ranking results can inform pharmaceutical scientists of the important influencing parameters in microsphere formulation development. Figure 5 showed the feature importance results calculated by XGBoost, RF, and LightGBM algorithms. These features covered from formulation components, process parameters, and microspheres characteristics to dissolution conditions. Besides dissolution time, the top 5 features in Fig. 5A, B, and C were similar, which included particle size, drug loading, encapsulation efficiency, API concentration, PLGA concentration, dissolution temperature, and stirring speed (See Discussion).

Experimental validation

To further investigate the model performance, risperidone was selected as the model drug and twenty microsphere formulations with different combinations of formulation and process parameters were prepared. The in vitro drug release testing results were compared with the predictions from the established consensus model. Obtained risperidone-loaded microspheres all possessed highly spherical shape with porous surface. Particle size ranged from 39.31 to 124.13 μm. All drug loading was higher than 25%, while encapsulation efficacy was above 80%. Supplemental materials showed the partial SEM pictures (Fig. S1) and characteristics of microspheres (Table S2).

The consensus model exhibited good performance to predict the release behavior of experimental results, in which R^2 reached 0.949, and MAE, MSE, and RMSE gave 5.3, 67.7, and 8.2 respectively (Table S3 in supplemental materials). Furthermore, the f_2 factor can compare similarity between two profiles. A successful prediction is an f_2 value higher than or equal to 50, indicating the average difference within 10%. The accuracy is the percentage of successful predictions in all predictions. The f_2 values between experimental and predicted released profiles of risperidone-load microspheres were shown in Table 5. Fourteen of the twenty formulations obtained an f_2 higher than 50, and the accuracy of the consensus model was 70% on this experimental validation set.

Figure 6 showed the specific predicted profiles from the consensus model. Most prediction profiles matched the experimental release profiles well, especially for the prediction of burst release (Fig. 6B and F). Agreed with our analysis above, conservative predictions were given for some formulations in the range of 80–100% (Fig. 6A, C, and F). Besides, Fig. 6I showed that the consensus model was able to provide more accurate predictions at 45 ℃ than at 37 ℃ directly because the data amount at 45 ℃ was double than at 37 ℃. Generally, these results above suggested the consensus model had good predictive ability and well discovered the influence of variables on the results, which could guide the experiment design.

Molecular dynamics simulation results

MD simulation was applied to provide the micro-level view of the microspheres preparation process. It can contribute to a better understanding of the microscopic changes and molecular mechanism involved in microsphere formation which the traditionally experimental methods are hard to investigate. Figure 7 revealed the dynamic preparation process of microspheres. Figure 7A and B showed the initial structure and 1 ns simulation process result of the oil phase system, in which polymers and drugs randomly dispersed in the organic solvent. Specifically, PLGAs exhibited obvious partially contracted into coil-like and partly expand (Fig. 7B). This result agreed with the Flory–Krigbaum theory [44, 45]. When in a good solvent, the interaction between polymer segments and solvent molecules would be energetically favorable and lead to polymer coils expanding, indicating the influence of solvent selection on the excipient. In the water phase, PVA molecules exhibited the tendency to self-fold and gather together (Fig. 7C). This is because PVA is an amphiphilic molecule with hydrophilic hydroxyl groups and hydrophobic acetyl groups. After adding the oil phase system into the water phase and removing the DCM, the simulation results in Fig. 7D and E indicated that drugs and polymers aggregated into a tight structure, which was agreed with our previous study [31]. Most drug molecules were encapsulated within the polymer sphere, while only a few drug molecules and PVA molecules were accumulated at the surface, indicating the need to wash the prepared microspheres.

