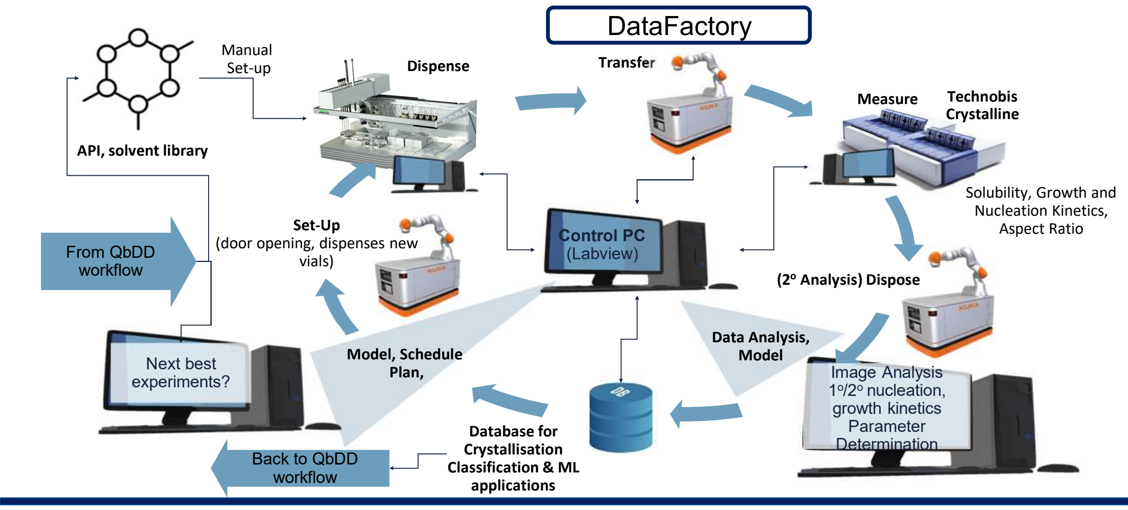
Data-Driven Approaches for the Classification of Crystallisation Outcomes from In-line Images

Parandeep Sandhu[1 & 2], **Christos Tachtatzis**[2], **Christopher Boyle**[1 & 3] **and Javier Cardona**[1, 2 & 3]

[1] **EPSRC Future Manufacturing Research Hub for Continuous Manufacturing and Advanced Crystallization (CMAC), Glasgow, UK.**
[2] **Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK.**
[3] **Department of Chemical and Process Engineering, University of Strathclyde, Glasgow, UK.**

DataFactory



Motivation

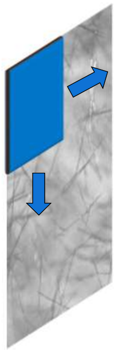
- Implementing a data-driven approach that can automate the classification of different phases of high-throughput crystallisation processes can be a powerful tool to **[1][2]**:
  - ☐ Predict the quality and consistency of the final pharmaceutical product
  - ☐ Optimise manufacturing processes
  - ☐ Reduce waste
- Advanced analytical techniques and machine learning algorithms can be leveraged to extract relevant features on crystallisation outcomes from in-line images.

[1] Borsos, A., Szilágyi, B., Agachi, P. S. & Nagy, Z. K. *Org Process Res Dev* 21, 511–519 (2017).
[2] Silva, A. F. T. et al. Eur J Pharm Biopharm 85, 1006–1018 (2013).

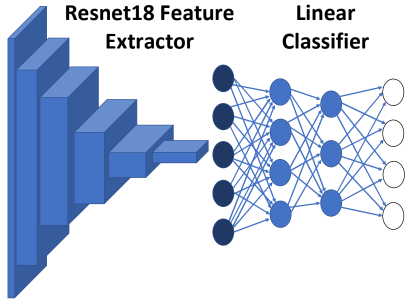
Method

- Classifying crystallisation outcomes from in-line images.
- Deep learning model comprised of pre-trained ResNet-18 feature extractor and linear classifier.
- 10 classes are categorised using 80% of images pair class for training and 20% for validation.
- Augmentations (image rotation, contrast and brightness change) applied to increase diversity, improving generalisation.

