

# The Detection and Characterisation of Molecular Crystals using Machine Learning

Malcolm Ramsay

School of Chemistry, The University of Sydney

## Abstract

The wide range of crystal structures of a molecular crystal can make it difficult to detect and characterise each of them individually. Presented, is a general approach to the detection and characterisation of local crystalline order using machine learning. This approach is able to categorise different crystalline orderings and the liquid phase with an accuracy in excess of 95% over a wide temperature range.

## Introduction

The molecule we are studying has three distinct crystal structures with potential energies all within 2% of each other.

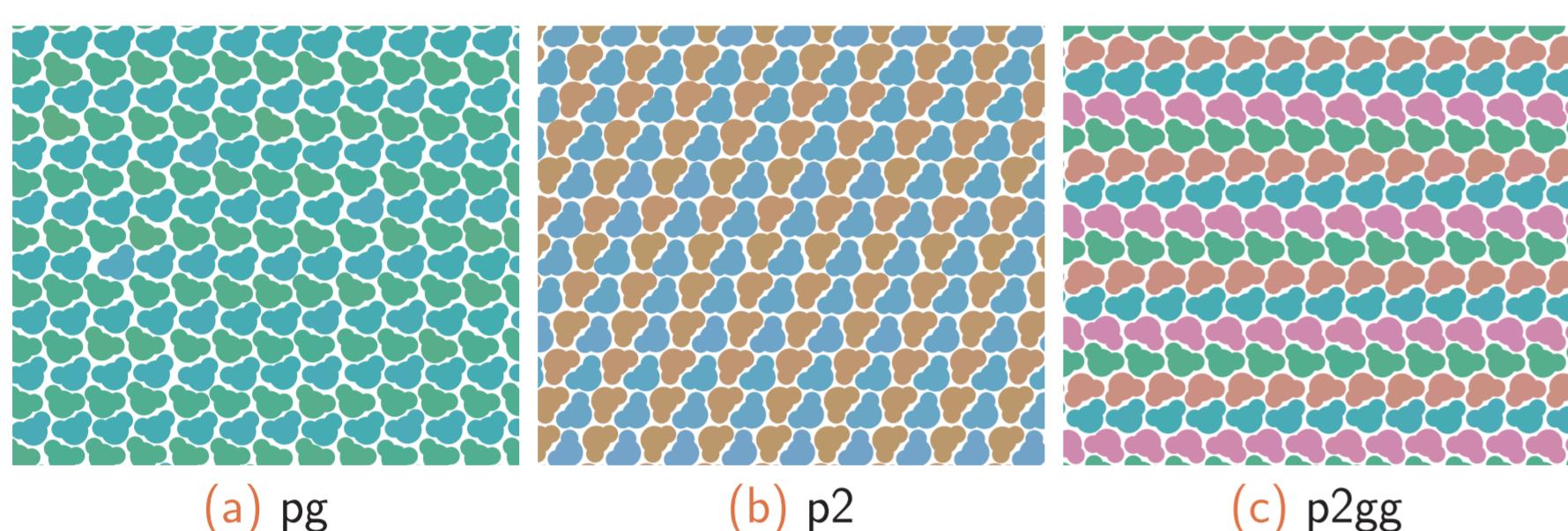


Figure: The three lowest energy structures of the Trimer molecule.

A complete understanding of the liquid–solid phase transition requires the detection and characterisation of all crystal structures. The diversity of structure required metrics and parameters tailored for each structure.

## Machine Learning Methodology

There has recently been work using machine learning to categorize fcc and bcc crystal structures<sup>1,2</sup> for single atom potentials. No comparable work was found for molecular crystals. The relative orientation of neighbouring molecules was chosen as the feature to distinguish crystal structures over a range of temperatures.

