



Science and
Technology
Facilities Council

Scientific Computing

Using AI and Machine Learning to Speed-Up Drug Development

A neural network algorithm developed by STFC Scientific Computing's AI for Science theme is being used by CrystalGrower Ltd, a spin out from the University of Manchester.

CrystalGrower clients in fields from pharmaceuticals to agrochemicals can now use the machine learning tool to analyse thousands of digitally simulated crystals faster and more accurately. This is enabling scientists to reach research insights in fields including drug development, chemicals and foods, far quicker than was previously possible.

The Challenge

CrystalGrower Ltd. is able to generate an infinite number and shapes of synthetic 3D crystals across many chemical systems for research purposes. However, analysing the results requires more scientists with specialist domain expertise than are available, so CrystalGrower approached STFC Scientific Computing to devise a novel way to auto cluster the crystals.

CrystalGrower scientists are active in areas from pharmaceuticals to agrochemicals and mining, and are actively using the synthetic crystals for studies in fields including paracetamols, glycine and other drug molecules.



CrystalGrower



Our Approach

CrystalGrower chose to partner with STFC Scientific Computing (SC) to be able to access their extensive Artificial Intelligence (AI) and Machine Learning (ML) skills. SC provides ML and AI analysis services to all of STFC's experimental facilities, as well as to external organisations.

“We chose to work with SC’s AI for Science Theme for its deep experience in the fundamental science side, as well as its impressive expertise in the fields of artificial intelligence and machine learning as applied to science. This has been a successful and productive collaboration.”

Dr. Adam Hill, Chief Scientific Officer, CrystalGrower Ltd

SC's AI for Science theme created a simulation software using machine learning and neural networks, to cluster chemical simulation systems comprising thousands of crystals for analysis purposes.

The automatic clustering devised by the SC team uses what's known as a variable autoencoder architecture – which consists of an encoder and decoder. The original input data is provided to the encoder which compresses the information as much as possible, from 10,000 or so datapoints, to 3 or 4 features per crystal, using X Y datapoints to identify each crystal's physical characteristics - such as size, texture and shape.

The features are critical, as they describe the physical characteristics of each crystal. Without this information, the computer cannot assess whether a crystal has a particular shape or compare it to another. The datapoints must, therefore, be converted into features that capture physical characteristics the computer can understand

The work takes complex, high-dimensional information and compresses it into a concise, low-dimensional form using an encoder. The decoder, which is the counterpart to the encoder, then reconstructs the original information from this compressed version.

“Using the decoder, we reconstruct the original information. A good reconstruction indicates that the compressed version contains concise and meaningful information from the original data.”

Dr Jaehoon Cha, Senior Data Scientist, STFC Scientific Computing

In the drug cycle, for example, the software could be used by scientists to study 2D projections of proteins and to improve understanding of the behaviour of proteins. The 3D proteins are too small to be visualised directly, but the auto clustering could be used to cluster the crystals by different shapes or colours and to build a good reconstruction of a 3D protein shape based on 2D analysis.



Benefits

Scientists can now use the SC neural network algorithm software powered by artificial intelligence in their research to better understand, control and modify crystals for developments in food, pharmaceutical and industrial fields.

The software clusters the crystals by shape easily and quickly, saving scientists at CrystalGrower considerable time distinguishing shapes by hand, an unmanageable task with sample sizes of 16,000 or more. Using the AI speeds up processing time and also reduces human error.

“Our software not only identifies the shape of crystals, it also helps us to understand how crystal shape changes in response to growth conditions. The shape affects how well the crystal works, so in drug development, using our algorithm to recognise the crystal shapes required for specific medicines could support more efficient exploration, reduce experimental effort, and potentially save significant time and cost.”

Dr Jaehoon Cha, Senior Data Scientist, STFC Scientific Computing

90% of CrystalGrower clients are pharmaceutical companies, working on developments across areas of drugs. The algorithm will allow future drug and other science developments to progress faster.

“The software is a great addition to our research infrastructure. It gives our scientists insights much quicker, such as identifying crystal shapes, and how they relate to one another.

“It adds the power of artificial intelligence and machine learning to our software – these kinds of data-driven insights are critical to the future of science.”

Steven Tendyra, PhD Researcher in Modelling Crystal Growth, University of Manchester

The Future

This pilot project with CrystalGrower was funded by Innovate UK AI4S. Since the method developed by the AI for Science Theme is independent of the software used to generate the crystal growth data, it can be extended to other studies involving three-dimensional data analysis.

STFC Scientific Computing hopes to secure additional funding to introduce some uncertainty quantification aspects to the software and is currently considering other analysis projects where they might apply this type of artificial intelligence and machine learning work.



Science and
Technology
Facilities Council

Scientific Computing

Read the associated research paper:

**Disentangling autoencoders and spherical harmonics for
efficient shape classification in crystal growth simulations**

[nature.com/articles/s42005-025-02129-7](https://www.nature.com/articles/s42005-025-02129-7)

Crystal Grower

crystalgrower.org

Contacts for more information:

Dr Jaehoon Cha: jaehoon.cha@stfc.ac.uk

Dr James Gebbie-Rayet: james.gebbie@stfc.ac.uk

Dr Jeyan Thiyaalingam: t.jeyan@stfc.ac.uk