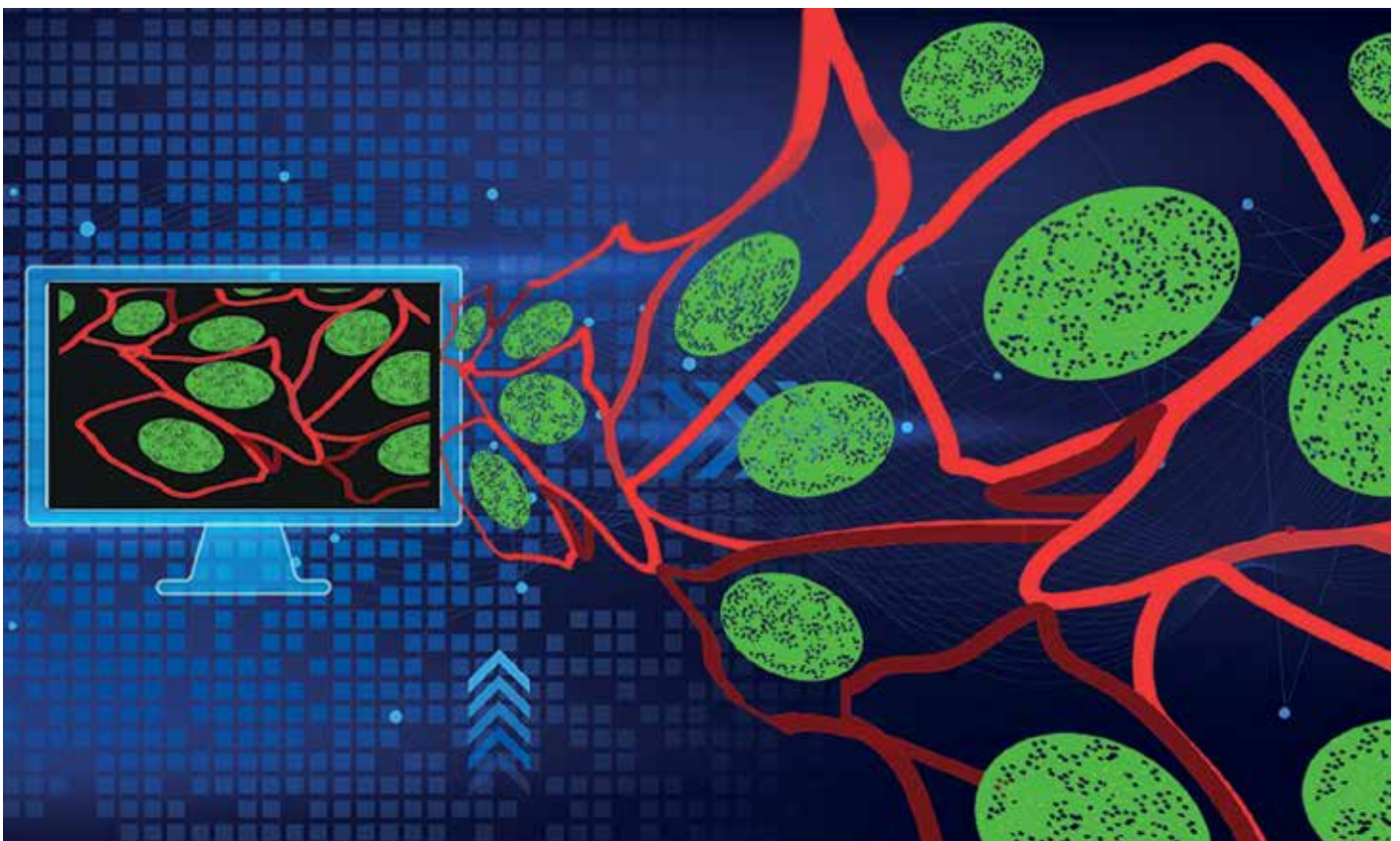


Efficient transfer and analysis of biological image data through web interfaces

Researchers can transfer their computing from Google's Colab to CSC's environment using a new application. A similar approach can be adapted if one wants to move from one supercomputing environment to another.



