Towards Data-Driven Development of Polymeric Long-Acting Injectables

Zeqing Bao[1], Pauric Bannigan[1], Riley J. Hickman[2,3,4], Matteo Aldeghi[2,3,4], Florian Häse[2,3,4], Alán Aspuru-Guzik[2,3,4,5], Christine Allen[1]

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

Introduction

Long-acting injectable formulations (LAIFs) can offer several advantages over conventional medicines, including increased bioavailability and improvements in patient compliance [1]. Several types of LAIFs have been approved for human use, including drug-loaded polymeric microparticles (MPs) [2]. Lupron Depot®, the first MP-based LAIF approved by the FDA, consists of leuprolide acetate encapsulated in poly(lactide-co-glycolide acid) (PLGA) MPs [1]. Today the majority of approved LAIFs are composed of PLGA (46%) [1]. Despite the success of PLGA-based systems they cannot be universally employed to deliver all drugs. Each drug has its own unique physico-chemical properties and drug-material interactions have a significant influence on the performance of LAIFs. Machine learning (ML) offers the ability to predict how drug-material combinations will affect final formulation properties, such as release kinetics, *a priori* [3]. This study has been submitted for publication (available on ChemRxiv [4]).

Methods

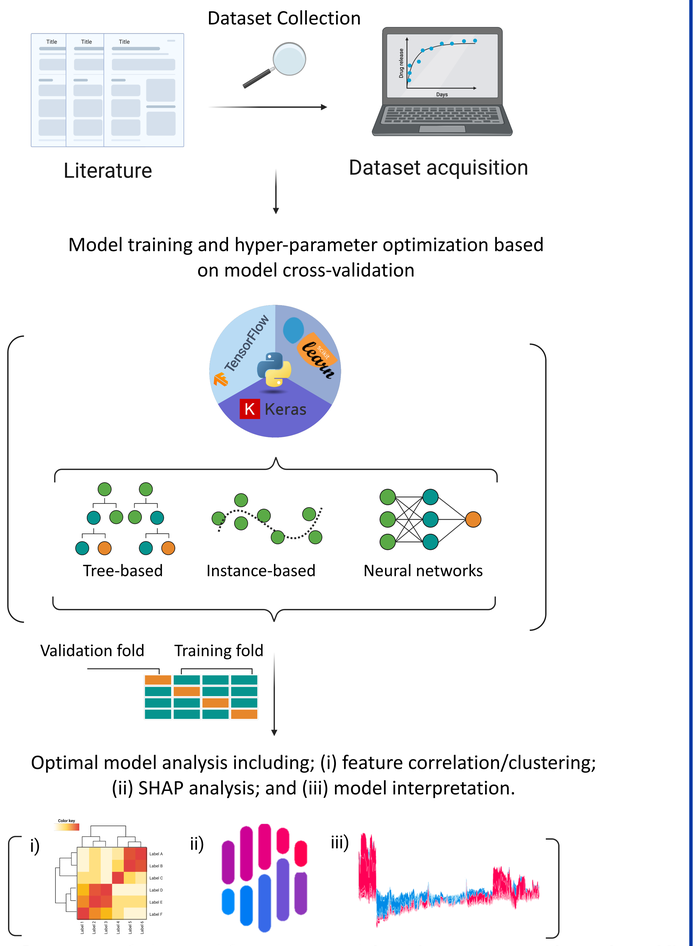

Figure 1. Summary of the workflow employed in this study to train various ML models to predict fractional drug release from LAIs.

Results

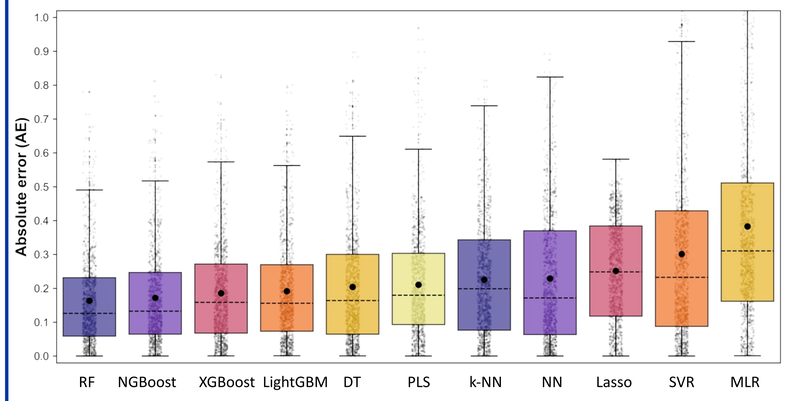

Figure 2. Summary of the overall predictive performance achieved by the various ML models during the pseudo-prospective study.

Table 1. Mean absolute error (MAE) values obtained for the different ML algorithms following prediction of fractional drug release for (a) the training dataset, and (b) the external validation dataset (pseudo-prospective study).

(a) MAE values obtained on training dataset following LOGO cross validation (n = 22 folds).

