Unravelling protein distributions within individual cells is vital to understanding their function and state and indispensable to developing new treatments. Here we present the Hybrid subCellular Protein Localiser (HCPL), which learns from weakly labelled data to robustly localise single-cell subcellular protein patterns. It comprises innovative DNN architectures exploiting wavelet filters and learnt parametric activations that successfully tackle drastic cell variability. HCPL features correlation-based ensembling of novel architectures that boosts performance and aids generalisation. Large-scale data annotation is made feasible by our AI-trains-AI approach, which determines the visual integrity of cells and emphasises reliable labels for efficient training. In the Human Protein Atlas context, we demonstrate that HCPL is best performing in the single-cell classification of protein localisation patterns. To better understand the inner workings of HCPL and assess its biological relevance, we analyse the contributions of each system component and dissect the emergent features from which the localisation predictions are derived.
Proteins play a vital role in most cellular processes—crucial to our survival. Their intracellular locations provide important insights about cell functions and state. The specific biological functions that proteins perform are closely tied to the subcellular compartments in which they are expressed. Therefore, the subcellular resolution is critical in determining functional information about proteins and understanding the regulation of individual cells. A highly promising direction in this field is automated analysis of immunofluorescence microscopy images to enable large-scale impactful discoveries. For example, image-based spatial analysis of proteomic cellular heterogeneity can uncover a valuable view of protein expression with subcellular resolution, aiding the identification of disease biomarkers and drug discovery.

Single-cell analysis is key to the detection of rare cells in heterogeneous populations, essential in the profiling of tumour biology and precision medicine. It constitutes a core strategy of the LifeTime Initiative, a large-scale, long-term initiative to implement cell-based interdisciplinary medicine in Europe. Although machine learning (ML) has been used to describe the location of human proteins in microscope images giving summary information on an entire population of cells, subcellular classification of proteins for individual cells is still an open research area with limited published work and limited publicly available high-quality data.

To address this gap, we have developed a deep-learning-based system, the Hybrid subCellular Protein Localiser (HCPL), for robust protein localisation with subcellular resolution. Our system uses an ensemble of diverse deep architectures to generate precise annotations of protein localisation patterns in individual cells. We develop and validate our approach using the Human Protein Atlas (HPA), which is the largest public dataset and forms an invaluable resource for studying cell biology (Methods). The HPA contains an extensive collection of four-channel images depicting specific protein localisations at a subcellular level, acquired using immunofluorescence staining followed by confocal microscopy imaging. This resource is vital for understanding human cells and the complex molecular mechanisms underpinning their functions, taking advantage of antibody-based multiplexed protein imaging methods.

Aimed at individual cell analysis, the HCPL system successfully addresses several major challenges that this task raises. Compared with methods that can only provide predictions for a collection of cells, the requirement to classify each cell individually is a far more difficult task. While single-cell localisation requires accurate predictions for each cell, on an image level it is sufficient to locate one relevant cell where the presence of a protein is most evident without classifying each of the remaining cells individually. From a ML perspective, this difficulty is further compounded by the frequent lack of accurate ground truth for training with typically only image-level labels available. For example, each HPA image comprises many cells jointly labelled with one set of labels, defined as the union of individual cell labels. Hence, the image-level labels are incorrect for some of the cells in an image. That is, in the same image of a genetically identical population, individual cells can exhibit different protein localisation patterns. This phenomenon is called weak labelling. Further difficulties arise from a dramatic variability in cells’ morphology, as well as from the use of different cell lines and inconsistent cell image quality caused by staining or segmentation failures. Finally, we must also contend with extremely imbalanced frequencies of the localisation classes, along with the multi-label setting where a single cell can take multiple labels. Details about HPA dataset class names and labels are presented in Supplementary Table 1.

Importantly, best image-level approaches were found unable to handle these challenges performing poorly on the task of cell-level protein localisation and only achieving circa 33.50% mAP. Recently presented unsupervised approach can learn latent space representations that loosely correspond to specific protein localisation patterns; however, this method cannot be used for the prediction of such patterns. Large-scale single-cell analysis and evaluation were made possible thanks to the efforts of the Human Protein Atlas team who ran an open challenge on this particular topic called Human Protein Atlas - Single Cell Classification. The results of this competition were analysed in ref. However, in contrast to ref. that reviews the best approaches in the competition, our work concentrates on the challenges that single-cell classification imposes and presents a systematic and in-depth solution. As a result, different parts of our system can be employed as stand-alone applications in the broader context (increasing the accuracy of existing labels and reducing the labour required for manual labelling tasks). In addition, the HCPL system produces superior results to those of and achieves the classification performance of 57.19% mAP in single-cell analysis. To better understand our system’s operation and the benefits brought by its innovative components, we perform extensive testing, including a series of ablation studies. Furthermore, we benchmark our system against the leading solutions developed during the recent Kaggle competition (Human Protein Atlas - Single Cell Classification). Finally, the work is concluded with an analysis that verifies the biological correctness and meaningfulness of the systems’ predictions.

We believe the HCPL system fills an important gap and is well-placed to contribute to our knowledge of spatial biology in health and disease, and its application to the development of therapeutics.