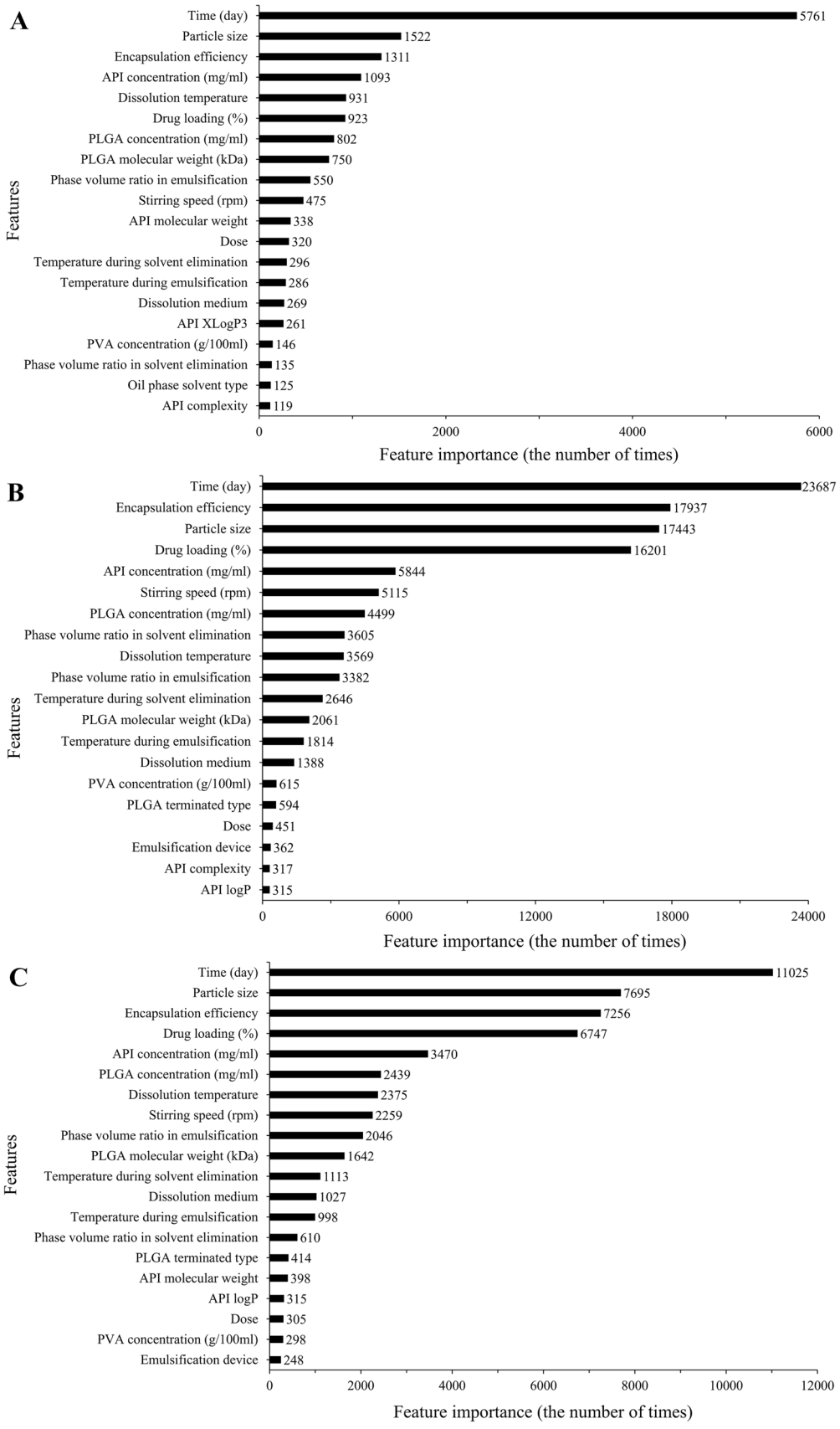

Fig. 5 Feature importance in microspheres dataset ranked by XGBoost **A**, RF **B**, and LightGBM **C** algorithm

Table 5 f_2 values between experimental and predicted released profiles

Formulation	f_2 value	Formulation	f_2 value
1	51.98	11	73.93
2	46.52	12	49.17
3	52.44	13	79.31
4	66.19	14	67.09
5	58.34	15	53.00
6	49.13	16	46.14
7	68.55	17	58.93
8	74.07	18	59.66
9	41.60	19	37.85
10	53.05	20	57.79

Figure 8A and B showed the result of RMSD and Rg. RMSD was used to measure the change of system, while Rg described the distribution of drugs in solution, indicating the compactness of the system. The whole system was stable after about 90 ns simulation, exhibiting a small fluctuation less than 1-Å in RMSD result and a small Rg value of about 17-Å. In addition, SASA was used to describe the contact area between drug and solvent molecules, and the averaged hydrogen bonds showed the number of hydrogen bonds formed between drug and PLGA molecules in the simulation process (Fig. 8C and D). The descending SASA and ascending averaged number of hydrogen bonds within the 180 ns simulation indicated the increasing aggregation of drug and PLGA molecules. These results showed that the stable and tight structure were obtained.

