© 2024. Published by The Company of Biologists Ltd | Journal of Cell Science (2024) 137, jcs262095. doi:10.1242/jcs.262095

REVIEW

SPECIAL ISSUE
IMAGING CELL ARCHITECTURE AND DYNAMICS

Machine learning in microscopy – insights, opportunities and challenges

Inês Cunha[1], Emma Latron[1], Sebastian Bauer[1], Daniel Sage[2] and Juliette Griffié[1,*]

ABSTRACT

Machine learning (ML) is transforming the field of image processing and analysis, from automation of laborious tasks to open-ended exploration of visual patterns. This has striking implications for image-driven life science research, particularly microscopy. In this Review, we focus on the opportunities and challenges associated with applying ML-based pipelines for microscopy datasets from a user point of view. We investigate the significance of different data characteristics – quantity, transferability and content – and how this determines which ML model(s) to use, as well as their output(s). Within the context of cell biological questions and applications, we further discuss ML utility range, namely data curation, exploration, prediction and explanation, and what they entail and translate to in the context of microscopy. Finally, we explore the challenges, common artefacts and risks associated with ML in microscopy. Building on insights from other fields, we propose how these pitfalls might be mitigated for in microscopy.

KEY WORDS: Data, Analysis, Machine learning, Microscopy, Bioinformatics, Image analysis

Introduction

Over the past two decades, light microscopy has benefited from significant advancements in optical design and sample preparation. For example, the development of super-resolution microscopy (Betzig et al., 2006; Gustafsson, 2000; Klar and Hell, 1999; Rust et al., 2006) has enabled the visualisation of molecules with a spatial resolution ranging from a few nanometres to 200 nm. Meanwhile, lattice light-sheet microscopy (Chen et al., 2014) has allowed 3D live-cell-friendly imaging of sensitive samples with high spatial-temporal resolution, and expansion microscopy allows the physical expansion of samples to visualize details below 200 nm with conventional microscopes (Wassie et al., 2019). In parallel, improvements have also been made in data acquisition diversity and scalability, including microchips assays to assess the cytotoxicity of cells in microwells (Guldevall et al., 2016) and cell painting setups (Bray et al., 2016), to study the effect of a variety of perturbations (e.g. different drugs) in different cellular structures. In this Review, we investigate how image processing and analysis can use artificial intelligence (AI) to embrace (and benefit from) these extensive and complex datasets.

Machine learning (ML) techniques represent a dominant subset within AI and are increasingly used in life science research, in particular for bioimage analysis (Jan et al., 2024). In contrast to conventional analytical pipelines, ML involves two stages: training and inference. During training, the model iteratively learns complex patterns and intricate relationships in the dataset, by estimating mathematical parameters. The entire process relies on the choice of model architecture, capacity and regularisation techniques, which significantly impact on the performance of the model. To ensure an optimal training process, a set of hyperparameters must be chosen, such as the learning rate (i.e. how quickly the model updates its predictions) or the loss function (i.e. a measure of how well the predictions of the model match the actual data) (Yang et al., 2020). When the model has converged, that is trained, assessment metrics are used on a separate test dataset to evaluate the accuracy of the model. Therefore, the model can then be used for inference, during which it applies its acquired knowledge to make either predictions or decisions typically on new unseen data, or generate its own synthetic datasets. The accuracy of the model measures the reliability of the inferences. Conventionally, ML models can be categorised into supervised and unsupervised. Supervised models are trained on paired datasets, meaning an input with its corresponding output targets. For example, raw images paired with output segmentation labels (Fig. 1A). In contrast, unsupervised models are fully data-driven; they mathematically learn the mapping between given inputs and outputs without provided labels (Fig. 1B). The aim in this case is to uncover intrinsic properties or groupings within the input images themselves, without explicit guidance.

A common feature, independent of the ML method applied, is that data is core and central. It is not only analysed or processed through the pipeline, but prior to that, it is required for training. This is a fundamental difference compared with conventional non-ML strategies. Thus, understanding data from users (e.g. its type, quality, content and quantity) and its impact on the models, both in terms of possible applications and challenges, are paramount if ML is to become a widely used tool in microscopy. In this Review, we explore how the transition of microscopy image analysis to ML impacts both data acquisition and analysis. We further reflect on inevitable risks and challenges associated with using AI in microscopy research, based on experience from other fields that have embraced this transition earlier. We learn from the strategies used by these fields to mitigate specific challenges and comment on how these can be translated to microscopy.