Results

**HCPL - Hybrid subcellular protein localiser.** Figure 1 presents an overview of the HPA dataset, the HPA challenge, and our HCPL solution. The HCPL system receives multi-channel images, segments individual cells using the HPA Cell Segmentator (Methods), and analyses each cell in turn to estimate its visual integrity and the probabilities of proteins being present in specific subcellular compartments. HCPL combines several DNNs (Deep Neural Networks) to capture the biological variability and richness of patterns present in the HPA data, specifically:

- A multi-task Dual-stream Actnet (DSA) DNN, which learns to extract and fuse information from both images and individual cells. The DSA mitigates the vast amounts of false-positive cell predictions caused by the weak cell labelling used for training.
- A robust Cell-level Hybrid model (CLH), which combines learnt deep features with handcrafted features from a set of well-defined filters.
- A Cell-level Actnet (CLA), which employs learnable parametric activations to robustly aggregate discriminatory image features.

For system training, we propose an AI-trains-AI framework to improve the quality of the weak labels. It employs two techniques: (1) DNN-powered re-labelling, where labels are iteratively improved during the training, and (2) automatic adjustment of localisation confidence factors based on a cell’s estimated visual integrity, limiting the impact of cells with segmentation errors.

We develop an algorithm to select and optimally ensemble multiple classifiers to fully benefit from the diversity in expert opinions provided by different DNNs. Our HCPL system
ensambles nine diverse architectures and achieves a protein localisation accuracy of 57.19% mAP, the highest among those evaluated. Execution time is below 1 second per cell image on a single thread CPU and NVIDIA-P100 GPU. It is essential to note that the test data is not available to anyone except competition organisers. Our models are trained using only the training dataset without visibility of the test set and without external datasets. All cellular images presented in the manuscript have been obtained from the publicly available HPA dataset.

**Fig. 1 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 labelling, 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.

**DSA model for efficient cell-level and image-level information fusion.** To maximise classification performance and increase robustness, our DSA model (Fig. 2a) jointly exploits local (cell-level) and global (image-level) visual cues. The DSA architecture comprises a baseline DNN (e.g., EfficientNet) followed by the dual-stream network. From each image, cells are extracted and forwarded to the DNN component, producing deep convolutional features. The features are passed to a learnable Weibull activation pooling to improve the discriminative power of the feature map. Specifically, the weak uninformative features are dampened, and a learnable proportion of strong informative features are equalised. The aggregated dense features are classified in two parallel streams—the image stream outputs predictions for the entire population of cells in an image, whereas the cell stream generates a set of predictions for each cell.

The DSA network is trained end-to-end using a weighted sum of Binary Cross-Entropy (Methods) losses from cell and image streams. At inference, for each cell, the probability of each class is computed as a product of relevant image and cell stream probabilities. For selected classes where image-level predictions are less reliable, the final probabilities given for such classes are the cell stream probabilities.

In Fig. 3a, experimental result show that the image stream individually achieves 42.11% mAP. This baseline performance comes from mapping image-level labels to all cells in that image, resulting in vast over labelling. The cell stream achieves a better 51.12% mAP, still relatively low because of the weak labels used in training. However, our dual-stream DSA architecture achieves 55.14%, a gain of (+4.02%) stemming from the intelligent fusion of both streams.

**Cell-level hybrid and actnet models for capturing cell variability.** We develop two innovative cell-level architectures to comprehend the richness of patterns in the HPA. The first network, CLH, uses an inductive hybrid system (Fig. 2b) to fuse hand-crafted features extracted by a scattering transform with
deep convolutional features. The scattering transform, formed by cascading wavelets, guarantees translation invariance and a linear response to deformations. The complementary nature of mathematically well-defined and data-adaptable filters yields a robust model. To ensure these properties are propagated into the DNN flow, two Hybrid Fusion blocks are inductively integrated into an EfficientNet architecture.

The Hybrid fusion blocks have been designed to incorporate scattering features into the training process to enhance the data flow with representative pre-defined features. As a result, the CLH benefits from well-defined and informative hand-crafted features and an adaptable and flexible deep learning-based approach. Experimental results show that the Hybrid fusion blocks improve the performance by 1.58% mAP compared with the same DNN without fusion. The Hybrid-DNN module works on cell-level images and outputs convolutional features, which are aggregated and passed to the classification module to generate predictions.

The second network, CLA, uses parametric activations for adaptive and robust aggregation (Fig. 2c). At its core, a base DNN extracts features which are fed to a learnable activation layer comprising the Weibull function. The Weibull function amplifies responses corresponding to distinctive features of cells that are important for the classification tasks relative to the background.

Fig. 2 Key components of the HCPL system. a The DSA comprises an image stream that models the entire population of cells and a cell stream that classifies patterns in each cell in the image. b CLH is a single-stream inductive hybrid architecture that uses a scattering transform as a complimentary source of information about each cell. c CLA is a single-stream architecture that employs Weibull activation pooling to aggregate deep features before classification.
The transformed feature vectors are forwarded to a global average pooling layer and power-normalisation layer to generate a global vector. Experiments showed that Weibull pooling improves the mAP score by 2.82%, 2.53%, 2.11%, and 1.95% compared with the Global max pooling, Global average pooling, Region of interest pooling, and Attention-based pooling, respectively. The global vector is passed to the classification module to compute predictions.

CLH and CLA are trained using a weighted Binary Cross-Entropy (BCE) loss. The BCE weights for each class are calculated as the inverse of the class frequency in the training dataset. At inference, the class probabilities for each cell are computed similarly to the DSA, that is as a product of relevant DSA image stream and cell-level stream probabilities (CLA or CLH outputs), conditioned on the reliability of the image stream.