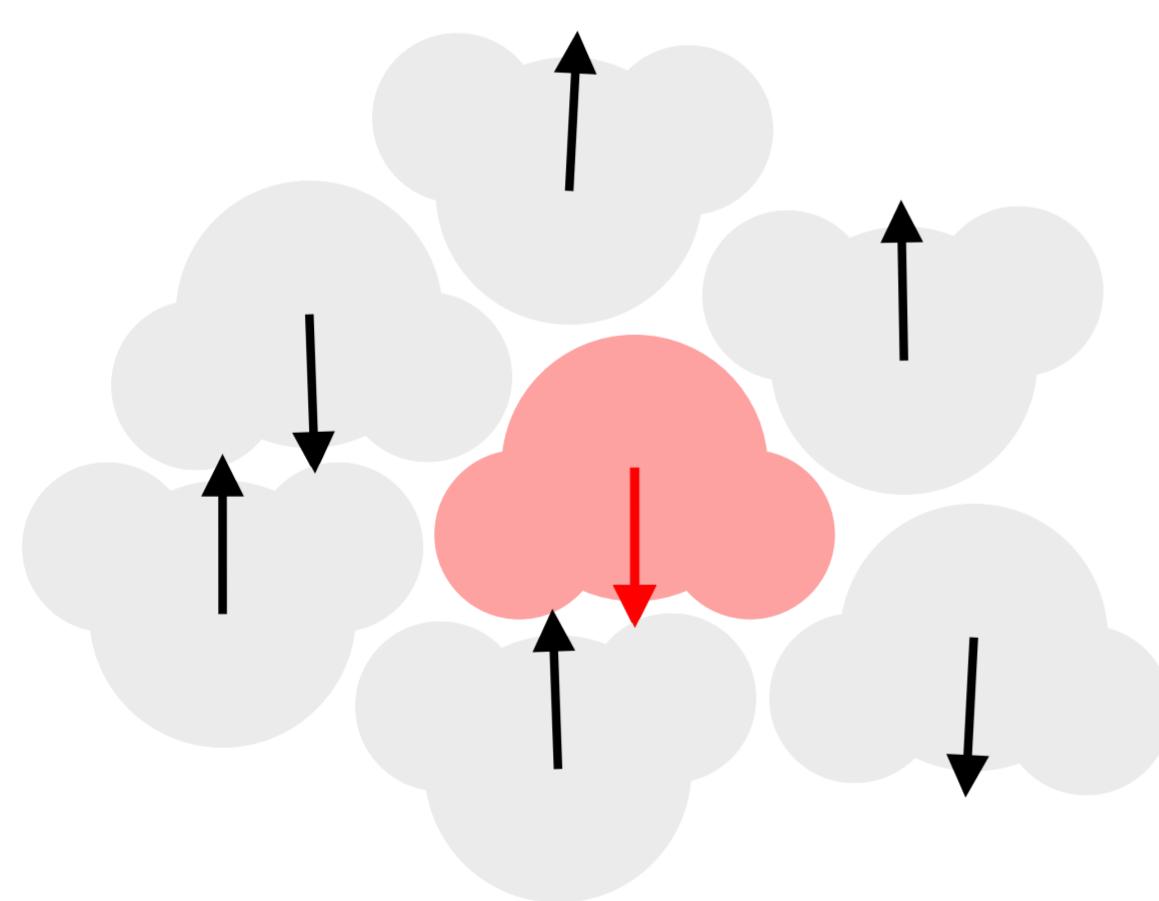


Figure: The features of a molecule (in red) are the relative orientations of the nearest neighbours (grey).