Data acquisition – what matters in a dataset for ML models?

"Any model is only as good as the data it is trained on". In this section, we delve into data requirements for successful ML implementation. More specifically, we discuss how data quantity, transferability and content affect the ML models used, as well as the limitations and trade-offs associated with these features.

Quantity

We use data quantity here to refer to the volume of data available for training. In microscopy, it often translates to the number of images

[1]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Tomtebodavägen 23, 171 65 Solna, Sweden. [2]Biomedical Imaging Group and EPFL Center for Imaging, École Polytechnique, Rte Cantonale, 1015 Lausanne, Switzerland.

*Author for correspondence (juliette.griffie@dbb.su.se)

I.C., 0000-0002-1327-9018; D.S., 0000-0002-1150-1623; J.G., 0000-0001-6438-0119

Journal of Cell Science (2024) 137, jcs262095. doi:10.1242/jcs.262095

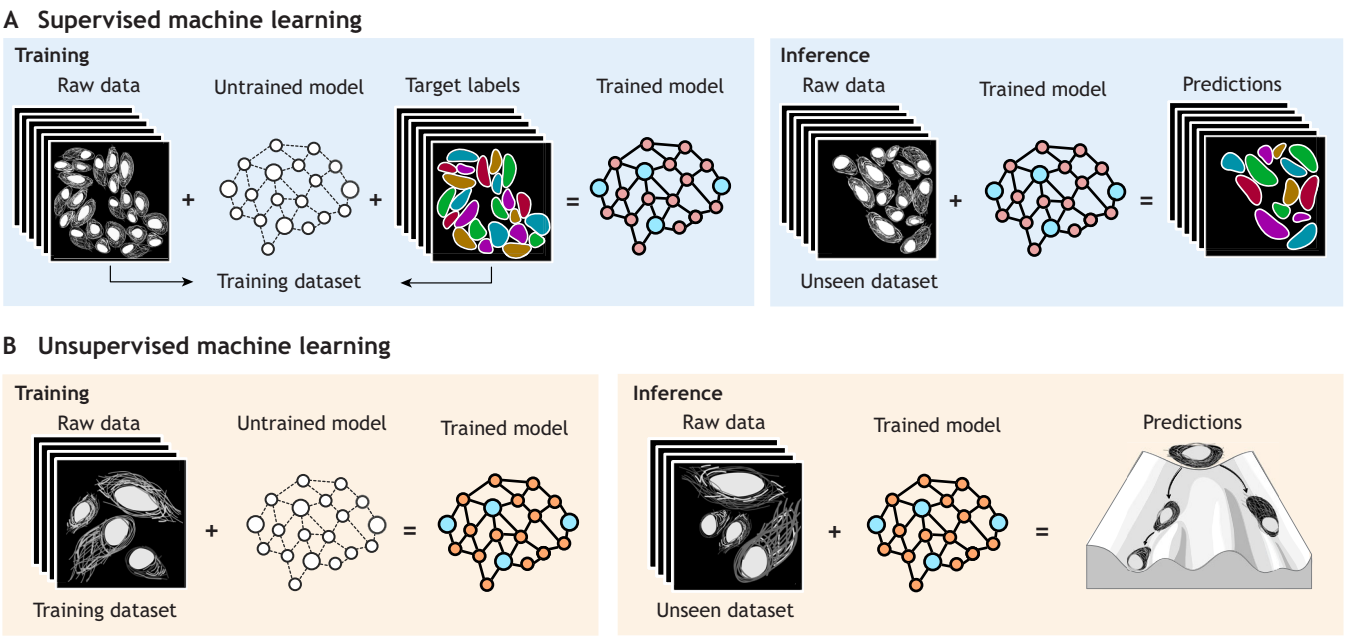

Fig. 1. ML principles. ML can be divided into two main subdivisions: supervised and unsupervised. (A) In supervised ML, during training (left), an untrained model takes raw training data and target labels as its inputs. The model learns a mathematical mapping from raw data to target labels. The trained model can then infer (right) predictions on unseen raw datasets. (B) In unsupervised ML, during training (left), the model learns intrinsic relationships within the raw input data, without target labels. In inference (right), the trained model can then be used to detect patterns within the unseen raw dataset.

