Lab-scale Machine Learning:
Tales of the good, the bad and the average

Prof. dr. dr. Danny E.P. Vanpoucke

QuATOMs group, Institute for Materials Research IMO-IMOMEC, UHasselt

Machine Learning and Artificial Intelligence are presented as the fix-all for current day problems. Also in research it is experiencing a golden age. However, before a Machine Learning model can be created, an enormous quantity of training data needs to be generated. This stands in stark contrast to general academic and industrial lab-scale data sets resulting from research projects. The latter give rise to small or even extremely small data sets (< 50 samples).

This makes many of us wonder: "Is it possible to train an ML model with 30 samples instead of 30.000.000?" Using some simple regression models, I'll show that these can be successful in creating a suitable model in the (very) small data regime. Real life use-cases considering adhesive coatings, solvable inks, and spray-coating are discussed. I'll present a strategy to always obtain the best model and highlight caveats and ways to deal with them.

[1] *"A machine learning approach for the design of hyperbranched polymeric dispersing agents based on aliphatic polyesters for radiation curable inks",*
Danny E. P. Vanpoucke, Marie A. F. Delgove, Jules Stouten, Jurrie Noordijk, Nils De Vos, Kamiel Matthyssen, Geert G. P. Deroover, Siamak Mehrkanoon, and Katrien V. Bernaerts
Polymer International , (2022) [doi: 10.1002/pi.6378]

[2] *"Small Data Materials Design with Machine Learning: When the Average Model Knows Best",*
Danny E. P. Vanpoucke, Onno S. J. van Knippenberg, Ko Hermans, and Katrien V. Bernaerts, and Siamak Mehrkanoon
J. Appl. Phys. 128, 054901 (2020) [doi: 10.1063/5.0012285]