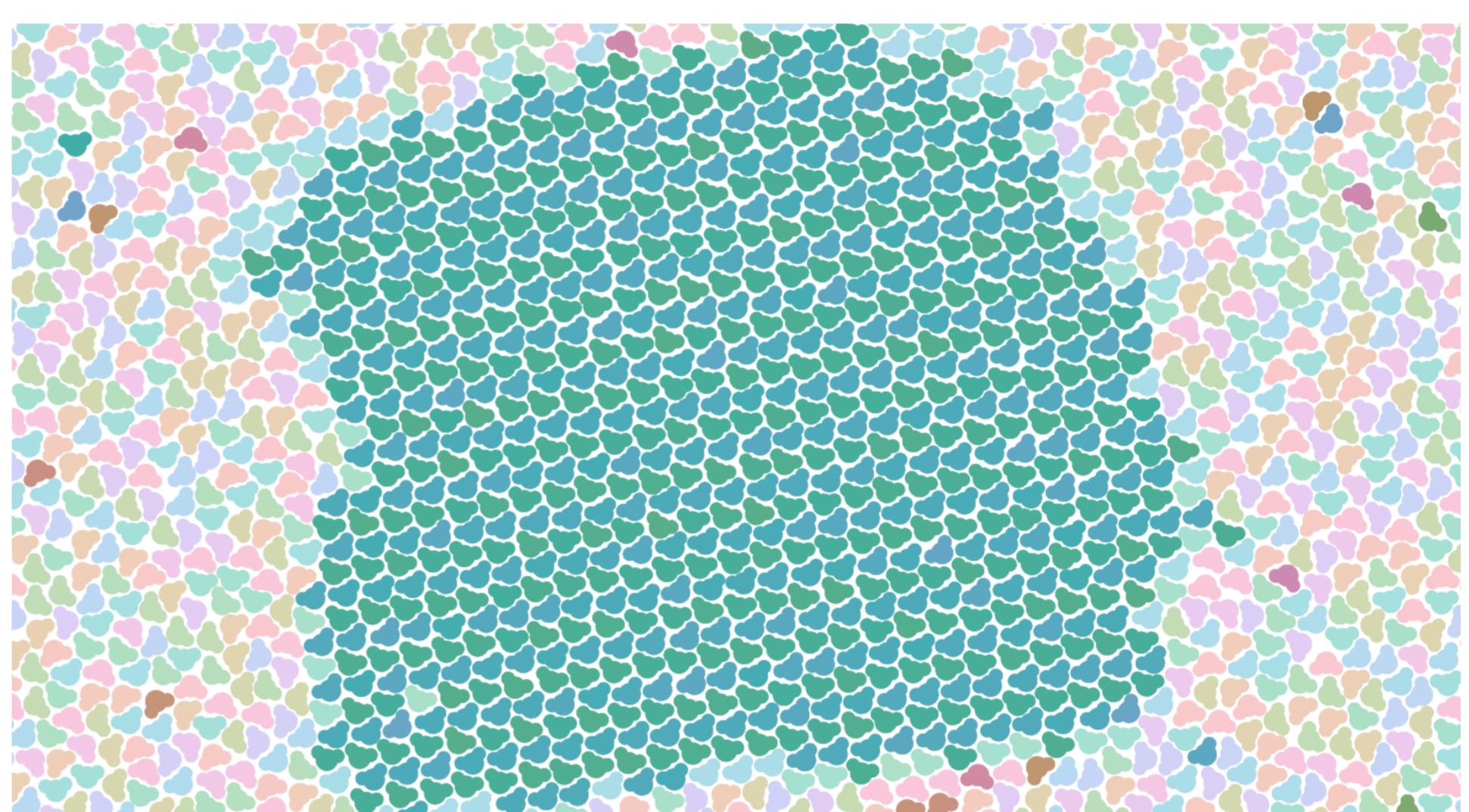
Machine learning algorithms from the scikit-learn<sup>3</sup> library were used for classification. The K-nearest neighbours algorithm gave the best results for this work.