(Fig. 2A). Intuitively, having more data for training in ML usually means that the model is exposed to more variability, dynamic range and noise. Overall, it enables ML models to focus more effectively on the data, discerning noise from meaningful signals (Lei et al., 2019).

Training with a broad dataset also helps to mitigate overfitting (Montesinos López et al., 2022), a phenomenon where the model captures irrelevant features on the training data rather than learning the generalisable patterns. By providing more acquired data, the model is less likely to 'memorise' specific instances in the training set and learns how to generalise across a broader spectrum of examples. For instance, in cell segmentation tasks (Stringer et al., 2021), overfitting is associated with wrongfully detected objects in the background, such as experimental artefacts (e.g. cell debris), instead of accurately segmenting the target biological structures.

The quantity of data required for training also scales with the capacity of the model, that is, how many parameters the model has to estimate during training. Typically, when an analysis task becomes more complex, so does the architecture of the ML model used, such as going from the binary pixel classification of cell types (e.g. healthy versus disease phenotypes), to the classification of multiple descriptive disease phenotypes, or going from pixel classification to more complex tasks such as the segmentation of individual cells, which involves the prediction of detailed and structured outputs (Stringer et al., 2021). This leads to learning a higher number of parameters of the model and hence a need for more training data.

It is pertinent to ask, "how much data do I need?". Unfortunately, there is no simple answer to this question, as the required amount of data highly depends on the application. Generally, large datasets are necessary, but in some cases, an excess of data can hinder the analysis. For instance, in developing a method for predicting fluorescent labels in unlabelled images using deep learning (Christiansen et al., 2018), researchers acquired ∼200,000 diverse images to train the model for sufficient generalisation. Conversely, for tasks involving the interpretation of specific biological data, such as identifying visual factors that determine embryo quality, only ∼2000 images were needed (Rotem et al., 2024). In cases like the latter, where the goal is to derive insights from highly specific datasets, exposing the model to extensive data diversity (with hundreds of thousands of cells) is not advisable, as it could lead to overgeneralisation. Nevertheless, for most applications in microscopy, the number of required images to train a model from scratch typically falls within the thousands, which is still a very high volume compared to the norm in biological research.

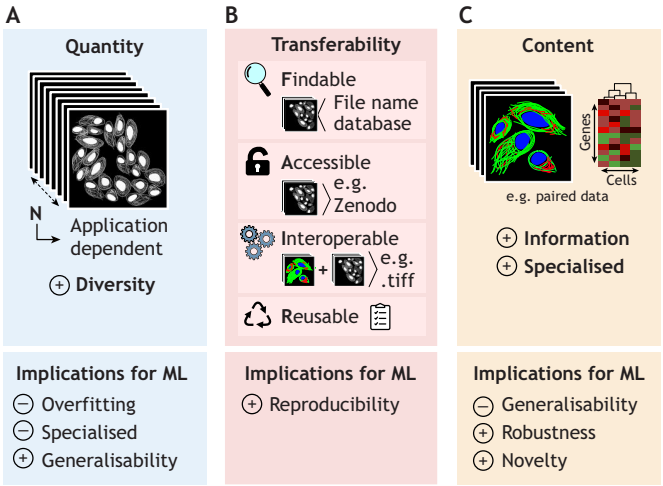

Fig. 2. Data characteristics for ML. Representation of specific data properties (top) and corresponding implications for ML models (bottom). Represents (A) data quantity, (B) data transferability and (C) data content.

In response to the need for acquiring extensive microscopy datasets, high-throughput technologies and automated acquisition platforms are increasingly being developed (Mahecic et al., 2019). These are especially useful for techniques that are inherently low throughput. For instance, single-molecule localization microscopy (SMLM) (Lelek et al., 2021) allows the acquisition of images with tens of nanometres in resolution, but each image is restricted to a small field-of-view (<1000 μm^2) and requires thousands of time frames to collect. For this, efforts have been made to automatically collect SMLM data in multi-well arrays, leading to a significant collection throughput (Beghin et al., 2017). Moreover, even for microscopy techniques that are not as low throughput as SMLM, such as confocal or widefield microscopy, efforts have been made to automate the acquisition process (Zehrer et al., 2024), with the help of tools that allow the microscope control (Moreno et al., 2021; Pinkard et al., 2021). Strikingly, researchers have recently started developing imaging farms, which consist of fully automated microscopes that acquire images in parallel. This falls within the scope of laboratory automation for autonomous and automated imaging (Bai et al., 2022), which promises to dramatically increase data acquisition speed and volume. Besides hardware-based approaches, synthetic datasets are a convenient strategy to increase the data size for model training, particularly in low-throughput techniques like SMLM (Speiser et al., 2021).