In Fig. 3b, we compute the class-wise Average Precision (AP) difference between CLH and CLA. It can be observed that the CLH better represents rare classes (0.22% and 0.11% AP gain over CLA on Aggresome and Mitotic). The second module consists of a base EfficientNet-B2 with a fully connected layer to output predictions for four classes. The data to train the DNN is generated by randomly cropping out some area on the cell’s border. The cells are classified into four classes based on the proportion of their cropped area to the original cell: less than 30% belongs to class 1, 30%–50% belongs to class 2, 50%–80% belongs to class 3, and 80%–100% belongs to class 4. The network is trained using cross-entropy loss. During inference, the network is applied to each cell, and the weight \( W_2 \) is assigned a value based on the output class, with \( W_2 = 0.1 \) for class 1, \( W_2 = 0.5 \) for class 2, and \( W_2 = 1 \) for class 3 or 4.

**Estimation of cell visual integrity with VID.** Poor imaging or segmentation failures lead to errors in classification. Hence, an important feature of our system is a Visual Integrity Detector (VID) that is trained to detect cell capture errors and to adjust classification confidence accordingly. It is generally agreed that at least 50% of a cell has to be captured in an image for a meaningful prediction. As presented in Fig. 4a, VID consists of two modules: (i) the first extracts hand-crafted geometric features from each cell and uses an XGBoost classifier to evaluate its structural integrity. (ii) An EfficientNet-B2 network, which predicts the ratio of the total cell body being captured. For training of the first VID module, we first compute the eight most representative features from each cell in the training dataset: bounding box height, width, aspect ratio, area, mask area, mask perimeter, the value of the largest dimension and a binary feature that is based on the pixel intensity and the ratio of blue and green to the total number of pixels. The VID dataset contains 10K cells hand-labelled as either ’good’, i.e., most of a cell is visible or ’bad’, i.e., a cell is severely occluded or a substantial part not visible. Note, this dataset with an extended set of properties and reference segmented cells is made publicly available.

We then train the XGBoost classifier on cell features using a fivefold cross-validation strategy. At inference time, each cell is passed to the XGBoost classifier to output the probability of the cell being bad \( P_b \). The VI weight \( W_1 \) is computed as \( 1 - P_b \).

Fig. 3 Performance details of selected elements of our system. a The classification performance of different modules of HCPL. The DSA*, CLH*, and CLA* modules are trained using weak image-level labels, whereas the DSA, CLH, and CLA modules are trained using improved labels obtained from the Cells Re-labelling Algorithm (CRA). The Ensemble VID and Ensemble systems represent the ensembling of DSA, CLA, and CLH models with and without the VID module, respectively. b Class-wise performance difference between cell-level Hybrid and cell-level Actnet. Positive values indicate better performance for CLH.
The total weight $W_c$ of VID is computed as the product of $W_1$ and $W_2$. Finally, the class probability of each cell is multiplied by $W_c$. The accuracy of the VID is 92.05% on 2K cell images of the VID dataset. In Fig. 3a, we can observe that the inclusion of the VID module mitigates false positive cell predictions and improves the system performance by 0.85% mAP.

Results in Fig. 3a show that the DSA retrained on improved labels achieved performance of 55.14% mAP compared with DSA trained on original weak labels (52.52% mAP). The above process is employed twice to refine label quality. Finally, the cell labels obtained after round two are used to train the CLA and CLH models, leading to improvements exceeding +4% mAP over the models trained on the original labels.

**Improving label quality using an AI-trains-AI approach.** We leverage the generalisation ability of our DSA model to learn from noisy data and assign confidence scores to ground-truth labels (image-level labels naively mapped to cells). A confidence score represents the probability that a ground-truth label is correct; hence it takes values in [0,1]. We use confidence scores to perform iterative training where subsequent models focus less on low-confidence cell labels while emphasising those with high confidence. This is achieved by our Cells Re-labelling Algorithm (CRA). It first computes the per-cell component probabilities using three DSAs trained on weak labels (bases EfficientNet-B4, NFNet-ECA, and ResNet50d). These three component probabilities are averaged across each label for each cell to obtain a new set of combined confidence factors.

In the second step, we re-evaluate the true-positive labels for each cell based on these combined confidence factors (Fig. 4b, left). The CRA transforms the factors using a power-normalisation operation ($c \Rightarrow c^\beta$, $c$ is a probability) to generate a new set of probabilities $K$. The probabilities $K$ then replace their associated labels, producing a new set of improved continuous valued cell labels. The histogram of improved cell labels is presented on the right.

**Multi-stream information fusion.** We exploit the fusion of information extracted by selected classifiers working on image and cell levels to maximise the system’s performance. Figure 5a shows 2D histograms of image-level and cell-level predictions for all nineteen classes. Each histogram is computed using images with their corresponding label.

We note that certain classes (Mitotic spindle, Centrosome, Aggresome, and Vesicles) show disagreement in predictions, where a bright spot is present at the top-left corner. This shows cell-level labelling rejecting cells classified as positive by the image-level labelling. This phenomenon can be summarised by computing the correlation coefficient $r$ between the image- and cell-level predictions. The aforementioned classes have very low correlation coefficient values, reflecting that these rare classes are prone to over-labelling by image classifiers.

Image and cell level fusion is performed as follows. For classes where the correlation is greater than a certain threshold $\rho_{th} = 0.32$, the localisation probabilities for a cell are obtained as per class products of image-level and cell-level predictions. For classes with low correlation, we assign the cell-level predictions to the final class probabilities. Please note that all cell-level networks are trained using improved cell-level labels.
Robust protein localisation using diversity-based ensembling.

