POSTER 30

CMAC FUTURE MANUFACTURING RESEARCH HUB

Prediction of mefenamic acid crystal shape by random forest classification

Siya Nakapraves[1,2], Monika Warzecha[1,2], Chantal Mustoe[1,2], Alastair J. Florence[2]
siya.nakapraves@strath.ac.uk

[1] Strathclyde Institute of Pharmacy and Biomedical Science (SIPBS), University of Strathclyde, UK
[2] EPSRC Future Manufacturing Research Hub for Continuous Manufacturing and Advanced Crystallisation (CMAC), University of Strathclyde, Technology and Innovation Centre, UK

Research Aims

- Research problem: Crystal shape is one of the key attributes affecting the bulk particle properties of a crystalline material as well as its downstream manufacturability[1]. However, the prediction of experimental crystal shapes remains very challenging.
- This research aims to explore the potential application of machine learning algorithms to solve this problem.

Figure 1. The expected outcome of the research

Research Methodology

Solubility Test

- Using a multi-reactor crystallisation platform (Crystal16): Measure real-time turbidity of sample solutions at different concentrations and determine solubility from clear points (%transmission = 100%) at specific temperature.

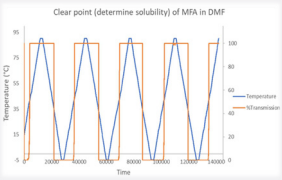


Figure 2. Crystallisation conditions used in Crystal16 for solubility analysis

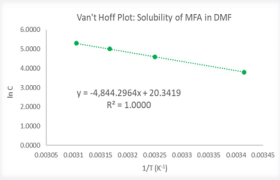


Figure 3. Van't Hoff plot of mefenamic acid in ethanol

The solubility curves were then used to calculate the degree of supersaturation for subsequent crystallisation screening.

Cooling crystallisation screening

- Sample solutions were prepared at different initial concentrations.
- The samples were heated on a hot plate to get clear solutions and then left in an incubator at 25 °C without disturbance until crystals occur.
- The crystal were observed under an optical microscope.
- Crystal images from 261 samples were collected and classified into either polyhedral crystals, needles, or no crystals.

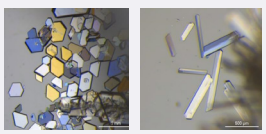


Figure 4. Polyhedral crystals

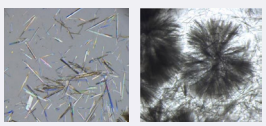


Figure 5. needle crystals
(The right-hand side picture shows needle crystal agglomerates or spherulitic crystals)

Random Forest Classification Models

- 87 models were prepared differently.
- The first 3 models were evaluated by 4-fold cross-validation, while the others were evaluated by using the data in which the samples were crystallised from 1 specific solvent as the test data.

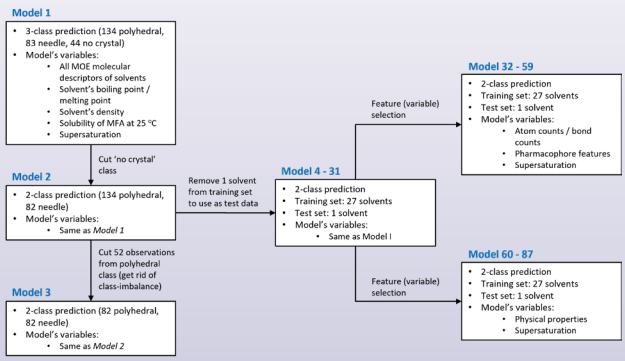


Figure 6. Diagram showing the dataset and variables used in each models

Results & Discussions

Model Evaluation

Model performance using crystal shape observations from all solvents in the training set

Three random forest classification models were built initially to determine the efficacy of this method and understand the extent to which the class imbalance present in the dataset would affect prediction accuracies.

Table 1 Model evaluation by 4-fold cross-validation of Models 1, 2 and 3. (SD = standard deviation)

Model	Model 1 (3 classes)	Model 2 (2 classes w/ class-imbalance	Model 3 (2 classes w/o class-imbalance)
Accuracy by 4-fold cross validation	82.4% (SD = 3.1%)	93.5% (SD = 2.1%)	93.3% (SD = 5.3%)

The result shows that *Model 1*, which predicts 3 crystal outcomes, has the lowest accuracy compared to the models predicting 2 crystal outcomes. Considering *Model 2* and *Model 3*, the model accuracies indicate that the class imbalance observed in *Model 2* did not noticeably affect the model performance. Consequently, the dataset used in *Model 2 will be used for the further models discussed below.*

Prediction of crystal shape from solvents not included in the training set

- To determine the ability of the models to predict crystal shapes from solvents for which no data was present in the training set, the observations for a single solvent were removed from the training data of each model (*Model 4 – 31*).
- Feature selection: *Model 32 – 59* and *Model 60 – 87* (see details in **Figure 6**)
- The performance accuracy was then assessed using the crystal shapes for the observations that were excluded from the training data.

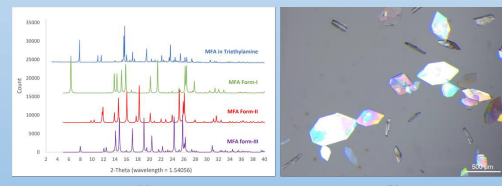
Table 2 Number of the models in which the prediction accuracy was 100% and less than 50%

Prediction accuracy	Numbers of models grouped by prediction accuracy		
	All solvent descriptors	Atom counts/bond counts + pharmacophore features	Physical properties
100%	10	10	12
Less than 50%	12	11	8

In total, 32 out of 84 models predicted the shape of mefenamic acid crystals with 100% accuracy, and the models using only physical property descriptors and supersaturations as the model variables resulted in the best overall prediction accuracy across all solvents.

Characterisation of mefenamic acid crystals grown in triethylamine

Further crystal characterisation was done for the crystals grown in solvents with the models showing low prediction accuracy. All samples were consistent with mefenamic acid form-I except the sample crystallised from trimethylamine which exhibited a notably distinct XRPD pattern (**Figure 7a**) and the crystals observed under a microscope also exhibit thinner flat plates than the plate crystals of mefenamic acid form-I crystallised from the other solvents (**Figure 7b**)



(a) (b)

Conclusion

RF classification can be a useful tool to predict the experimental crystal shape of MFA. Our two-class RF classification model resulted in 93% of prediction accuracy. For solvents that were excluded from the training set, the model which expressed poor performance also suggested the possibility to discover a new solid form of crystals.

Whilst demonstrated only for mefenamic acid it is expected that with the appropriate data, the application of this tool can be broadened to cover a wider range of active ingredient molecular and crystal attributes.

References

1. Docherty, R., et al. Application of bravais-friedel-donnay-harker, attachment energy and ising models to predicting and understanding the morphology of molecular crystals. *J. Phys. D. Appl. Phys.* **24**, 89–99 (1991).

Acknowledgement

Many thanks to Dr. Monika Warzecha, and Dr. Cameron Brown, EPSRC Future Manufacturing Research Hub for Continuous Manufacturing and Advanced Crystallisation (CMAC), University of Strathclyde, Technology and Innovation Centre, for their expertise and kind support along this work.