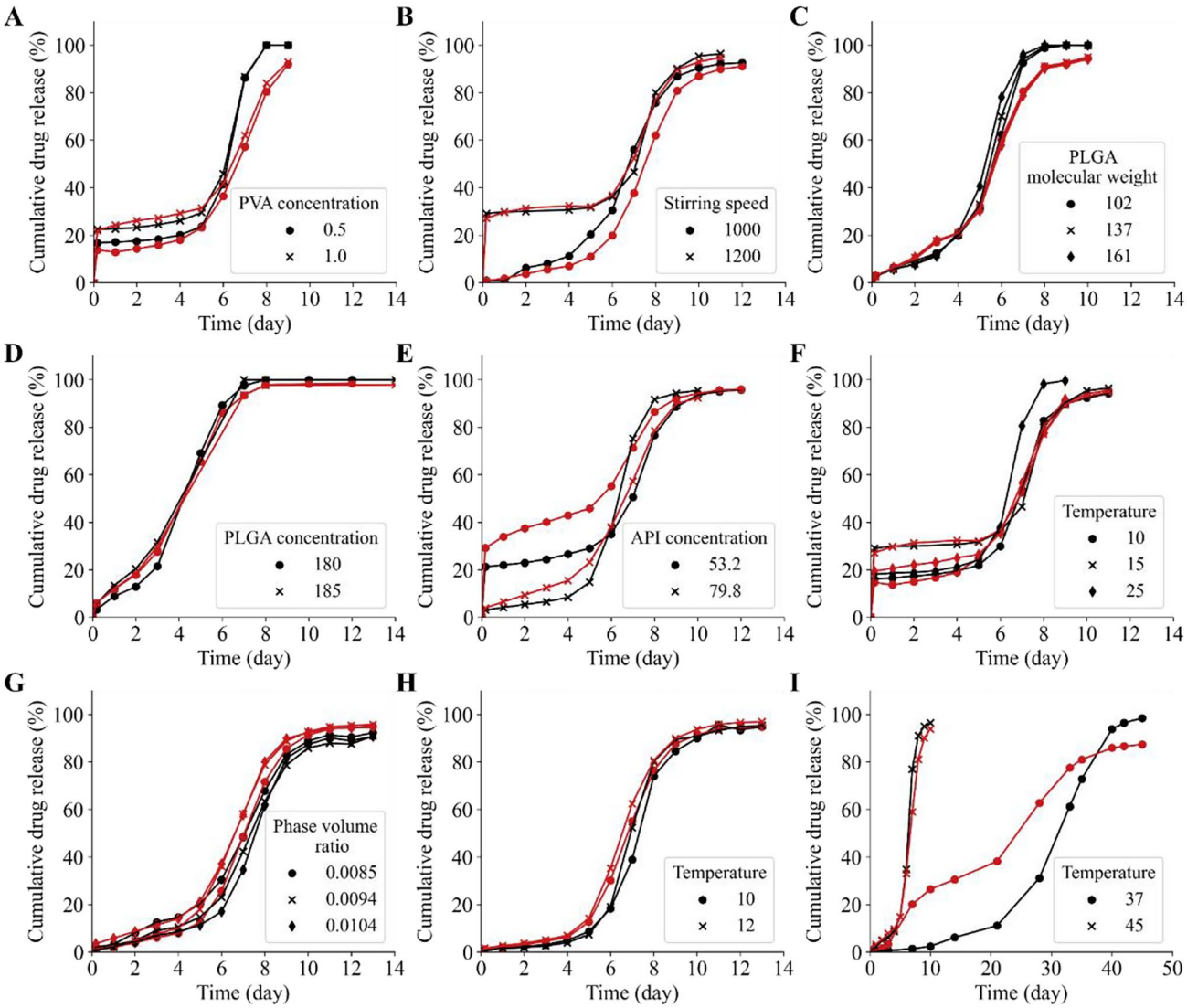

Fig. 6 Experimental in vitro release profiles of risperidone-loaded microspheres versus predicted profiles of the consensus model. Red, predicted values; black, experimental values. **A** PVA concentration, **B** stirring speed, **C** PLGA molecular weight, **D** PLGA concentration, **E** API concentration, **F** temperature during solvent elimination, **G** phase volume ratio, **H** temperature during emulsification, **I** dissolution temperature