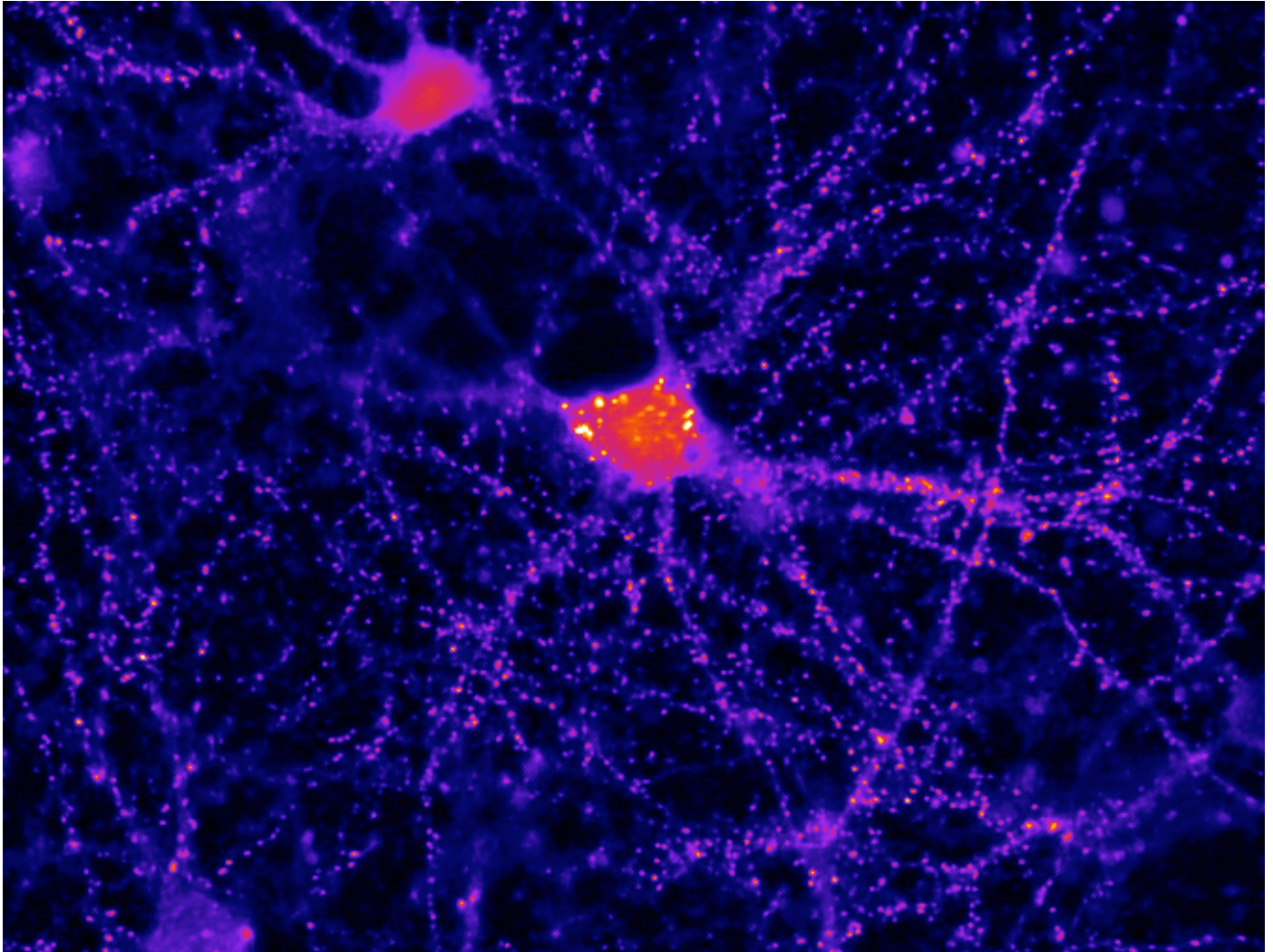
Application Specialist **Laxmana Yetukuri** at CSC – IT Centre for Science and Specialist Researcher **Michael Courtney** at Turku Bioscience Centre have customised GPU-empowered notebooks originally developed by the Centre's team, that can apply deep learning models for biological image data in the CSC supercomputing environment. There is also an open-source toolkit, ImageJ/Fiji, for the deep learning models in microscopy and the usage of the toolkit is under exploration at CSC, the Finnish node of the ELIXIR infrastructure.

Turku Bioscience Centre has been using Google's Colab Notebook cloud ser-

vice to analyse and visualise data. Large scale data-intensive research activities, however, have their limitations in using Google services in free-of-cost models. Researchers often require significantly more storage and computing capacity for data processing than basic users. CSC's supercomputing environment provides vastly superior storage and computing capacity free-of charge for academic users, and now it's possible to make the switch from Google Colab environment to CSC environment. A personal laptop/PC can easily access CSC's supercomputing environment through a web browser, thanks to recently

developed user-friendly web interfaces to supercomputers at CSC.