Experimental results of our models show that the maximum performance that a single model can achieve is 55.84% mAP (Fig. 6a). Classification accuracy can be substantially improved by optimal ensembling of DNNs (classifiers) outputs. However, meticulous selection of diverse classifiers is required to achieve this improvement. A set of classifiers are considered diverse if they perform well on different examples or classes.

The diversity inherent in the DNNs can be visualised with a correlation matrix of probabilities produced by each network (Fig. 5b). Here, we observe a block diagonal structure in the correlation matrix (green diagonal blocks), demonstrating that DNNs of similar architectures (EfficientNets, NFNets, ResNets, DenseNets and Transformers) have strong correlations. The selection of the final network set (Fig. 5c) is accomplished by selecting the best-performing DNN, on the Kaggle public leaderboard (Methods), from each class of architectures.

The HCPL system, utilising nine diverse DNNs, is illustrated in Fig. 5c. The first phase is to extract individual cells from each image using HPA Cell Segmentator. Next, the cells are forwarded to the individual DSA and Cell-level networks to compute the predictions. The outputs from these diverse networks are hierarchically aggregated to compute the probabilities of all classes. Simultaneously, the cells are passed to the VID to compute the cell visual integrity weighting. Finally, the
probabilities are multiplied by the visual integrity weights to generate the final vector of class probabilities.

Quantifying HCPL performance and identifying its essential components. We conduct ablation studies to evaluate strengths of different models and the improvements brought by diversity-driven multi-DNN ensembling. According to Fig. 6a, the individual DNNs are not able to improve beyond 55.84% mAP. The mAP is improved to 56.64% by ensembling four Dual-stream architectures with bases EfficientNet-B4 (54.72%), NFNet-ECA (54.84%), ResNet50d (55.14%), and DenseNet161 (54.67%). We can achieve a mAP of 56.75% by ensembling the predictions from CLAs with bases EfficientNet-B4 (55.61%), SwinT (54.95%), ResNet200d (55.84%), ResNet50d (55.22%) and CLH (55.23%). Importantly, the ensembling of all DSA and cell-level networks obtain a classification score of 57.19%.

Benchmarking the HCPL system. We start our analysis with the evaluation of existing image-level systems. For this, we take the best-performing image-level classification DNN (best fitting, DenseNet-based) as a starting reference point. This model is fine-tuned on the HPA single-cell classification dataset using the original training configuration. The resulting performance is 33.50% mAP on single-cell classification, indicating that image-level classifiers are not directly applicable to the single-cell classification task. A more meaningful comparison can be obtained by evaluating the performance of HCPL against the methods employed in the Kaggle (HPA Single Cell Classification) competition. The details about the different architectures are presented in Supplementary Note 1.

A limitation of this comparison is that existing solutions were designed during the four-month Kaggle competition, with feedback based on public leaderboard scores, while we developed HCPL after the competition. While we had no access to test data, submitting code after the competition showed scores on public and private leaderboards. For a fair comparison, we only used the public leaderboard to develop our algorithm and avoided any optimisation based on the private leaderboard. Once we finalised the architecture based on the public leaderboard, we reported the performance on the private leaderboard.

We first compare the performances of single models without ensembling. From Fig. 6b, we observe that the proposed DSA performs better than Kaggle multi-head models, including Fair Cell Activation Network (FCAN), Duo-branch, modified Puzzle-CAM, and PfNet. This DSA improvement can be attributed to the use of Weibull activation pooling to aggregate deep features, correlation-dependent thresholding for the fusion of the two streams, along with training on improved labels from CRA.

AmonThang cell-level architectures, Fig. 6c demonstrates that our CLA achieves better classification performance (55.84%) than Cell model InceptionV3, Cell model SwinT, Data-centric CNN, Gridify and Gapmask, and Bag-of-Cells. The advantage of CLA is associated with the use of learnable and parametric activation pooling to aggregate discriminatory features.
When ensemble-based architectures are considered (Fig. 6d), our HCPL system attains the highest classification accuracy of 57.19% mAP. The best currently published result (56.67%, bestfitting) was trained with antibody-id information which was scraped from the Human Protein Atlas website and re-linked with the competition dataset.

The antibody-ids were excluded from the competition datasets by the organisers because their use in training presents three issues: (1) the model may use irrelevant hidden variables (for example, background noise) or focus on biologically meaningless features to improve the model performance; (2) the gains achieved might be due to the exploitation of batch effects and statistical correlations between the training and test sets; and (3) the model trained using antibody-id information may generalise poorly to unseen cell lines, new antibody-ids, and unknown proteins.

In fact, competition organisers suggested limiting the set of information allowed for training to protein location labels only (as supplied with the datasets). This prevents models from exploiting spurious correlations between the predicted protein localisation labels and additional information such as antibody-ids, tissue-ids and cell lines, helping to ensure that the cell localisation predictions are unbiased and generalise well.

The evaluation results show that without using the antibody-id information, the performance of bestfitting is 55.54%. To achieve this outcome, we followed the code and algorithm description of the bestfitting solution, and retrained the models without incorporating the antibody-id information (Methods). Importantly, no other team, including our HCPL, used antibody-id information during model training. When a fair comparison is performed, i.e., antibody-id data is not employed, HCPL (57.19%) demonstrates a gain of +1.65% and +1.86% over the first (55.54%) and second (55.33%) Kaggle teams. The HCPL’s better performance is predominantly due to the combination of its component parts, which includes utilising DNN architectures with wavelet filters and Weibull activation pooling to effectively handle large cell variations, employing the CRA algorithm to produce enhanced cell labels, using the VID system to evaluate the visual quality of cells, and combining diverse deep architectures through correlation-based ensembling.