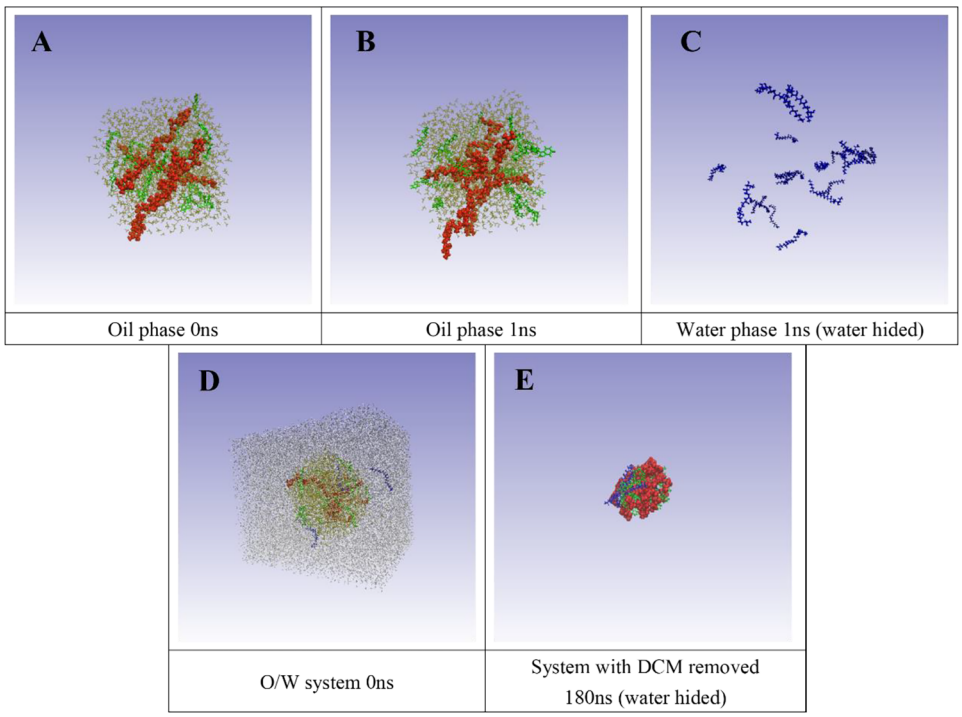

Fig. 7 Snapshots of MD simulations in the oil phase at **A** 0 and **B** 1 ns, in the water phase at 1 ns **C**, in the O/W system at 0 ns **D**, in the system with DCM removed at 100 ns. Green molecules, risperidone; red molecules, PLGA; blue molecules, PVA

Discussion

Prediction model in microsphere formulation development

Formulation development is commonly full of uncertainty. The consensus model was established on 286 release profiles from publications and pharmaceutical company and has shown good predictive capability on drug release behaviors, which can provide instruction and help ease such uncertainty. Besides, the development of microspheres required the consideration of various variables, especially the process parameters which were crucial for scalability investigation. Our prediction model pointed out the key factors in the microsphere formulation and process, which was able to facilitate the formulation and process design. Moreover, the elevated temperature method has been used by the industry for rapid quality control.

