SUPPLEMENTARY INFORMATION

Machine Learning Predictions of Drug Release from Polymeric Long Acting Injectables

Pauric Bannigan†, Florian Häse²,³,⁴, Matteo Aldeghi²,³,⁴, Zeqing Bao¹, Alán Aspuru-Guzik²,³,⁴,⁵*, Christine Allen¹*

¹ Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON M5S 3M2, Canada.
² Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada.
³ Department of Computer Science, University of Toronto, Toronto, ON M5S 3H6, Canada.
⁴ Vector Institute for Artificial Intelligence, Toronto, ON M5S 1M1, Canada.
⁵ Lebovic Fellow, Canadian Institute for Advanced Research, Toronto ON, M5S 1M1, Canada.

† P. Bannigan and F. Häse contributed equally to this work.

CORRESPONDING AUTHORS

*Christine Allen
Leslie Dan Faculty of Pharmacy,
University of Toronto,
Toronto, ON, M5S3M2,
Canada
Email: cj.allen@utoronto.ca

*Alan Aspuru-Guzik
Department of Chemistry,
University of Toronto,
Toronto, ON, M5S 3H6
Canada
E-mail: alan@aspuru.com
1. Materials and Methods:

1.1. Dataset construction

The training dataset used for model development consisted of 102 drug release profiles of 13 drugs. Some of this data has been previously collected by the Allen lab, and all data has been previously published. Data taken exclusively from the literature was obtained through a “Web of Science” database search using the keyword combinations of “polymeric microparticle” and “drug delivery”. Data was extracted from published release profiles using the “GetData Graph Digitizer” application. The external validation dataset used for the pseudo-prospective study consists of 79 release profiles collected from the literature in a similar way. Both of these datasets are available as supplementary information files and also online at the Aspuru-Guzik Group’s GitHub page (https://github.com/aspuru-guzik-group/to-be-released-upon-acceptance).

1.2. Model architectures

The neural network (NN) architectures deployed in this study were trained to minimize the root mean square deviation (RMSD) between their predictions, \( \{p_i\}_{i=1}^n \), and the targeted release fractions, \( \{t_i\}_{i=1}^n \). Mean variance estimate (MVE) networks, in contrast, follow a different approach to reproducing the target data from the input features. Instead of predicting the target value for a given input feature directly, MVEs parameterize a distribution from which the target value is sampled with high likelihood. The MVEs constructed in this study assume that target values are sampled from normal distributions, \( t \sim N(\mu, \sigma) \), parameterized with a location, \( \mu \), and a scale, \( \sigma \), which are both predicted from a conventional NN architecture. Instead of minimizing the mean square deviation between the predicted location \( \mu_i \) and the associated target \( t_i \), MVEs are trained to maximize the likelihood, which is equivalent to minimizing the negative log-likelihood, defined as Eq. S1:

\[
l(\{x\}_{i=1}^n|\{\mu\}_{i=1}^n, \{\sigma\}_{i=1}^n) = -\frac{n}{2} \ln(2\pi) - \sum_{i=1}^{n} \ln(\sigma_i) - \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{2\sigma_i^2}
\]

Equation S1

MVEs can maximize the likelihood by increasing its prediction accuracy and thus minimize the \((x_i - \mu_i)^2\) term in Eq. S1 or alternatively, if high prediction accuracies cannot be achieved for certain inputs, by maximizing the associated scale parameter, \( \sigma_i \), thus actively learning for which
inputs the computed predictions are inaccurate. This feature allows interpretation of the predicted scale parameter as a quantitative degree of confidence on the prediction. However, the predicted scale parameter should not be interpreted as an absolute uncertainty on the prediction. The NN and MVE models predicted release curves iteratively, by estimating the additional release day by day. Hence, their output is guaranteed to be positive but can exceed 100% release (i.e., a release fraction of 1). Therefore, an additional hyperbolic tangent activation function was used to ensure their predictions would also lie in the [0,1] interval, as expected. In addition to these NN that predicted release iteratively, models that would predict the whole curve in one shot (i.e., $\text{NN}_{\text{sig}}$ and $\text{MVE}_{\text{sig}}$) by estimating the parameters of a sigmoid function were also investigated. In these cases, an additional learnable parameter was included to account for deviations from the ideal sigmoidal curve. In all cases, L2 regularization, batch normalization, and early stopping was used. The Adam algorithm (https://arxiv.org/abs/1412.6980) was used for the optimization of the models’ parameters. The hyper-parameters of the different model architectures are summarized in Table S1.