Researchers can transfer their computing from Google's Colab to CSC's environment using an in-house developed container wrapper. The container wrapper allows researchers to define a standardised environment in which they can run scientific software. The program's code, along with its libraries and settings, is placed within the container. Once the software and its dependent packages are installed as part of the wrapper tool, other users can start using the application without any pre-installations. A similar approach can be adapted



An example image of a neuron that has a labelled form of the protein, captured by the automated microscope. The protein distributes in a rather specific manner, where the SynGAP1 protein forms puncta of concentration along processes, probably representing synaptic contacts, the signal is highlighted bright purple or orange.

if one wants to move from one supercomputing environment to another.

“We make the work of researchers easier by providing easy-to-follow instructions for the installation of user’s custom notebooks. Once a project member installs an application in their project area, other researchers don’t need to install any software to use custom notebooks – they can start working immediately. Accessing CSC’s notebooks in supercomputing environment requires just a few clicks on the web interface (www.puhti.csc.fi),” says Yetukuri.

“Biological image analysis typically needs larger disk space for storing image

data. CSC object storage ALLAS provides a good storage environment. The computing environment can be accessed only with a user account obtained from CSC,” Yetukuri says.

Yetukuri and Courtney were able to utilise the existing CSC infrastructure by customising a Google Colab notebook to CSC’s computing environment to analyse microscopic image data. The custom notebooks were used to build machine learning models using microscopy image data. The notebooks were accessed through the Puhti web interface. The researchers are now exploring containerised deployment of im-

aging-related software, Fiji/ImageJ in CSC supercomputing environment to perform downstream analysis.

Biological imaging and image data analysis use algorithms to extract a significant amount of quantitative information from the images. This information can be used for pattern recognition and classification of image data, providing biologically significant insights. Using the system, Yetukuri and Courtney aim to develop and apply machine learning models for identifying SYNGAP1 gene variants causing brain disorders, and, in the future, for drug screening.

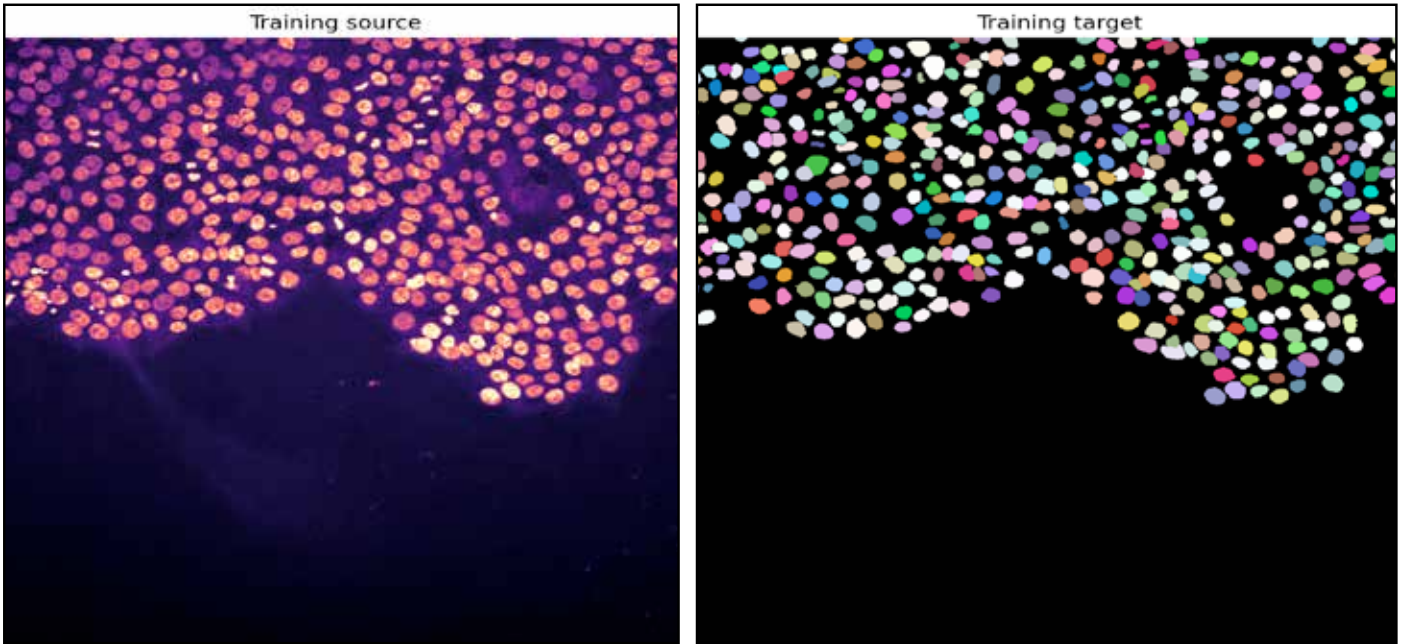


Image shows epithelial cells, as example data to demonstrate development and application of machine learning models. (Image: training data of ZeroCostDL4Mic-team who developed the Google Colab notebooks).