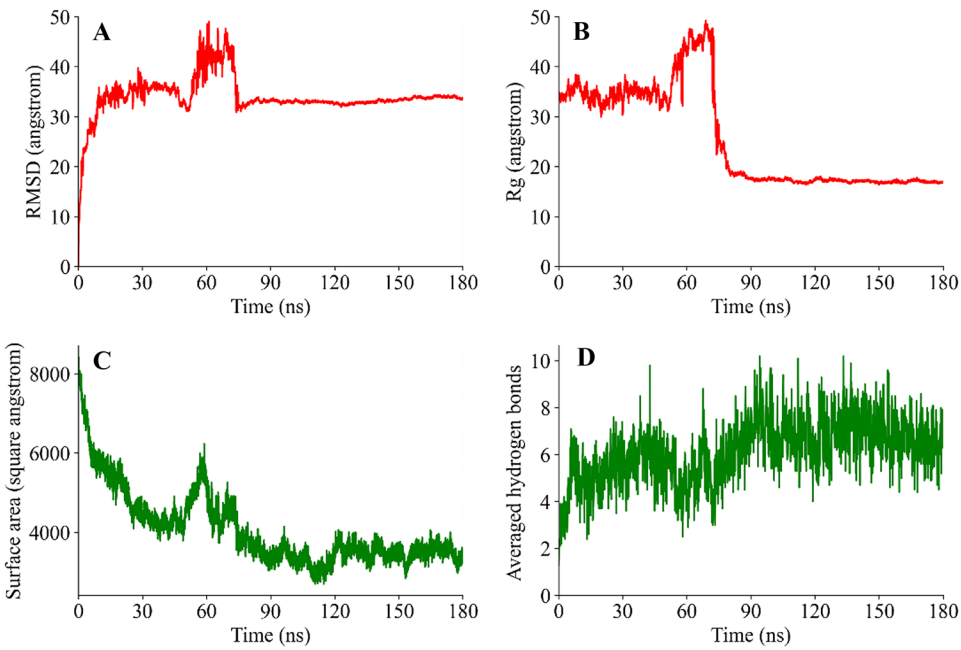

Fig. 8 CPPTRAJ analysis results of solvent removed system. **A** Root-mean-square deviation vs. time. **B** Radius of gyration of the system vs. time. **C** Solvent accessible surface areas vs. time. **D** The averaged hydrogen bonds between drug and PLGA molecules during the simulation

The good predictions for drug release at both 37 °C and 45 °C were provided by consensus model, which can further reduce the time required and better meet the needs of pharmaceutical industry.

Feature importance of prediction model

The microspheres research already revealed numerous influencing factors on drug release but hard to tell the priorities involved. Our consensus model can provide not only the predictions but also the ranking result of the input features. Besides dissolution time and dissolution temperature, some formulation and process parameters, such as the encapsulation efficiency, drug loading, particle size, API concentration, stirring rate, and PLGA concentration, were identified as the key factors affecting the drug release behavior of microspheres.

Drug loading was described as the amount of drug encapsulated in the microspheres. The increased drug loading was considered to increase the release rate and the release duration [8, 42]. API concentration can also affect the drug release behavior by influencing the drug loading. High API concentration may lead to inadequate homogenization and result in the uneven drug distribution in microspheres. Encapsulation efficiency was associated with the drug loading and concentration of API. The effect of particle size has been widely recognized. As the particle size decreases, the surface-area-to-volume ratio and release rate increase. High stirring speed can lower the particle size by decreasing the size of microdroplets formed. The elevated dissolution temperature results in improved the polymer mobility and faster drug diffusion, hence achieve increased drug release. Drug dissolution rate would increase greatly, when temperature was around the glass transition temperature [46]. PLGA concentration was also an imperative factor on drug release. High PLGA concentration would cause the rapid polymer precipitation and accompanying with chain entanglement, microspheres with low porosity and high density were obtained [47]. Moreover, the emulsification with higher PLGA concentration would exhibit relatively less shrinkage during the solvent evaporation process, which resulted in the microspheres with bigger size [38, 48].

Accelerated release testing by elevating temperature

For sustained-release microspheres, dissolution testing at physiological temperature (typically 37 °C) often takes weeks for each batch, which is inefficient and unacceptable for the preliminary formulation screening. To speed up, pharmaceutical companies usually apply the elevated temperature (mostly 45 °C), which was also mentioned in the FDA regulatory science program and EUFEPS workshop [9, 10]. The relevance between the release profiles at the elevated temperature and time-scaled ones at 37°C is constrained by the glass transition temperature (T_g) of microspheres [46]. The T_g refers to the temperature range at which the material transforms into the rubbery state from the glassy one [49], and the T_g of microspheres typically depends on multiple factors, such as polymer properties, drug amount, solvent residue, and others [50]. Generally, when the dissolution temperature exceeds T_g, the polymer mobility and free volume will increase, and microspheres would suffer from structural reconfiguration, such as pores closure and structural deformation, which consequently changes the drug release kinetics [51, 52], and therefore decreases the correlation between accelerated and real-time condition. Recent studies have already investigated the T_g of risperidone-loaded microspheres [51, 53], and release profiles under accelerated condition at 45 °C and physiological condition were demonstrated to be with high correlation ($R^2 > 0.98$).

Although the elevated temperature approach significantly shortens the release time, the initial burst release is difficult to assess under the accelerated condition, which is a commonly existing and crucial concern in depot formulations [10]. The product microspheres are in thermodynamic disequilibrium and PLGA would spontaneously undergo physical aging to reach equilibrium [54, 55]. Physical aging could alter the PLGA matrix structure, reducing free space and causing the structure to relax heterogeneously, subsequently resulting in the development of microvoids. During the dissolution process, the water instantly rushes into the voids, resulting in a significant initial drug release, known as the burst release [55, 56]. Therefore, initial phase investigations at physiological temperature for burst release evaluation are indispensable [10].