Although such technological developments indeed lead to an increase in the amount of collected raw images, they do not provide the labels necessary for supervised ML applications. For example, producing pixel-level segmentation of cells typically requires extensive and tedious manual annotation and curation (Thul et al., 2017). This often requires interdisciplinary collaborations, time and resources, and thus continues to constitute a major bottleneck for ML implementations. One way to circumvent this is by the use of active learning, where a small labelled subset of data is used for initial training. The model identifies uncertain predictions, which are then corrected by an external source (e.g. a human). These newly labelled data are added to the training set, iteratively improving the performance of the model without requiring a fully labelled dataset upfront (Kutsuna et al., 2012). Another inventive way of tackling the annotation bottleneck involves citizen science, where the general public contributes to scientific data labelling. 'Project Discovery' is a mini-game where players annotate images provided from The Human Protein Atlas (https://www.proteinatlas.org/; Sullivan et al., 2018). These annotations were then used for a subcellular map of the human proteome (Thul et al., 2017). After having data annotations, a common strategy to increase the overall data used for training in supervised ML (i.e. images and corresponding target labels) is by using artificial data augmentation techniques. These include rotations, flips and noise injection to expand the existing dataset.

In practice, a combination of high-throughput imaging, synthetic data generation, annotation and data augmentation is often employed to create comprehensive datasets for training ML models in microscopy. This often requires interdisciplinary collaborations, time and resources, and thus continues to constitute a major bottleneck for ML implementation. Increasing data quantity is also not always an available option. This could be due to the rarity of certain biological samples, or limitations associated with the acquisition method itself, such as phototoxicity (Icha et al., 2017).

Transferability and availability

Image processing and analysis tasks are often repetitive in nature, or at least partly overlapping. In this context, ML models are very promising solutions to provide 'fit for most' general tools. This, however, requires presenting the ML models with broad collections of datasets during the initial training phase. ML models used to segment nuclei (Schmidt et al., 2018 preprint), for instance, provide reliable results independently of the cell type, as long as they have been trained with a broad range of cell types in the first place. Because acquiring large-scale and diverse datasets is costly and remains challenging overall (Ellenberg et al., 2018) (e.g. due to the time and financial costs of multi-disciplinary collaborations), researchers have been putting efforts towards making data more transferable and openly shared within the community (Bagheri et al., 2022; Hohlbein et al., 2022). For data to be inherently transferable (Fig. 2B), it should follow specific standardised formats and principles, such as 'Findable, Accessible, Interoperable, Reusable' (FAIR) (Wilkinson et al., 2016).

Findable data is data that is properly identified (e.g. by cell type and imaging technique), with sufficient metadata that describes conditions, such as cell treatments and imaging parameters, and is easily searchable in a resource (e.g. database). Data accessibility ensures datasets are readily available to use through public repositories and data sharing platforms, such as Zenodo (https://zenodo.org/). Interoperable data can be seamlessly used and combined with other data in workflows without any prior modifications (e.g. using standardised data formats; Goldberg et al., 2005; Linkert et al., 2010). Reusable data is structured and documented for easy future use, with clear documentation of protocols and workflows used for acquisition and processing, to ensure findings can be easily replicated.