Assistance in research on disorders of brain development

Courtney, Co-principal Investigator Li-Li Li and their colleagues at Turku Bioscience Centre are investigating disease-causing variants of SynGAP1 proteins. The SynGAP1 gene is located on the sixth chromosome and produces the SynGAP protein. The protein regulates synapses, the junctions through which nerve cells communicate with one another. A variant of the SynGAP1 gene causes the production of SynGAP protein to drop below a sufficient level. This leads to abnormal communication between nerve cells, leading in turn to various neurological disorders. In order to develop in a normal way, the brain requires two correct genes that encode the SynGAP1 protein. Mutations can lead to one of these not being expressed, resulting in developmental delays.

SynGAP1 encephalopathy is an early-onset intellectual disability. The developmental delay characteristic of the condition is typically observed during the first or second year of life. Additionally, about eight out of ten encephalopathy patients are diagnosed with epilepsy. The symptoms of epilepsy vary individually, and it can be

difficult to treat. Behavioural disorders and autism occur in half of the patients.

Turku Bioscience Centre's microscopy screening unit analyses normal SynGAP1 genes and point mutations that may impair protein function, sometimes to nearly non-existent levels. Point mutations only change one amino acid in the protein, but the consequences of this require further clarification.

The SynGAP1 protein is present only in nerve cells. Using high-throughput microscopes, it is possible to simultaneously examine 384 living neuron circuits over time. In the neurons SynGAP1 is labelled with a fluorescent protein tag that can be detected with frequency-tuned light. In each circuit normal SynGAP1 or a different pathogenic form can be studied. Aberrations in protein function can be observed based on the images.

This approach is valuable for future arrayed drug screens

The microscopes perform automatic image capture and can sample changes in circuits every 20 seconds. When studying different variants of the protein, it is possible

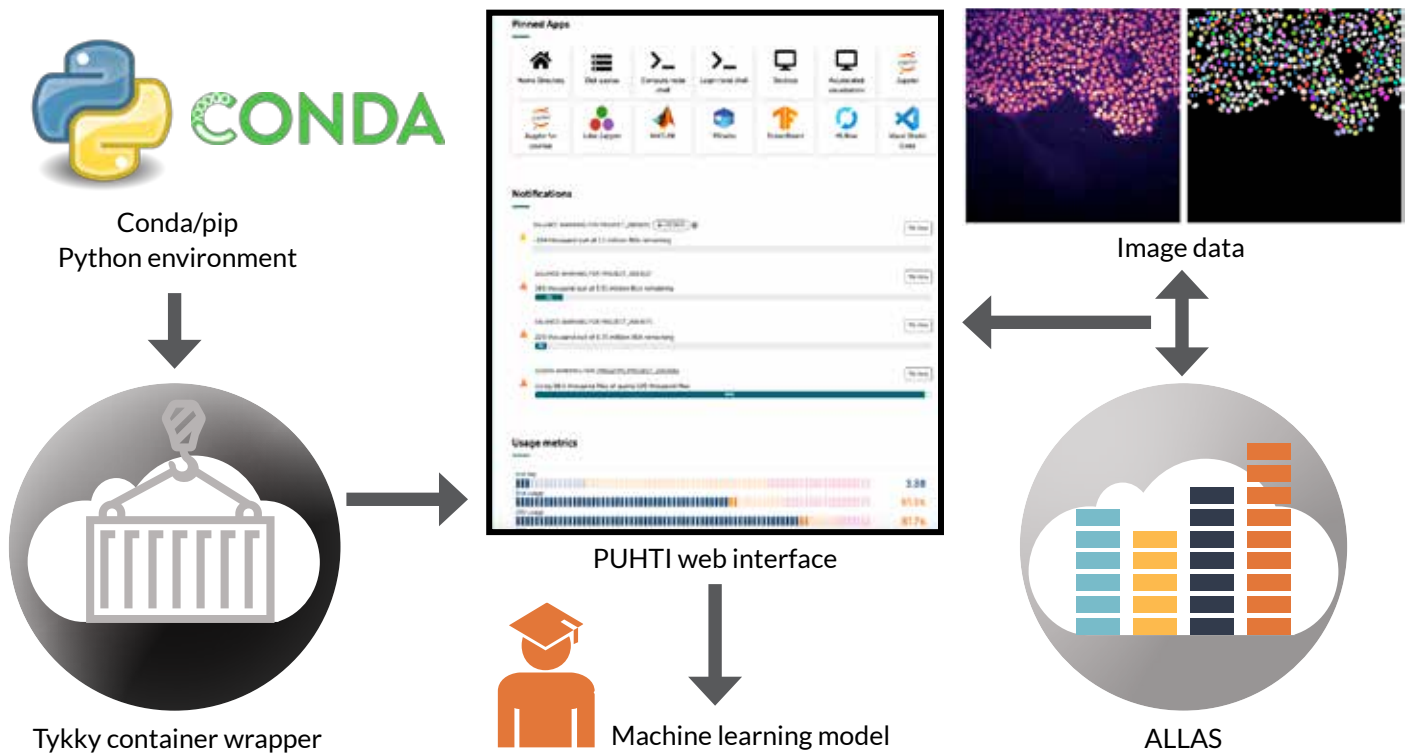
to compare whether its function is normal, or whether is enhanced or has completely ceased, both of which could lead to pathology.