	RF	NGBoost	XGBoost	LightGBM	DT	PLS	k-NN	NN	LASSO	SVR	MLR
MAE	0.164	0.170	0.181	0.167	0.182	0.311	0.210	0.218	0.330	0.234	0.359

(b) MAE values obtained during the pseudo-prospective study (n = 1017 samples)

	RF	NGBoost	XGBoost	LightGBM	DT	PLS	k-NN	NN	LASSO	SVR	MLR
MAE	0.164	0.172	0.186	0.191	0.204	0.210	0.226	0.229	0.252	0.301	0.383
Standard Deviation	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.006	0.005	0.009	0.010
Median AE	0.126	0.133	0.158	0.156	0.164	0.179	0.199	0.172	0.249	0.233	0.311

Results

Table 2. Table summarizing the impact of feature cluster removal on the predictive performance of the RF model.

Effect of features engineering on prediction accuracy		
No. of Features	**Feature(s) removed**	**MAE**
14-features	n/a	0.164
12-feature	Drug_NHA, Drug_TPSA	0.164
10-feature	SE, Initial D/M ratio	0.166
6-features	Drug_PKa, SA-V, CL Ratio, DLC	0.178
3-features	Time, Polymer_MW, Drug_logP	0.256

Results

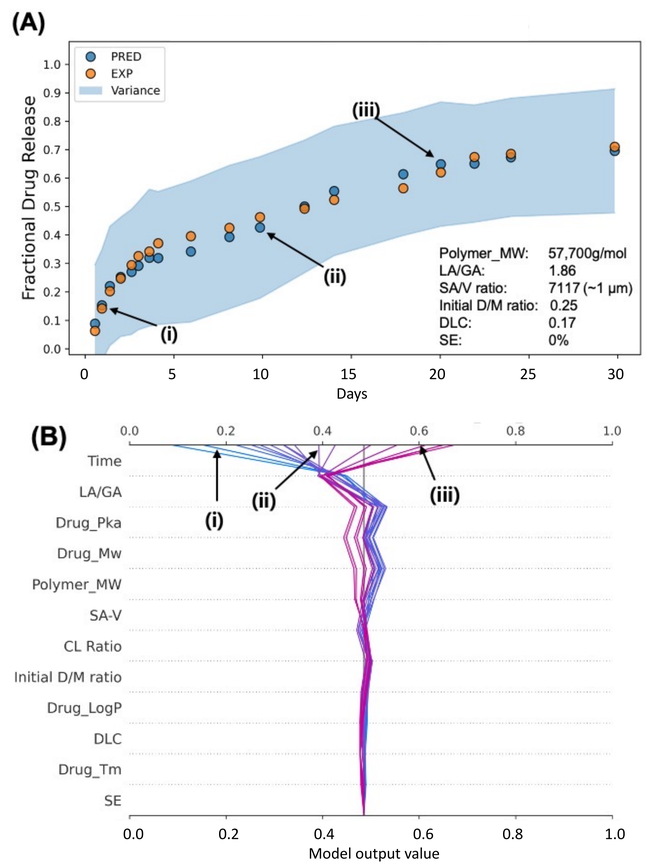

Figure 4. (A) Experimental fractional drug release profile (orange circles) for LTZ-PLGA Experimental_index39 plotted against the predicted fractional drug release profile (blue circles) and associated variance (blue halo) [5] generated by the RF model. Three fractional drug release predictions (i, ii, and iii) are highlighted. (B) Decision path taken for each fractional drug release prediction in the LTZ-PLGA Experimental_index39 example, with the same three fractional drug release predictions for (A) highlighted.

Conclusion

The 12-feature RF model returned the most accurate predictions, as well as a measure of confidence in those predictions. The integration of ML models into drug formulation development may enable us to move away from the current trial-and-error-based approach and adopt a more efficient and sustainable formulation development process that is "data-driven".

Acknowledgments

This work was funded through an NSERC Discovery Grant awarded to CA.

[1] K. Park *et al.*, "Injectable, long-acting PLGA formulations: Analyzing PLGA and understanding microparticle formation," *J. Controlled Release*, vol. 304, pp. 125–134, Jun. 2019, doi: 10.1016/j.jconrel.2019.05.003.
[2] C. I. Nkanga *et al.*, "Clinically established biodegradable long acting injectables: An industry perspective," *Adv. Drug Deliv. Rev.*, vol. 167, pp. 19–46, Dec. 2020, doi: 10.1016/j.addr.2020.11.008.
[3] P. Bannigan, M. Aldeghi, Z. Bao, F. Häse, A. Aspuru-Guzik, and C. Allen, "Machine learning directed drug formulation development," *Adv. Drug Deliv. Rev.*, May 2021, doi: 10.1016/j.addr.2021.05.016.
[4] P. Bannigan, Z. Bao, R. Hickman, M. Aldeghi, F. Häse, A. Aspuru-Guzik, C. Allen, Machine Learning Models to Accelerate the Design of Polymeric Long-Acting Injectables, (2022). https://doi.org/10.26434/chemrxiv-2021-mxrxw-v2.
[5] K. Polimis, A. Rokem, B. Hazelton, Confidence Intervals for Random Forests in Python, Journal of Open Source Software. 2 (2017) 124. https://doi.org/10.21105/joss.00124.