In Fig. 7a, we show the performance of our system in each of the 19 classes. The performance depends on several variables, such as the difficulty of recognising different localisation patterns, the number of training samples, and extreme visual variations. Classes with distinct visual patterns have high AP: Nucleoplasm 76.55%, Nuclear membrane 76.25%, and microtubules 71.10%. Despite a lower number of training samples, the AP is higher for Aggresome (64.75%) than Plasma membrane (49.02%) because Aggresome has visually distinguishable features, whereas Plasma membrane is often confused with Cytosol. The Endoplasmic reticulum has the lowest AP of 36.30%; it is also confused with the Mitotic spindle and Aggresome (64.75%) than Plasma membrane (49.02%) because these features of HCPL combined with VID help achieve a considerable advantage in challenging classes over the red.ai system.

The ability of HCPL to interpret biological information. We use the Grad-CAM method to identify the parts of an input image that impact the classification score. This highlights regions contributing features that support predictions of the target label, helping us understand whether localisation predictions are biologically meaningful (by comparing to the ground truth staining patterns). These attention regions will vary for each image and location label.

Deep features visualisation using UMAP. To investigate the ability of a DNN to distinguish subcellular structures, we visualise the high-dimensional feature vector extracted from the penultimate layer using uniform manifold approximation and projection for dimension reduction (UMAP). The visualisation results for two different DNNs (DSA and CLH) are presented in Fig. 9. Here, each point on the plot represents a single cell. Only cells with single labels are chosen, coloured by their respective label.

Firstly, we observe that the DNNs cluster the majority of cells from the same class together, which shows their underlying discriminative power. Secondly, we note the presence of two larger clusters at a global level: one for nuclear sub-compartment (Nucleoplasm, Nuclear membrane, Nucleoli) and another for locations outside the nucleus (e.g., Cytosol, Mitochondria). Thirdly, instances of some classes are located within clusters belonging to other classes, which links to the relative decrease in performance (Fig. 7a), for example, Cytosol partially overlaps with Endoplasmic reticulum and Plasma membrane, Nuclear bodies with Nucleoplasm and Vesicles with Golgi apparatus. Lastly, the different amount of overlap observed in the clusters generated by DSA and CLH confirms the diversity and complementarity in the deep features.

Discussion

Our core contributions include a multi-task DSA architecture for improved image and cell information fusion and two cell-level architectures (CLA & CLH) to capture the wide variability between cells. Further, we introduce an AI-trains-AI approach...
with DNN-powered re-labelling and automatic cell visual integrity weighting and propose an effective strategy to optimally ensemble diverse DNNs. The evaluation shows that HCPL has better single-cell classification performance than existing approaches and is expected to generalise well to unseen cell lines and proteins. Further improvements could be achieved by obtaining more examples of rare localisation classes, accurate cell labels, and feedback from human experts.

HCPL narrows the performance gap between AI methods and human experts and provides a toolbox of methods to tackle the challenges of single-cell protein localisation successfully. This will help accelerate the characterisation of unknown proteins and our understanding of cellular function and biology to advance our knowledge of disease-related phenotypes and drug discovery.

### Methods

#### HPA dataset

Our work uses the HPA dataset provided by the Human Protein Atlas – Single Cell Classification Kaggle challenge. This consists of images from the freely accessible Human Protein Atlas project. In particular, images from the subcellular section of the HPA were used. A total of 104307 images (internal and external dataset) were made available for training purposes while an additional 1776 unseen images were retained by the challenge organisers for testing purposes, further split into 559 images for the public leaderboard and 1217 images for the private leaderboard. Each of the provided images contains multiple cells and consists of four channels, displayed in red, green, blue, and yellow. The task considered in this paper is to localise the protein of interest (green channel) in 18 possible subcellular organelles in each cell in an image. An additional negative class is added for negative staining and unspecific patterns.

A cell can have multiple labels (specifically, we work with 19 localisation classes with up to 6 protein locations per cell). The 19 labels and their names are shown in Fig. 1a. Our DNNs are trained on approximately 1.2 Million cell images segmented from HPA images using the HPA Cell Segmentator.

#### Evaluation metric

To ensure a fair comparison of results, all experiments were evaluated by computing mean Average Precision (mAP). The mean value was calculated over the 19 segmentable classes ($C$) of the challenge with a mask-to-mask IoU > 0.6 as described below:

$$mAP = \frac{1}{C} \sum_{i} Prec_i$$

where $Prec_i$ is Precision for class $i$ which is calculated according to the two-stage method described in article.

All mAP scores are reported based on the Kaggle private leaderboard.

#### HPA cell segmentator (HCS)

The HCS (https://github.com/CellProfiling/HPA-Cell-Segmentation) segments input images into individual cell instances for multi-label classification. Since the procedure of the HCS software provided by organisers occupied 67% of the total permitted processing time (9 hours for the entire system’s inference), improving the efficiency of the segmentation algorithm is important. The algorithm consists of three main stages: (i) prediction of the nuclei; (ii) general cell prediction; and (iii) post-processing procedure. However, cell segmentation training data was not publicly available, and so gains could only be obtained by modifying the post-processing procedure rather than training a new model. In the segmentation architecture, the nuclei and cell maps are first obtained via the corresponding predictor U-Net. The next step is the post-processing of the outputs. To improve the efficiency of HCS, we introduce...
down-scaling and up-scaling blocks at the start and the end of the process respectively. Since post-processing is largely based on various morphological operations, its complexity is proportional to the product of the processed image dimensions. Therefore, reducing the spatial dimensions by 50% resulted in a speedup of a factor of at least 2. Another effect of the reduced dimensionality was a large reduction in high-frequency noise. This allowed simplification of the pipeline by removing extra processing in two fine-tuning blocks and in the Segmentation and gradual object removal block. These changes resulted in a 2x speed-up. However, such speed improvement resulted in a deterioration in system accuracy of around 0.2% mAP.