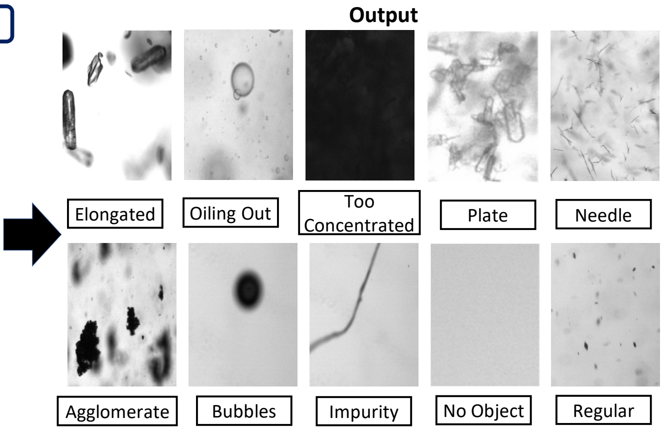
Input Images



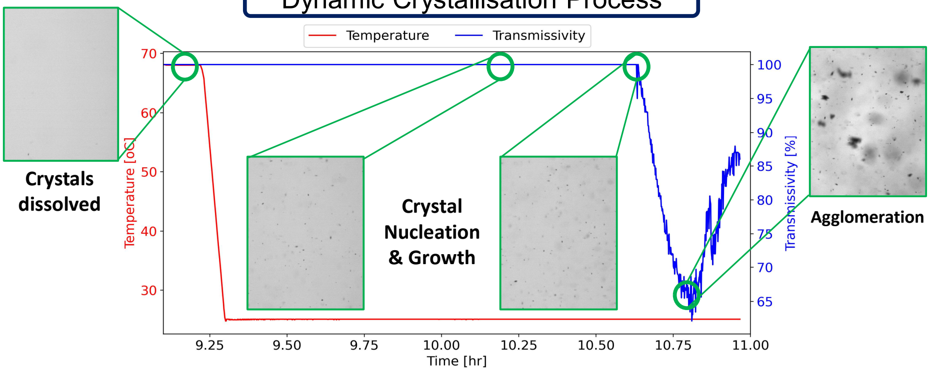
Deep Learning Model



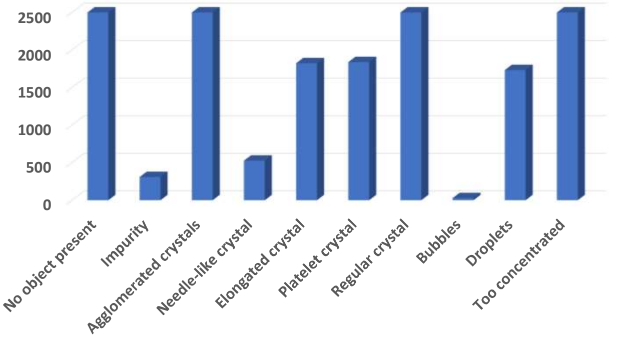
Output



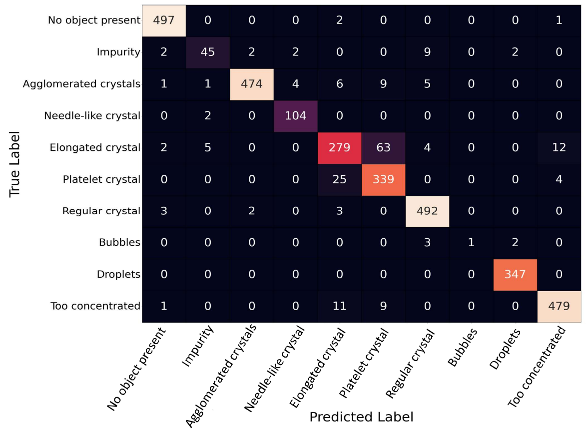
Dynamic Crystallisation Process



Available Data



Results



True Label \ Predicted Label	No object present	Impurity	Agglomerated crystals	Needle-like crystal	Elongated crystal	Platelet crystal	Regular crystal	Bubbles	Droplets	Too concentrated
No object present	497	0	0	0	2	0	0	0	0	1
Impurity	2	45	2	2	0	0	9	0	2	0
Agglomerated crystals	1	1	474	4	6	9	5	0	0	0
Needle-like crystal	0	2	0	104	0	0	0	0	0	0
Elongated crystal	2	5	0	0	279	63	4	0	0	12
Platelet crystal	0	0	0	0	25	339	0	0	0	4
Regular crystal	3	0	2	0	3	0	492	0	0	0
Bubbles	0	0	0	0	0	0	3	1	2	0
Droplets	0	0	0	0	0	0	0	0	347	0
Too concentrated	1	0	0	0	11	9	0	0	0	479

	Precision (%)	**Recall (%)**	**F1-Score (%)**
No object present	98.2	99.4	98.8
Impurity	84.9	72.6	78.3
Agglomerated crystals	99.2	94.8	96.9
Needle-like crystal	94.5	98.1	96.3
Elongated crystal	85.6	76.4	80.8
Platelet crystal	80.7	92.1	86.0
Regular crystal	95.9	98.4	97.1
Bubbles	100.0	16.7	28.6
Droplets	98.9	100.0	99.4
Too concentrated	96.6	95.8	96.2
Macro-average	**93.5**	**84.4**	**85.8**
Weighted-average	**94.1**	**93.9**	**93.9**

Conclusion

- The model achieves a macro-average and weighted-average for F1-score greater than 85% and 93%, respectively.
- The model has distinguished the majority of classes very well however, further data is needed to allow more variety in training as there is a lack of data for some classes.
- This approach is not limited to the Technobis Crystalline and can be applied to any sensor capable of taking microscopic imaging of crystals.

Engineering and Physical Sciences Research Council