## References

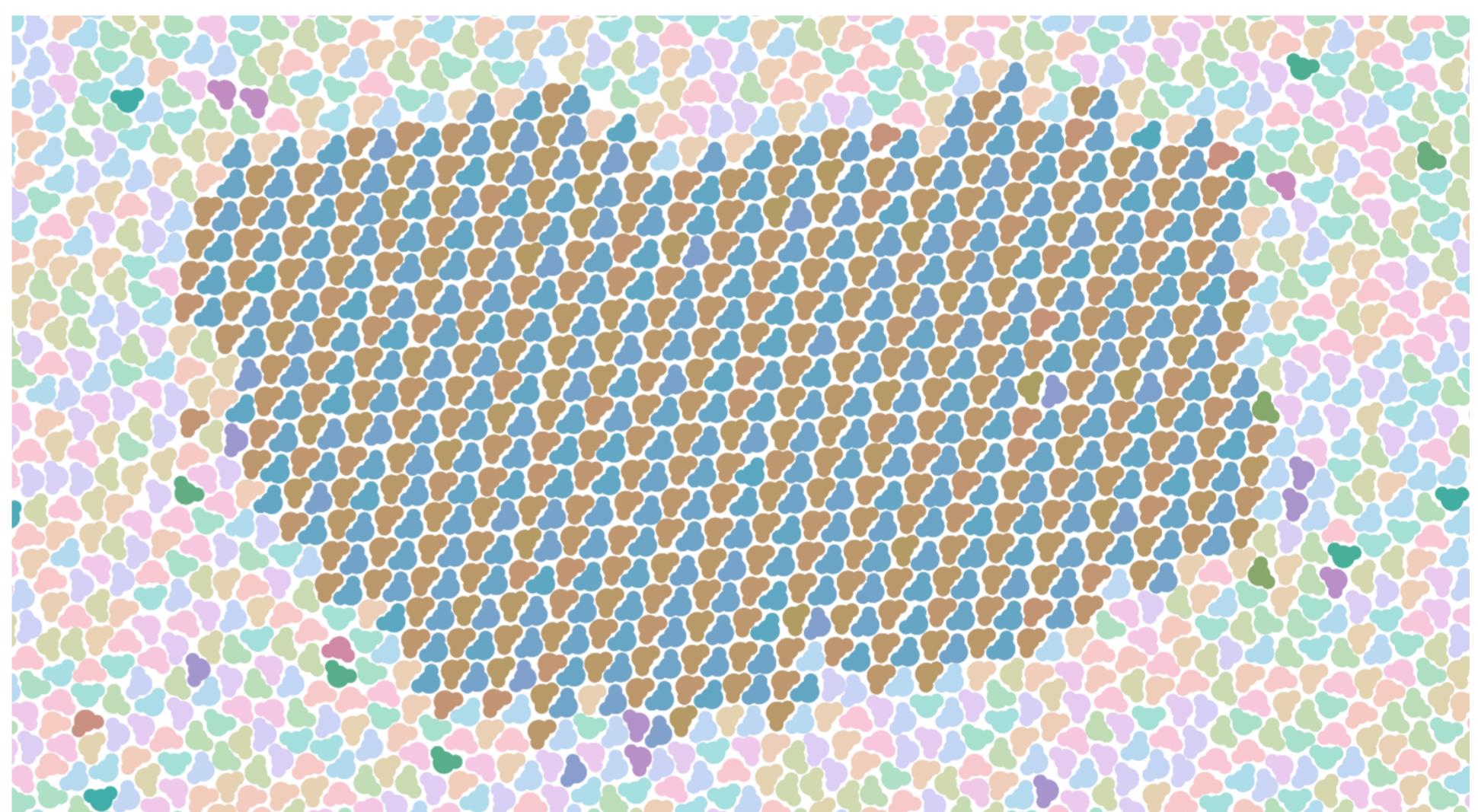
- <sup>1</sup>W. F. Reinhart, A. W. Long, M. P. Howard, A. L. Ferguson and A. Z. Panagiotopoulos, "Machine learning for autonomous crystal structure identification", *Soft Matter* **13**, 4733–4745 (2017).  
<sup>2</sup>C. Dietz, T. Kretz and M. H. Thoma, "Machine-learning approach for local classification of crystalline structures in multiphase systems", *Phys. Rev. E* **96**, 1–5 (2017).  
<sup>3</sup>F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: machine learning in Python", *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

