Machine learning

Generate Data

- Crystallographic Information
- Structural descriptors
- Other descriptors (chemical, topological, thermodynamic, etc.)
- Figures of merit for **training set**

Assess Data

- Correlation analysis
- Data visualisation
- Draw qualitative conclusions
- Similarity and diversity study
- Creating **diverse training set**

Train Model

- Several configurations of descriptors
- Several machine learning algorithms (MLR, RF, SVM, NN, etc.)
- Hyperparameter optimisation
- **Bias *vs*. Variance**: Test set / Train set performance metrics

Evaluate model

- Cross-validation or direct assessment to test **performance** on unseen data
- Comparison to simulation methods
- Interpretability *vs.* Accuracy
- Transferability to new dataset

Data Analysis

Structure-property

- Quantitative structure-property relationship (QSPR) using ML
- **Design rules** of high-performing materials

Descriptor analysis

- **Feature importance**: the key explanatory descriptors
- Principal component analysis (PCA)

Interpretation

- Interpretable models
- Global Model-Agnostic Methods
- Local Model-Agnostic Methods
- Neural Network interpretation
- Use the above methods to **explain** ML

Relation to theory

- **Physico-chemical theories** helps understand the previous analysis
- More detailed computational methods can be used on specific case

HTP Screenings

Application range

- Define the **conditions** of application of the ML model
- Class of materials, adsorbent type
- Targeted industrial applications

Transferability

- Assess the **transferability** outside the application range
- Possibility to use transfer learning

Theoretical limits

- Fast screening of larger dataset gives more reliable theoretical limits of performance
- Informs on the **limits of the material design process**

Materials discovery

- Fast screening of new database can unveil **new best-performing** materials
- Further theoretical studies can be performed to confirm
- Experiments can finally concretise the applicability of such material