| Hyperparameter               | Selected value/output |
|------------------------------|-----------------------|
| Number of layers             | 4                     |
| Number of neurons per layer  | 48                    |
| L2 regularization weight     | 1e-3                  |
| Learning rate                | 1e-4                  |
| Batch size                   | 100                   |
| Hidden layer activation      | Leaky ReLU            |
| Output layer activation      | Softplus              |

2. **Additional Results:**

2.1. **Machine learning**

The initial training dataset consisted of 25 input features that described the various LAI systems. During the model development phase, NN structures with various numbers of input features (i.e., 9, 12, 13 and 25) were trained and evaluated using leave-one-group-out (LOGO) cross-validation. Briefly, LOGO involved grouping the data into drug-polymer combinations, withholding one of
these groups, training the model on the remaining groups and evaluating the accuracy of the predicted versus actual drug release profiles for the withheld group in terms of the Pearson correlation coefficient (PCC) and RMSD (Figure S1).

**Figure S1:** Performance of several deep network architectures and descriptor sets on the task of predicting drug release profiles. Upper panel: achieved root mean square deviations. Lower panel: achieved Pearson correlation coefficients.

While feature sets should overall be comprehensive and encode all relevant physical information to make accurate predictions, adding too many, possibly irrelevant features, increases the risk of introducing spurious correlations between individual features and observed release profiles which eventually degrade the generalizability of the data-driven models. Four different sets of features: (i) [9 features], (ii) [12 features], (iii) [13 features], and (iv) [25 features], were generated based on pharmaceutical formulation experience as well as feature correlation (Figure S2) to probe the effects of feature sets on the model performance. The 9-feature set includes: LogP, pKa, and melting temperature of drugs, cross-linking ratio of polymers, surface area-volume-ratio and drug loading capacity of formulations, drug solubility enhancer concentration in release media, and the
days and gaps of sampling; the 12-feature set includes: 9-feature set, molecular weight of drugs and molecular weight and Csp3 fraction of polymers; the 13-feature set includes: 9-feature set, topological polar surface area of polymers and total number of rotatable bonds of drugs/polymers; the 25-feature set includes: 13-feature set as well as total number of proton acceptors/donors, min/max partial charge, total number of heteroatoms, and molecular weight of drugs/polymers. Both NN and MVE architectures are constructed to satisfy two physical constraints: (i) predicted release fractions must be between 0 and 1, and (ii) release fractions must increase over time. To satisfy these constraints, the network models are set up to predict the increase in the release fraction during a given period of time at a given number of days after the preparation of the drug-polymer system. Output activations are chosen to be strictly positive functions and drug release profiles are obtained from integrated model predictions which are transformed by hyperbolic tangents to satisfy constraint (i).
Figure S2: Correlation between features in a) 25, b) 13, c) 12, and d) 9 -feature set. The p-value equal to 0 suggests no correlation and equal to 1 suggests perfect correlation.

2.2. Pseudo-prospective study

A pseudo-prospective study was conducted using the external validation dataset to further examine the prediction accuracy and the limitation of the model. This study involved several steps. First, the selected (and trained) 13-feature MVE was used to predict the 79 release profiles contained in the external validation dataset. The resulting predictions are shown plotted against their respective experimental release profiles in Figure S3 and Figure S4.
Figure S3: Release profiles for the LAI formulations included in the external validation dataset (black) in comparison to their corresponding predictions (orange) with uncertainty (blue)-Part I. Numbers correspond to the uploaded raw dataset.
Figure S4: Release profiles for the LAI formulations included in the external validation dataset (black) in comparison to their corresponding predictions (orange) with uncertainty (blue)-Part II. Numbers correspond to the uploaded raw dataset.
Second, following the prediction of these drug release profiles the overall performance of the 13-feature MVE was determined. Here model performance was quantified in terms of prediction accuracy, as defined in Eq.2. Briefly, the differences between actual drug release percent and the corresponding predicted drug release percent were calculated, and the model accuracy was defined by 1 minus the mean of these absolute differences for each individual drug release profile. The overall accuracy of the model on the external validation dataset as well as the Pearson’s correlation coefficient (PCC) is shown in Figure S5.

\[
\text{Accuracy} = 1 - \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}
\]  

Equation S2

\(y_i = \text{actual drug release percent}\)

\(x_i = \text{predicted drug release percent}\)

\(n = \text{total number of timepoints}\)

![Figure S5: Summary of 13-feature MVE model accuracy and PCC as determined through the pseudo-prospective study.](image)