## Results

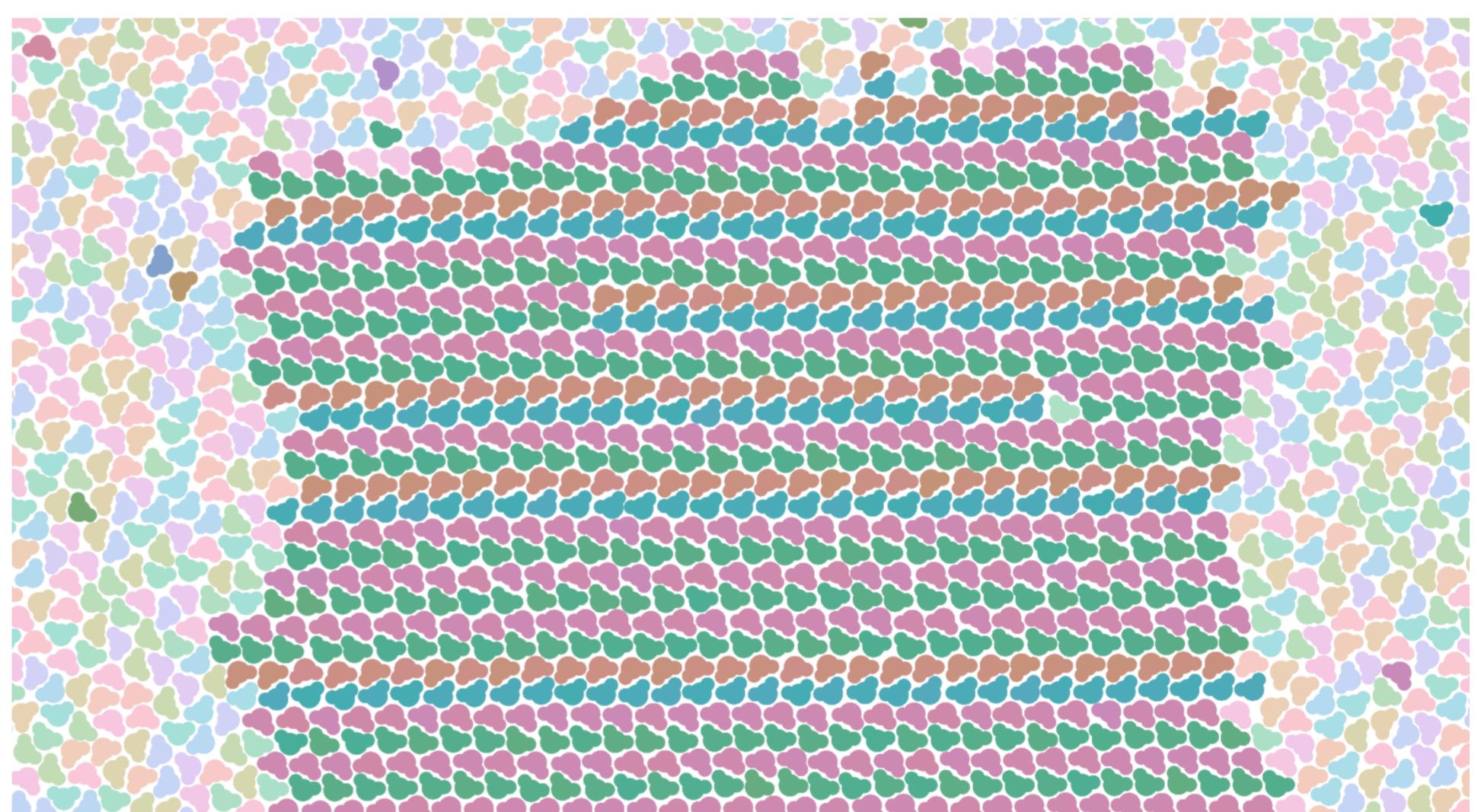
A single classification algorithm can be used to monitor melting of all crystals with minimal classification errors or noise from thermal fluctuations.



(a) pg



(b) p2



(c) p2gg

Figure: Each of the crystal structures characterised using the same machine learning algorithm. Regions classified as crystalline are darker than those classified as liquid.

Note that most of the single molecules classified as crystalline in the liquid have a local environment that matches one of the crystals.

## Future Work

This work could easily be extended to a wide range of crystal structures using the relative orientations, a scaled distance, and angle to neighbours. This extension requires finding the crystal structures of a number of 3D molecules to generate the training data.

Further work would incorporate this functionality into a simple to use open source library so that this can become a standard analysis tool for many types of crystals.