**Fig. 8 CAMs of the high-scoring CLA trained on CRA improved labels, and a low-scoring CLA trained on original weak labels.** (1) The CAMs for the Nucleoplasm staining demonstrate biologically meaningless attention for the low-scoring model as shown by red circles. (2) The Microtubule staining overlaps well with the CAM for the high-scoring model. In contrast, the low-scoring model uses visual features of the Mitotic spindle class to assign a high probability for the Mitotic spindle class, as indicated by red circles. (3) For the Mitotic spindle staining patterns, CAMs for the high-scoring CLA highlight relevant cellular regions indicated by green circles. (4) The red circle shows where the Aggresome staining pattern is not captured by the low-scoring DNN, whereas the green circle depicts the high-scoring model correctly localises it. Scale bars, 10 μm.

**DSA, CLH, and CLA training and inference configurations.** The DSA comprises a baseline DNN followed by the dual stream network. From each image, N cells are selected, resized and flattened as a batch (typically N = 20). Let \( X \in \mathbb{R}^{A \times B \times 4} \) denote a cell image of resolution \( A \times B \). Each cell \( X \) is processed by a base DNN, which embeds an input into the space of compact deep features. The output tensor of the final convolutional layer, denoted as \( R \in \mathbb{R}^{W \times H \times D} \), is forwarded to a learnable Weibull activation layer \( \hat{R} \), where \( W \) and \( H \) are the width and height of the feature map and \( D \) is the feature dimensionality. The output of the activation layer is forwarded to a Global Average-Pooling (GAP) layer and power-normalisation layer to generate global descriptors, which are then passed to the image stream and cell stream.

The image stream applies Global Max-Pooling to a bag of \( N \) cell descriptors originating from a single image to generate a unified image representation \( V \), which is then passed to a fully connected layer \( FC_1 \) and Softmax to generate an image-level prediction. The cell stream takes \( N \) cell descriptors as an input and outputs the predictions for each cell using a fully-connected layer \( FC_2 \) and Softmax. The predictions from the image stream are passed to classification loss layer. The loss layer computes the weighted Binary Cross-Entropy loss \( L_f \) between the image label and bag prediction. Similarly, the cell stream’s weighted Binary Cross-Entropy loss \( L_c \) is calculated between cell predictions and cell labels. The final loss \( L_f \) is the weighted sum of cell stream loss and image stream loss \( L_f = W_c \times L_c + W_i \times L_i \). In the first round of DSA training, the cell-level labels are weak, and we therefore intuitively assign a much lower weight to cell stream loss compared with image stream loss. The final loss is computed as \( L_f = 1 \times L_i + 0.2 \times L_c \). In the second round, the DSA is trained on improved cell labels obtained from our CRA algorithm. Since we now have more trust in the cell-level labels the final loss is computed as \( L_f = 1 \times L_i + 1 \times L_c \). The DSA is trained using an Adam optimiser and cosine annealing learning rate scheduler. The Cell-level Hybrid takes cell images as an input and outputs convolutional features denoted as \( R \in \mathbb{R}^{W \times H \times D} \). The features are aggregated using Global Average Pooling (GAP) layer and forwarded to classification module (fully-connected layer and Softmax).

In Cell-level ActNet, the convolutions features extracted from cell images are passed to learnable Weibull activation pooling. The transformed features are aggregated using GAP and power-normalisation layers and forwarded to the classification module.

The training of CLH and CLA is performed using weighted Binary Cross-Entropy loss, Focal loss, Adam optimiser, and a cosine annealing scheduler with initial learning rate \( 2e^{-4} \). We applied data augmentation in the form of random cropping, flipping, shifting, rotation, scaling, and cutout to train all models.
Each element of the tensor Weibull function is transformed by the Weibull function resulting in the output tensor $R_{fi}$, the output tensor of the final convolutional layer. Each element of the tensor $R$ is transformed by the Weibull function resulting in the output tensor $T \in \mathbb{R}^{W \times H \times D}$ (where $0 \leq i < W \times H \times D$):

$$T_i = \left( \frac{R_i}{\lambda} \right)^{\gamma} \exp\left(-\left(\frac{R_i}{\gamma}\right)^\eta\right),$$  \hspace{1cm} (2)$$

The learnable parameters of the activation layer are $\lambda, \gamma$, and $\eta$. The output of the activation layer is fed to the Global Average Pooling (GAP) layer, denoted as $P(T)$, to compute the global vector $S$:

$$S = \left( \frac{1}{WH} \sum_{j=1}^{W} \sum_{i=1}^{H} T_{kj} \right)_{k=1}^{D}$$  \hspace{1cm} (3)$$

Each element ($s \in S$) of the tensor $S$ is power normalised to balance the non-linear scaling of the Weibull function. The power-normalisation function is represented as $\delta: \mathbb{R}^D \rightarrow \mathbb{R}^I$, with the rule:

$$\delta(s) = a_1 s_1^{\beta_1}, ..., a_D s_D^{\beta_D}$$  \hspace{1cm} (4)$$

where $a, \beta$ are learnable scaling parameters.