The data of the two conditions was integrated in the modeling, which enabled the predictions at both 37 °C and 45 °C, and the performance of 45 °C condition was slightly better thanks to the better quality and quantity of data. The users can employ our models following the idea of experimental practice, where 37 °C condition is used for burst release assessment and 45 °C condition for rapid screening of formulation and process based on the correlation between these two conditions. This highly practical prediction model will provide valuable guidance for the investigation of microsphere formulations.

Conclusion

In this study, a constructed consensus model by the XGBoost, RF, ResNet, and LightGBM algorithms was successfully built with good predictive capability in predicting release profiles at both 37 ℃ and 45 ℃ for microsphere

formulation with small-molecule drugs. The formulation feature importance provided by predictive models gave hints and instruction for formulation experimental design. Further experimental results validated the capability of our consensus model, and the application of MD simulation revealed the molecular mechanism of the risperidone-loaded microsphere formation. In vitro dissolution results obtained rapidly using the prediction model and the understanding of formation mechanism at microscopic level derived from the MD simulation results can assist the formulation researchers in the design and screening of formulations, which is promising to help accelerate the rational design and development of microspheres.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13346-022-01253-z.

Acknowledgements Molecular modeling was performed at the High-Performance Computing Cluster (HPCC) which is supported by Information and Communication Technology Office (ICTO) of the University of Macau.

Author contribution Jiayin Deng: data collection, molecular dynamic simulation, and writing—original draft preparation; Zhuyifan Ye: machine learning modeling and writing—reviewing and editing; Wenwen Zheng: study design and revision; Jian Chen: experimental work; Haoshi Gao: data analysis; Zheng Wu: platform construction; Ging Chan: supervision; Yongjun Wang: supervision; Dongsheng Cao: supervision; Yanqing Wang: study conception and design; Simon Ming-Yuen Lee: study conception and design; Defang Ouyang: study conception and design and writing—reviewing and editing.

Funding Current research is financially supported by the Zhuhai-HongKong-Macau Collaboration Project (ZH22017002210010PWC) and the Macau FDCT research grant (0108/2021/A).

Availability of data and materials All data and code will be available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate No ethical approval is required for this study.

Consent for publication All authors have agreed with publication of the manuscript.

Competing interests The authors declare no competing interests.