Data transferability still grapples with limitations, despite the efforts of researchers to adhere to it. In microscopy, the data sharing 'culture' only emerged relatively recently, and publishing open access data often comes with limited gains. Arguably, a major issue hindering data reuse is trust. Publication bias (Lee et al., 2024) (i.e. publishing unrepresentative 'good looking' datasets), is often an aspect research groups are very wary of when looking for data. In addition, researchers are still frequently incentivised to acquire fresh data for publications, even when the novelty of their work lies in the analysis pipeline or software development, for which a systematic requirement of new datasets is highly debatable. In an effort to counteract this, researchers and developers are increasingly using and developing platforms for data sharing. These can be more general or multi-purpose platforms [e.g. Zenodo and GitHub (https://github.com/)] or field-specific (Ouyang et al., 2022b). For example, the Image Data Resource (IDR; https://idr.openmicroscopy.org/) (Williams et al., 2017) and the Bioimage archive (https://www.ebi.ac.uk/bioimage-archive/; EMBL) are open-source platforms for publishing imaging data. Besides, depending on the specific analytical task, it is possible today to directly use pre-trained models available on open cloud platforms (Chamier et al., 2021). BioImage Model Zoo (https://bioimage.io; Ouyang et al., 2022a preprint), for instance, is an online platform where standardised data and pre-trained models can be easily shared and used, improving accessibility and reproducibility of ML workflows in microscopy.

Transferability has enabled scientists to take advantage of large amounts of data and train robust ML models for a specific task. By making these models available, transfer learning, where pre-trained models are reused and 'readjusted' to new data, also becomes possible (Alzubaidi et al., 2021). Furthermore, these pre-trained models can be fine-tuned with a significantly smaller amount of new data, often requiring only tens of samples rather than thousands, making the process more efficient and accessible for specific use cases.

Content – information-rich datasets for novel knowledge with ML

High-content data refers to datasets rich in information, often captured using specialised equipment, techniques and assays (Fig. 2C). In recent decades, there has been a notable shift towards acquiring high-content data, symptomatic of a 'race for information-rich datasets' across fields, with the hope of gaining new insights and generating novelty. This has been made possible by notable progress in data acquisition, curation and annotation. In microscopy, this typically translates to an increased number of dimensions, such as the number of markers, as well as longer imaging timescales, improved resolution, and imaging of multi-cellular structures or full organisms. We discuss here how ML might become instrumental in enhancing how these 'pretty pictures' can be integrated into understanding key biological processes.

High-content datasets indeed provide ML models with a wealth of diverse features and patterns, enabling them to learn complex relationships and capture subtle variations, leading to more accurate and robust predictions (Gong et al., 2019). Interestingly, high-content datasets might also consist of a combination of results obtained with multiple microscopy modalities (i.e. paired data). ML analysis pipelines easily scale to varied data types as inputs (Schwartz et al., 2023). Effectively, this has the potential to provide a holistic understanding of complex biological processes, with inputs collected over multiple scales from genomes and molecules to cells and whole organisms.

High-content screening (HCS) assays (Abraham, 2004; Boutros et al., 2015; Lin et al., 2020) generate high-content datasets in which a very high number of conditions, perturbations and/or targets or markers can be investigated. For example, a genome-wide HCS method has been used to systematically test the function of genes through automated analysis of different phenotypic features in cells (Chia et al., 2010). This involved the use of siRNAs to silence the expression of thousands of genes and collecting data on the resulting phenotypic changes in the human embryonic stem cells. Using ML trained on this complex dataset, genes involved in regulating stem cell identity were successfully identified. Image-based cell profiling uses a similar strategy and is used in fields like drug discovery and functional genomics (Caicedo et al., 2016, 2017). For example, fixed cell features (e.g. cell morphology or staining intensity profiles of the cell) can be extracted from high-throughput microscopy acquisition and analysed with ML, and be reliably used to identify promising new drug compounds and their effects (Bray et al., 2016). Considering the success of these implementations, future developments are likely to diversify the range of the source of measurements or features. Morphological features, relying on stains (such as for membrane, actin or nuclei, among others), can be enriched with measurements like biosensors (Dunn et al., 2021), polarisation of the transmitted light (De Angelis et al., 2019) or Förster resonance energy transfer (FRET) measurements (Sekar and Periasamy, 2003), to generate quantitative phenotypic signatures for cells.

Another compelling example of an image-based profiling assay is the multiplexed error-robust fluorescence *in situ* hybridisation (MERFISH) system (Chen et al., 2015). MERFISH is a highly multiplexed imaging technique, allowing thousands of RNA species to be imaged in single cells through hybridisation of encoded RNA probes. The technique provides a unique insight into the intertwined relationship between RNA spatial distribution and genetic content with cellular processes outcome.

