Abstract

The rapid development and increasing use of computer science and new data analysis methods have gained significant interest also in the pharmaceutical field. In this field, traditionally used data analysis approach is based on design of experiments combined with multivariate regression and response surface methodology. The added value of new data analysis methods, such as machine learning, is their accuracy and predictive power. Machine learning methods can handle non-linearity, a high number of (correlated) variables, and large amounts of data, which is a common scenario within pharmaceutical sciences. However, the utilization of machine learning methods in pharmaceutical formulation and process development discipline is not as prominent as in other scientific fields. The current challenges of machine learning are the interpretability of generated models and the significant technical effort needed for implementation.

This thesis aims to combine a risk-based approach and machine learning to explore a complex pharmaceutical system of injectable aqueous crystalline suspensions (ACS). Complex pharmaceutical systems have typically many variables related to range of responses. These variables and responses do often interact and the relationships between them are non-linear. Firstly, this thesis addresses various machine learning methods, their advantages, disadvantages, and applications in the pharmaceutical field. The understanding of machine learning methods is paramount to utilization of these methods in pharmaceutical formulation and process development. Secondly, this thesis is focusing on investigating the variation in solid form properties of a hydrophobic drug. Thousands of particles were characterized with a combinatory method based on both Raman spectroscopy and imaging. The large volume of data resulting from this experimental part was qualitatively and quantitatively explored using machine learning. Additionally, a risk-based approach and machine learning were implemented to evaluate the complex relationships between the formulation and process parameters of a model ACS. An interpretable machine learning approach based on influence networks was identified as an ideal solution. The influence network demonstrated the flexibility of multi-objective optimization based on Monte Carlo predictions.

The work presented realized the superior application of a risk-based approach combined with machine learning on the analytical-, formulation-, and process-related pharmaceutical challenges.