**Experimental evaluation.** We followed the bestfitting solution code (https://github.com/topics/hpa-challenge-2021) and the algorithm description\(^1\)\(^1\)\(^3\) to retrain the models without using the antibody-id information. The base model for the bestfitting solution, Fair Cell Activation Network (FCAN), was trained to jointly minimise three loss functions: reconstructing regularisation loss between the image and cells Class Activation Maps (CAMs), classification losses supervised by nineteen class labels and metric learning losses supervised by antibodies. We removed the metric learning loss layer and retrained the FCAN using the original hyperparameters. Following this, we used the original post-processing algorithm and an ensemble of FCANs and cell-level models to obtain the final solution.

The class-wise performances of Kaggle teams were obtained from Supplementary Tables 1 to 11 of ref.\(^1\)\(^1\)\(^3\).

**Statistics and reproducibility.** The study did not employ statistical analysis when evaluating the HCPL performance on Kaggle platform. As this is a Kaggle code competition, all code submission for inference were collected and graded automatically, which allows for reproduction of the scores. All training data and HPAv20 dataset are publicly available for model training and testing dataset is available on Kaggle platform. Results reported for the HCPL system can be reproduced using the standard submission process for the Human Protein Atlas - Single Cell Classification Kaggle competition\(^1\)\(^2\) and by utilising the source code and models provided in the Code Availability section. All the data required to reproduce the figures are in Figshare https://doi.org/10.6084/m9.figshare.2251208.

**Data availability**

The HPA training dataset analysed during the current study are available from Kaggle, https://www.kaggle.com/competitions/hsa-single-cell-image-classification/data. The Bad Cell dataset generated during the current study are available at https://www.kaggle.com/datasets/anokas/hpabadcellxgboost. The HPA test set is not openly available and was not accessible to the authors; code submissions can be made for scoring through the Kaggle page.

**Code availability**

The Python source code for HCPL and model weights suitable for inference are available online via the project’s web page at https://bit.ly/hpa-hcpl.

Received: 21 October 2022; Accepted: 12 April 2023; Published online: 05 May 2023

**References**

1. Paul, I., White, C., Turcinovic, I. & Emili, A. Imaging the future: the emerging era of single-cell spatial proteomics. FEBS J. 288, 6990–7001 (2021).
2. Gnann, C., Cesnik, A. J. & Lundberg, E. Illuminating non-genetic cellular heterogeneity with imaging-based spatial proteomics. Trends Cancer 7, 278–282 (2021).
3. Mahdessian, D. et al. Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. Nature 590, 649–654 (2021).
4. Chandrasekaran, S. N., Ceulemans, H., Boyd, J. & Carpenter, A. E. Image-based profiling for drug discovery: due for a machine-learning upgrade? Nature Rev. Drug Discov. 20, 145–159 (2020).
5. Tellez-Gabriel, M., Ory, B., Lamoureux, F., Heymann, M.-F. & Heymann, D. Tumour heterogeneity: the key advantages of single-cell analysis. Int. J. Mol. Sci. 17, 1661–16596 (2016).
6. Bodenmiller, B. Multiplexed epitope-based tissue imaging for discovery and healthcare applications. Cell Syst. 2, 225–238 (2016).

7. Rajewsky, N. et al. Lifetime and improving european healthcare through cell-based interdictive medicine. Nature 587, 377–386 (2020).

8. Sullivan, D. et al. Deep learning is combined with massive-scale citizen science to improve large-scale image classification. Nat. Biotechnol. 36, 820–828 (2018).

9. Ouayang, W. et al. Analysis of the human protein atlas image classification competition. Nat. Methods 16, 1254–1261 (2019).

10. Kobayashi, H., Cheveralls, K. C., Leonetti, M. D. & Royer, L. A. Self-supervised deep learning encodes high-resolution features of protein subcellular localization. Nat. Methods 19, 995–1003 (2022).

11. Le, T. et al. Analysis of the human protein atlas weakly supervised single-cell localization. Nat. Methods 19, 1221–1229 (2022).

12. Winsnes, C., Lundberg, E., Phil, C.M., Le, T. & Wei, O.U. Human Protein Atlas - Single Cell Classification. https://www.kaggle.com/competitions/hpa-single-cell-image-classification (2021).

13. Uhlen, M. et al. Tissue-based map of the human proteome. Science 347, 1266419 (2015).

14. Thal, P. J. et al. A subcellular map of the human proteome. Science 356, eaal3321 (2017).

15. Lundberg, E. & Borner, G. Spatial proteomics: a powerful discovery tool for cell biology. Nat. Rev. Mol. Cell Biol. 20, 283–302 (2019).

16. Christopher, J. et al. Subcellular proteomics. Nat. Rev. Methods Primers 1, 32 (2021).

17. Hickey, J. et al. Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging. Nat. Methods 19, 284–295 (2022).

18. Tan, M. & Le, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proc. International Conference on Machine Learning. PMLR 119, 636–647 (2022).

19. Husain, S. S., Ong, E. & Bober, M. ACTNET: end-to-end learning of feature activations and multi-stream aggregation for effective instance image retrieval. Int. J. Comput. Vis. 129, 1432–1450 (2021).

20. Minský, D. & Bober, M. Efficient hybrid network: Inducting scattering features. In Proc. International Conference on Pattern Recognition, 2300–2306 (2022).

21. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proc. International Conference on Learning Representations. 730–734 (2015).