References

1. O'Brien MN, Jiang W, Wang Y, Loffredo DM. Challenges and opportunities in the development of complex generic long-acting injectable drug products. J Control Release. 2021;336:144–58.
2. Burgess DJ, Wright JC. An introduction to long acting injections and implants. In: Wright JC, Burgess DJ, editors. Long Acting Injections and Implants. US, Boston, MA: Springer; 2012. p. 1–9.
3. Wang Y, Burgess DJ. Microsphere Technologies. In: Wright JC, Burgess DJ, editors. Long acting injections and implants. US, Boston, MA: Springer; 2012. p. 167–94.
4. Wong CY, Al-Salami H, Dass CR. Microparticles, microcapsules and microspheres: a review of recent developments and prospects for oral delivery of insulin. Int J Pharm. 2018;537:223–44.
5. Su Y, Zhang B, Sun R, Liu W, Zhu Q, Zhang X, Wang R, Chen C. PLGA-based biodegradable microspheres in drug delivery: recent advances in research and application. Drug Deliv. 2021;28:1397–418.
6. Göpferich A. Mechanisms of polymer degradation and erosion. Biomaterials. 1996;17:103–14.
7. Yoo J, Won Y-Y. Phenomenology of the initial burst release of drugs from PLGA microparticles. ACS Biomater Sci Eng. 2020;6:6053–62.
8. Fredenberg S, Wahlgren M, Reslow M, Axelsson A. The mechanisms of drug release in poly(lactic-co-glycolic acid)-based drug delivery systems—a review. Int J Pharm. 2011;415:34–52.
9. Wang Y, Qu W, Choi SH. FDA's regulatory science program for generic PLA/ PLGA-based drug products. Am Pharm Rev. 2016.
10. Burgess DJ, Crommelin DJA, Hussain AS, Chen M-L. Assuring quality and performance of sustained and controlled release parenterals. Eur J Pharm Sci. 2004;21:679–90.
11. Lundström EA, Rencken RK, van Wyk JH, Coetzee LJ, Bahlmann JC, Reif S, Strasheim EA, Bigalke MC, Pontin AR, Goedhals L, Steyn DG, Heyns CF, Aldera LA, Mackenzie TM, Purcea D, Grosgurin PY, Porchet HC. Triptorelin 6-month formulation in the management of patients with locally advanced and metastatic prostate cancer: an open-label, non-comparative, multicentre, phase III study. Clin Drug Investig. 2009;29:757–65.
12. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, Zhao S. Applications of machine learning in drug discovery and development. Nat Rev Drug Discovery. 2019;18:463–77.
13. Han R, Xiong H, Ye Z, Yang Y, Huang T, Jing Q, Lu J, Pan H, Ren F, Ouyang D. Predicting physical stability of solid dispersions by machine learning techniques. J Control Release. 2019;311–312:16–25.
14. Gao H, Wang W, Dong J, Ye Z, Ouyang D. An integrated computational methodology with data-driven machine learning, molecular modeling and PBPK modeling to accelerate solid dispersion formulation design. Eur J Pharm Biopharm. 2021;158:336–46.
15. He Y, Ye Z, Liu X, Wei Z, Qiu F, Li H-F, Zheng Y, Ouyang D. Can machine learning predict drug nanocrystals? J Control Release. 2020;322:274–85.
16. Wang W, Feng S, Ye Z, Gao H, Lin J, Ouyang D. Prediction of lipid nanoparticles for mRNA vaccines by the machine learning algorithm. Acta Pharm Sin B. 2021.
17. Gao H, Jia H, Dong J, Yang X, Li H, Ouyang D. Integrated in silico formulation design of self-emulsifying drug delivery systems. Acta Pharmaceutica Sinica B. 2021.
18. Ye Z, Yang W, Yang Y, Ouyang D. Interpretable machine learning methods for in vitro pharmaceutical formulation development, Food. Frontiers. 2021;2:195–207.
19. Versypt AN, Pack DW, Braatz RD. Mathematical modeling of drug delivery from autocatalytically degradable PLGA microspheres — a review. J Controlled Rel. 2013;165:29–37.
20. Zawbaa HM, Szlęk J, Grosan C, Jachowicz R, Mendyk A. Computational intelligence modeling of the macromolecules release from PLGA microspheres-focus on feature selection. PLoS ONE. 2016;11: e0157610.
21. Rodrigues de Azevedo C, von Stosch M, Costa MS, Ramos AM, Cardoso MM, Danhier F, Préat V, Oliveira R. Modeling of the burst release from PLGA micro- and nanoparticles as function of physicochemical parameters and formulation characteristics. Int J Pharm. 2017;532:229–40.

