Poster 14

Acceleration of drug development using artificial intelligence and machine learning

Authors: Hansol Lee, Shubham Sharma, Marissa Lee Puzan, Joseph Krzyzaniak
Presenting Author: Hansol Lee
Affiliation: Pfizer Worldwide Research and Development, Drug Product Design
Corresponding Author's email: hansol.lee@pfizer.com

Purpose

Developing the next generation of pharmaceutical solid oral dosage forms requires accurate understanding of active pharmaceutical ingredients' (APIs) performance attributes and their impact on bioavailability, manufacturability, and stability. As drug development becomes increasingly more complex, the failure to predict problems with the API performance attributes like flow, sticking or dissolution has resulted in increased cost of development. To mitigate this, computational and data science tools can be leveraged to understand and predict these attributes early in a drug development program. This is an industry-recognized challenge and opportunity. In fact, Phamaceutical scientists have previously published on the topic, demonstrating a significant need in supporting accelerated drug development. There has also been recent academic research demonstrating the potential use of machine learning in this endeavor.

As such, the goal for this poster is to demonstrate the latest attempts at developing computational tools to predict performance attributes of pharmaceutical powders, their shortcomings, and potential ways to develop improved models.

Methods

The methodology to develop a predictive model will involve generating experimental data on subsets of pharmaceutically relevant materials and leveraging machine learning on the dataset to develop standards that can better represent particle diversity. These standards will constitute the training data set for machine learning.

Results

We present and discuss various models from industry and academia that have sought to predict performance attributes from API particles. The general approach from the research was to distill down the particle complexity by reporting the particle properties such as size into D10, D50, and D90 values and relate them directly to the powder flow. This is illustrated clearly in Diaz et al. 2022. While they curated approximately 100 different model systems to train the machine learning model to predict flow from particle size, the inability to accurately capture the particle diversity resulted in less than 62% accuracy. To date, applying similar modeling approaches on Pfizer API datasets yielded inadequate results.

Instead, by developing internal standards that represent the particle diversity observed in pharmaceutical materials as well as generating subsets of data that will be used to train the machine learning model, a more accurate predictive tool could be developed.

Conclusion

In this poster, we summarize the latest industrial and academic works that have sought to predict the performance attributes of APIs from their particle properties. We discuss their shortcomings and our plan to solve this problem by developing standards that represent the particle diversity observed in pharmaceutical materials as well as generating subsets of data that will be used to train the machine learning model.

Keywords: Machine learning, crystals, particles, powder, pharmaceutical, development, models, computational, experimental