22. Babenko, A. & Lemhtyski, V. Aggregating local deep features for image retrieval. In Proc. IEEE International Conference on Computer Vision. 1269–1277 (2015).

23. Gordo, A., Almazan, J., Revaud, I. & Larlus, D. End-to-end learning of deep visual representations for image retrieval. Int. J. Comput. Vision 124, 237–254 (2017).

24. Makarov, I. Second Place Solution for Human Protein Atlas - Single Cell Classification. https://www.kaggle.com/competitions/hpa-single-cell-image-classification/discussion/238645 (2021).

25. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In Proc. International Conference on Machine Learning. 685–693 (2016).

26. Ong, E.-J., Minský, D., Bober-Irizar, M., Husain, S. S. & Bober, M. HPA Bad Cell Detection. https://www.kaggle.com/datasets/anokas/hpabaddetection (2022).

27. Brock, A., De, S., Smith, S. L. & Simonyan, K. High-performance large-scale image recognition without normalization. In Proc. International Conference on Machine Learning. 1059–1071 (2021).

28. Zhang, H. et al. ResNeSt: Split-attention Networks. In Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2735–2745 (2022).

29. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In Proc. IEEE Conference on Computer Vision and Pattern Recognition. 2261–2269 (2017).

30. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In Proc. International Conference on Computer Vision. 10012–10022 (2021).

31. He, T. et al. Bag of tricks for image classification with convolutional neural networks. In Proc. IEEE Conference on Computer Vision and Pattern Recognition. 558–567 (2019).

32. Winsnes, C., Lundberg, E., Phil Culliton, M., Le, T. & Wei Ouayang, U. Human Protein Atlas - Single Cell Classification Private Leaderboard. https://www.kaggle.com/competitions/hpa-single-cell-image-classification/leaderboard (2021).

33. Dai, S. Fair Cell Activation Network and Swim Transformer, the first place solution for Human Protein Atlas - Single Cell Classification. https://www.kaggle.com/competitions/hpa-single-cell-image-classification (2021).

34. Jo, S. & Yu, I.-J. Puzzle-cam: Improved localization via matching partial and full features. In Proc. IEEE International Conference on Image Processing. 639–642 (2021).

35. Benhamou, E. Third place solution MPWARE part for Human Protein Atlas - Single Cell Classification. https://www.kaggle.com/competitions/hpa-single-cell-image-classification/discussion/238862 (2021).

36. Chubachi, K. Seventh place solution for Human Protein Atlas - Single Cell Classification. https://www.kaggle.com/competitions/hpa-single-cell-image-classification/discussion/238507 (2021).

37. Baric, N. Fourth Place Solution for Human Protein Atlas - Single Cell Classification. https://www.kaggle.com/competitions/hpa-single-cell-image-classification/discussion/239071 (2021).

38. Yang, S. Sixth Place Solution for Human Protein Atlas - Single Cell Classification. https://www.kaggle.com/competitions/hpa-single-cell-image-classification/discussion/239166 (2021).

39. Henkel, C. Third place Dieter part for Human Protein Atlas - Single Cell Classification. https://www.kaggle.com/competitions/hpa-single-cell-image-classification/discussion/238989 (2021).

40. HPA: The Human Protein Atlas. https://www.proteinatlas.org/.

41. Dai, S. Where to get the antibody information (Kaggle Forums). https://www.kaggle.com/competitions/hpa-single-cell-image-classification/discussion/239001#1308869 (2021).

42. Lundberg, E. Human Protein Atlas - Single Cell Classification Antibody-id Discussion. https://www.kaggle.com/competitions/hpa-single-cell-image-classification/discussion/215675/1218794 (2021).

43. Winsnes, C. Human Protein Atlas - Single Cell Classification Cell lines Discussion. https://www.kaggle.com/competitions/hpa-single-cell-image-classification/discussion/227862 (2021).

44. Selvaraju, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proc. IEEE International Conference on Computer Vision. 610–614 (2019).

45. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction (2020). https://arxiv.org/abs/1802.03426.

46. Su, W., Yuan, Y. & Zhu, M. A relationship between the average precision and the area under the roc curve. In Proc. International Conference on The Theory of Information Retrieval. 349–352 (2015).

47. Everingham, M., Gool, L. V., Williams, C. K. I., Winn, J. & Zisserman, A. The pascal visual object classes (voc) challenge. Int. J. Comput. Vis. 88, 303–338 (2009).

48. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F. (eds.), Medical Image Computing and Computer-Assisted Intervention, 234–241 (2015).

Acknowledgements

The authors thank the organisers and all participants of the Kaggle competition. Funding support was provided by the CVSSP Research Group at the University of Surrey and ForecomAI. We thank the Human Protein Atlas for their continued work on open proteomic data. Examples from the HPA (proteinatlas.org) dataset within Figs. 1, 2, 4, and 8 are available under a CC BY-SA 3.0 license (https://creativecommons.org/licenses/by-sa/3.0/) license.

Author contributions

S.H., E.J.O., D.M., M.B.-L., and M.B.: methodology, software, validation analysis, data curation, and visualisation. A.I.: conceptualisation, visualisation, and funding acquisition. M.B.: conceptualisation, supervision, funding acquisition. All authors contributed equally to writing, review, and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-023-04840-z.

Correspondence and requests for materials should be addressed to Syed Sameed Husain.

Peer review information Communications Biology thanks Ying-Ying Xu and Hirofumi Kobayashi for their contribution to the peer review of this work. Primary Handling Editor: Manuel Breuer. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