Microscope systems, such as the lattice light-sheet microscope (Chen et al., 2014) and MoNaLISA (Masullo et al., 2018), a nanoscope capable of imaging the entire cell volume at low light intensities at a scale of 45–65 nm, have been developed specifically to produce 3D movies with improved temporal resolution and lowered phototoxicity. Although movies of live cells or organisms generated using these techniques often follow a high number of molecular targets and provide unique and highly qualitative information, they cannot be efficiently analysed with state-of-the-art conventional analysis tools (Liu et al., 2021). We anticipate that such imaging techniques will vastly benefit from being combined with ML holistic analytics that can scale to high dimensional data, retain temporal information and incorporate nonlinear relationships.

Although most of the aforementioned methods aim to increase information content overall, a new avenue of research in microscopy focuses on designing smart 'content-driven' microscopes (Alvelid et al., 2022; André et al., 2023; Morgado et al., 2024; Shi et al., 2024), which can be used to overcome certain limitations. For example, acquisition limitations (e.g. photobleaching and phototoxicity) can be partially circumvented by acquiring content-enriched (i.e. better spatial and/or temporal resolution) data only when the biological process of interest occurs (Mahecic et al., 2022). These smart microscopes already heavily rely on ML and have the potential to address well-defined biological questions, such as mitochondria division events (Mahecic et al., 2022).

Overall, researchers face the challenge of navigating data acquisition trade-offs, balancing the need for rich, comprehensive data with considerations of cost, scalability and often human-made annotations. Ultimately, the selection of acquisition methods depends on the specific goals, constraints and priorities of each research project, highlighting the importance of thoughtful decision-making in data acquisition strategies for ML applications.

What does ML bring to microscopy?

ML serves several key utilities, including data curation, exploration, prediction and explanation. In this section, we dive into each ML utility, providing examples of its use in several fields and how these strategies are or could be implemented for microscopy.

Data curation – improving microscopy data

The definition of ML-based data curation might vary across different fields and disciplines. Here, we adopt a broader definition, where curation consists of automated techniques, taking raw data as input, and aiming at increasing quality and usefulness (e.g. data cleaning, the process of removing incorrect or uninformative data within a dataset), as well as reproducibility (Fig. 3A). These pre-analysis pipelines are a requirement for reliable quantification, scalability and decision-making processes at later stages. Representative examples of ML-based data curation in microscopy include methods based on deep learning (DL), a subcategory of ML. This involves image filtering, segmentation and cell tracking (Chai et al., 2023), as well as improving the signal-to-noise ratio. Noise2Void (Krull et al., 2018 preprint), for example, is a widely adopted unsupervised DL model for microscopy image denoising. Tasks such as cell or nuclei segmentation are automated with the help of supervised ML models such as Stardist (Schmidt et al., 2018 preprint). Cellpose (Stringer et al., 2021), a DL-based generalist model, allows the segmentation of cells from a wide range of image types. More recently, a computer vision foundation model capable of generalising segmentation of any object – the Segment Anything Model (SAM) – was released (Kirillov et al., 2023 preprint). SAM has been successfully applied to microscopy datasets (Archit et al., 2023 preprint), accelerating the annotation of images and their ground truth masks – the correct outlines – for cells or organelles thanks to the user prompt. DL

