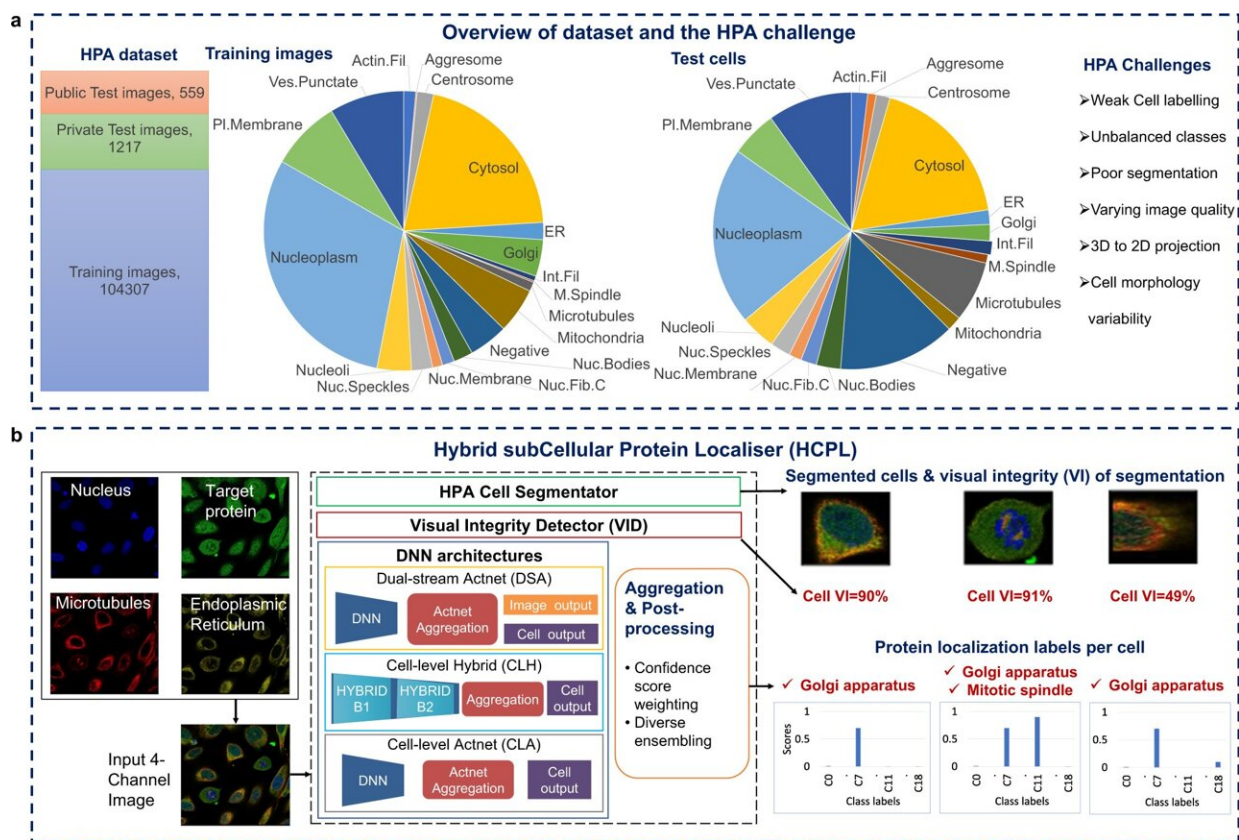


Newly developed AI shows speed and accuracy in identifying the location and expression of proteins

May 10 2023



Overview of the HPA dataset and the proposed solution. a The HPA dataset is the largest collection of images depicting specific protein localisations at a subcellular level, acquired using immunofluorescence staining followed by confocal microscopy imaging. The training dataset consists of 104307 images and corresponding image-level labels. To evaluate the system's performance, the test set comprises 1776 images of 41,597 single cells. The test set is divided into

a public test set (559 images) and a private test set (1217 images). The pie charts illustrate the numerical proportion of images and cells per class in the training and test set. Developing ML models for protein localisation is challenging due to issues from weak labeling, prevalent multi-label classifications, 3D-2D projection ambiguities, and severe class imbalance. b Each HPA image is represented by four channels, the nucleus (blue), the protein of interest (green), microtubules (red), and the endoplasmic reticulum (yellow). Our HCPL system takes 4-channel images as input and outputs segmented cells, protein localisation labels with associated probabilities, and the visual integrity scores for each cell. Experimental evaluation shows that HCPL achieves the classification performance of 57.19% mAP in single-cell analysis. Credit: *Communications Biology* (2023). DOI: 10.1038/s42003-023-04840-z

A new advanced artificial intelligence (AI) system has shown world-leading accuracy and speed in identifying protein patterns within individual cells. The new system, developed at the University of Surrey's Institute for People-Centered AI, could help scientists understand differences in cancer tumors and identify new drugs for diseases.

In a study published in *Communications Biology*, researchers demonstrate how the HCPL (Hybrid subCellular Protein Localizer) requires only partially labeled data to learn how to decipher the locations of proteins within cellular structures and their behavior in different cells.

The team tested the HCPL on the Human Protein Atlas and found it to be the most accurate tool for identifying the location of proteins within individual cells.

Professor Miroslaw Bober, leader of the HCPL project from the University of Surrey, said, "To understand how proteins work inside cells, scientists need to study where they are located, but this can be a time-consuming and complicated process. HCPL makes this process

easier.

"This program uses a [deep-learning](#) model to quickly and accurately identify subcellular structures where proteins are present inside individual cells. We are hopeful that HCPL can help scientists study how proteins work and develop new treatments for diseases."

Spatial proteomics is a research area that studies the distribution of proteins in cells or tissues using a combination of experimental techniques and computational approaches. Fluorescence microscopy is a common method in this field where proteins are physically tagged with fluorescent markers. AI maps the proteins onto individual cell compartments (subcellular structures or organelles). This helps scientists to understand the roles and functions of proteins and possibly reveal the complex inner workings of cells.

HCPL was developed in partnership with ForecomAI, a research and development company with world-class expertise in machine and deep learning providing solutions in health care and biosciences.

Dr. Amaia Irizar, director of ForecomAI said, "Proteins play a key role in most cellular processes crucial to our survival. Unraveling [protein](#) distributions and interactions within [individual cells](#) is vital to understanding their functions and indispensable to developing new treatments.

"Our work with the University of Surrey enables scaling up of this process and opens new frontiers. The partnership between Surrey and ForecomAI has been a successful interdisciplinary collaboration in scientific research, paving the way for further initiatives."

More information: Syed Sameed Husain et al, Single-cell subcellular protein localisation using novel ensembles of diverse deep architectures,

Communications Biology (2023). [DOI: 10.1038/s42003-023-04840-z](https://doi.org/10.1038/s42003-023-04840-z)

Provided by University of Surrey

Citation: Newly developed AI shows speed and accuracy in identifying the location and expression of proteins (2023, May 10) retrieved 25 April 2024 from <https://phys.org/news/2023-05-newly-ai-accuracy-proteins.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.