"We have been able to create experimental and analytic setups that investigate the functions of damaged SynGAP1 that deviate from the normal. This may potentially provide a pathway for drug screening in the future. There are also gene variants that are found in patients with the disease, but it's unknown how or even if they are actually causing it," Michael Courtney says.

"With our method, we can determine if these gene variants have a similar impaired function as known pathogenic variants."

Once the sample preparation requirements were satisfied, a key step was to develop a deep learning model that automates the identification of SynGAP1 puncta (a bunch of dots) usually located at the synapses, the sites of communication between the neurons. Puncta are discrete regions of images where the fluorescent tag is visible.

"Once these are identified, their number and some 25 properties of each puncta can be extracted. Once demonstrated, this approach will be extremely valuable for



future arrayed drug screens where each missense variant will be exposed to each of the up to 4,000 separate drugs in the screening library.”

According to Courtney only by testing drugs there is a hope to rapidly find a com-

pound that already has known clinical safety and tolerability data. This information is crucial to clinicians and a potential short-cut to achieving a benefit for patients.

“As we are studying a rare disease with so many different variants of an essential

protein, it is very difficult to carry out any kind of clinical trial to find effective drugs. Even generating animal models for this diversity of disease-causing variants is highly challenging.”

20.2.2023 | Ari Turunen

MORE INFORMATION:

Research is funded by the patient advocacy groups SynGAP Research Fund US and EU, and Leon and friends e.V.

FIRI

Article was supported by the Research council of Finland grant No: 345591 for ELIXIR European Life-Sciences Infrastructure for Biological Information (FIRI 2021)

A free and open-source notebook for Deep Learning in microscopy at CSC. Possibility to run Google Colab notebooks at CSC HPC environment via the web interface

GitHub: <https://github.com/yetulaxman/ZeroCostDL4Mic>

The story behind ZeroCostDL4Mic, or How to get started with using Deep Learning for your microscopy data

<https://focalplane.biologists.com/2021/01/18/the-story-behind-zero-cost-dl-4-mic-or-how-to-get-started-with-using-deep-learning-for-your-microscopy-data/>

Democratising deep learning for microscopy with ZeroCostDL4Mic

<https://www.nature.com/articles/s41467-021-22518-0>

Turku Bioscience Centre

<https://bioscience.fi>

CSC – IT Center for Science

is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned, centralised IT infrastructure.

<http://www.csc.fi>

<https://research.csc.fi/cloud-computing>

ELIXIR

builds infrastructure in support of the biological sector. It brings together the leading organisations of 21 European countries and the EMBL European Molecular Biology Laboratory to form a common infrastructure for biological information. CSC – IT Center for Science is the Finnish centre within this infrastructure.

<http://www.elixir-finland.org>

<http://www.elixir-europe.org>

ELIXIR FINLAND

Tel. +358 9 457 2821s • e-mail: servicedesk@csc.fi
www.elixir-europe.org/about-us/who-we-are/nodes/finland

www.elixir-finland.org

ELIXIR HEAD OFFICE

EMBL-European Bioinformatics Institute
www.elixir-europe.org