22. Bannigan P, Häse F, Aldeghi M, Bao Z, Aspuru-Guzik A, Allen C. Machine learning predictions of drug release from polymeric long acting injectables. ChemRxiv. Cambridge: Cambridge Open Engage. 2021.
23. Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin VA, Radchenko EV, Zefirov NS, Makarenko AS, Tanchuk VY, Prokopenko VV. Virtual computational chemistry laboratory – design and description. J Comput Aided Mol Des. 2005;19:453–63.
24. Han R, Yang Y, Li X, Ouyang D. Predicting oral disintegrating tablet formulations by neural network techniques, Asian. J Pharm Sci. 2018;13:336–42.
25. Yang Y, Ye Z, Su Y, Zhao Q, Li X, Ouyang D. Deep learning for in vitro prediction of pharmaceutical formulations. Acta Pharm Sin B. 2018.
26. Shah VP, Tsong Y, Sathe P, Liu J-P. In vitro dissolution profile comparison—statistics and analysis of the similarity factor, f2. Pharm Res. 1998;15:889–96.
27. FDA, Guidance for Industry. SUPAC-IR. Immediate release solid oral dosage forms. Scale-up and post approval changes. Chemistry, manufacturing and controls. In vitro dissolution testing and in vivo bioequivalence documentation. 1995.
28. Bastien F, Lamblin P, Pascanu R, Bergstra J, Goodfellow I, Bergeron A, Bouchard N, Warde-Farley D, Bengio Y. Theano: new features and speed improvements. 2012. arXiv:1211.5590.
29. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X. TensorFlow: a system for large-scale machine learning. In: Proc 12th USENIX Conf Operating Sys Des Implement, USENIX Association, Savannah, GA, USA. 2016;265–283.
30. Lei T, Li Y, Song Y, Li D, Sun H, Hou T. ADMET evaluation in drug discovery: 15. Accurate prediction of rat oral acute toxicity using relevance vector machine and consensus modeling, Journal of cheminformatics. 2016;8:6.
31. Wang Y, Huang W, Wang N, Ouyang D, Xiao L, Zhang S, Ou X, He T, Yu R, Song L. Development of arteannuin B sustained-release microspheres for anti-tumor therapy by integrated experimental and molecular modeling approaches Pharmaceutics. 2021;13.
32. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field 2004;25:1157–1174.
33. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC Model: I Method, J Comput Chem. 2000;21:132–146.
34. Martínez L, Andrade R, Birgin EG, Martínez JM. PACKMOL: a package for building initial configurations for molecular dynamics simulations. J Comput Chem. 2009;30:2157–64.
35. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. 1996;14:33–38.
36. Liu J, Li D, Liu X. A simple and accurate algorithm for path integral molecular dynamics with the Langevin thermostat. J Chem Phys. 2016;145.
37. Berendsen HJ, Postma JV, Van Gunsteren WF, DiNola AR, Haak JR. Molecular dynamics with coupling to an external bath. 1984;81:3684–3690.
38. Park K, Skidmore S, Hadar J, Garner J, Park H, Otte A, Soh BK, Yoon G, Yu D, Yun Y, Lee BK, Jiang X, Wang Y. Injectable, long-acting PLGA formulations: analyzing PLGA and understanding microparticle formation. J Control Release. 2019;304:125–34.
39. Skiveren J, Nordahl Larsen H, Kjaerby E, Larsen R. The influence of needle size on pain perception in patients treated with botulinum toxin A injections for axillary hyperhidrosis. Acta dermato-venereologica. 2011;91:72–74.
40. Gill HS, Prausnitz MR. Does needle size matter? J Diabetes Sci Technol. 2007;1:725–9.
41. Haas P, Falkner-Radler C, Wimpissinger B, Malina M, Binder S. Needle size in intravitreal injections - pain evaluation of a randomized clinical trial. Acta Ophthalmol. 2016;94:198–202.
42. Wischke C, Schwendeman SP. Principles of encapsulating hydrophobic drugs in PLA/PLGA microparticles. Int J Pharm. 2008;364:298–327.
43. Wang J, Wang BM, Schwendeman SP. Characterization of the initial burst release of a model peptide from poly(d, l-lactide-co-glycolide) microspheres. J Control Release. 2002;82:289–307.
44. Flory PJ. Thermodynamics of high polymer solutions. J Chem Phys. 1941;9:660–1.
45. Flory PJ, Krigbaum WR. Statistical mechanics of dilute polymer solutions II. 1950;18:1086–1094.
46. Zolnik BS, Leary PE, Burgess DJ. Elevated temperature accelerated release testing of PLGA microspheres. J Control Release. 2006;112:293–300.
47. Yeo Y, Park K. Control of encapsulation efficiency and initial burst in polymeric microparticle systems. Arch Pharmacal Res. 2004;27:1.
48. Rosca ID, Watari F, Uo M. Microparticle formation and its mechanism in single and double emulsion solvent evaporation. J Control Release. 2004;99:271–80.
49. Shrivastava A. 1 - Introduction to plastics engineering. In: Shrivastava A, editor. Introduction to Plastics Engineering. William Andrew Publishing; 2018. p. 1–16.
50. Prudic A, Lesniak AK, Ji Y, Sadowski G. Thermodynamic phase behaviour of indomethacin/PLGA formulations. Eur J Pharm Biopharm: Official J Arbeitsgemeinschaft Fur Pharmazeutische Verfahrenstechnik e.V. 2015;93:88–94.
51. Rawat A, Bhardwaj U, Burgess DJ. Comparison of in vitro–in vivo release of Risperdal® Consta® microspheres. Int J Pharm. 2012;434:115–21.
52. Andhariya JV, Burgess DJ. Recent advances in testing of microsphere drug delivery systems. Expert Opin Drug Deliv. 2016;13:593–608.
53. Shen J, Lee K, Choi S, Qu W, Wang Y, Burgess DJ. A reproducible accelerated in vitro release testing method for PLGA microspheres. Int J Pharm. 2016;498:274–82.
54. Park K, Otte A, Sharifi F, Garner J, Skidmore S, Park H, Jhon YK, Qin B, Wang Y. Potential roles of the glass transition temperature of PLGA microparticles in drug release kinetics. Mol Pharm. 2021;18:18–32.
55. Struik LCE. Physical aging in amorphous polymers and other materials. 1977.
56. Yoshioka T, Kawazoe N, Tateishi T, Chen GJMM. Engineering. Effects of structural change induced by physical aging on the biodegradation behavior of PLGA films at physiological temperature. 2011;296:1028–34.