Journal of Cell Science (2024) 137, jcs262095. doi:10.1242/jcs.262095

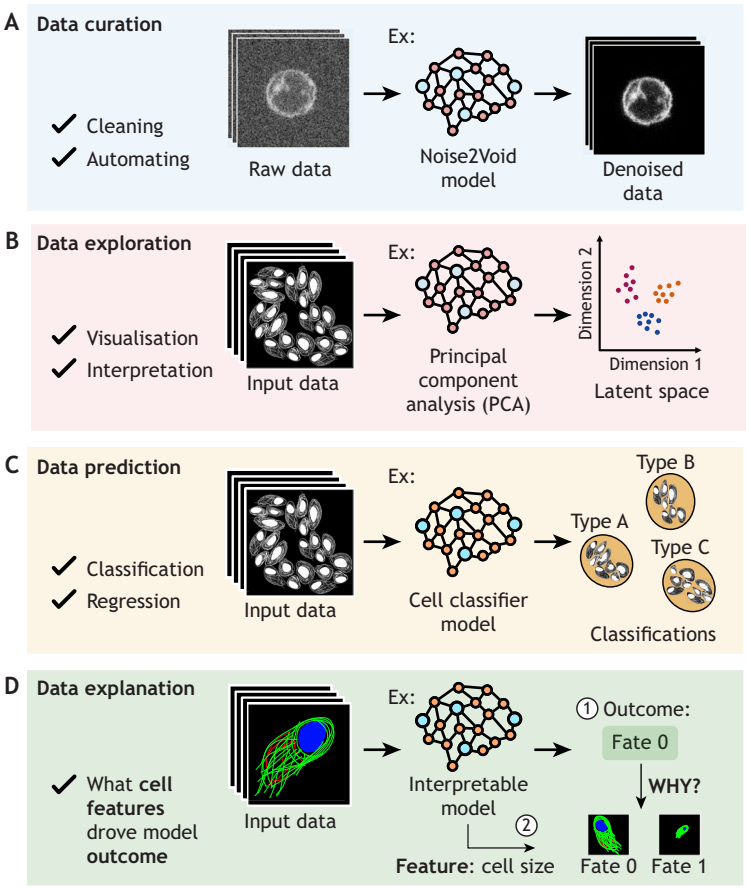

Fig. 3. ML utilities for microscopy. (A) Data curation, which includes data cleaning and processing task automation. One use-case example in microscopy is data denoising using the Noise2Void ML model. (B) Data exploration, encompassing data visualisation and interpretation. In microscopy, this can be done by doing a principal component analysis (PCA) for dimensionality reduction of the data into a latent space representation, where data is plotted as a 2D point pattern. (C) Data prediction, which involves classification and regression. One prominent example of microscopy data prediction is the classification of cell types, by using a cell classifier model. (D) Data explanation, which aims at finding what cell features drove a specific model outcome. In this example, given a model outcome (cell fate 0), using an interpretable model would help interpret why the model chose fate 0 for the input data (in this case, it was due to the cell size). Ex, example.

methods have also been employed for the removal of outliers from image datasets, with models being capable of generating a low-dimensional representation of the data representing the shapes of cells. This representation allowed the profiling of cellular phenotypes and consequent discrimination of outliers (Burgess et al., 2024). For all the above-mentioned examples, pre-trained models are readily available (Chamier et al., 2021; Ouyang et al., 2022a preprint, 2022b), allowing researchers to use models built for specific tasks for inference in their own data.

Data exploration – applications in cell biology

Microscopy often produces highly dimensional and complex datasets. ML-based exploration provides tools to explore these datasets, with the aim of unravelling novel patterns, relationships and trends within the data itself that are not directly accessible from the collected images. Unlike manual feature engineering, where domain experts predefine features to be extracted based on prior knowledge, explorative models autonomously discover relevant features inherent to the data and generate a novel representation of the data without requiring explicit human intervention (Fig. 3B). They consist of a fully data-driven approach.

Unsupervised ML algorithms like principal component analysis (PCA) (Jolliffe and Cadima, 2016) or uniform manifold approximation and projection (UMAP) (McInnes et al., 2018), aim to convert the input data into a low-dimensional space representation, called latent space (Fig. 3B). Latent spaces typically provide users with a much simpler visualisation (typically a 2D or 3D point pattern) of the original high-dimensional dataset (i.e. images). By exploring the latent space generated by these dimensionality reduction algorithms, where clustering and trends of the original data are easily accessible, it is possible to unravel cellular phenotypes or states with no *a priori* knowledge. Dimensionality reduction algorithms allowed the discovery of differentiation pathways of dendritic cells (See et al., 2017), the quantification of myofibroblast cells activation (Hillsley et al., 2022), and to differentiate between excitatory and inhibitory neurological synapses (Unterauer et al., 2024). However, these dimensionality reduction techniques come with inherent limitations. Though they aim to preserve the most relevant information, the process of reducing high-dimensional data to a low-dimensional representation inevitably results in some information loss. Traditional methods like PCA and UMAP might struggle to capture complex, non-linear relationships in the data, potentially overlooking subtle but important features (Diaz-Papkovich et al., 2021; Fernandes et al., 2024). To address this issue, autoencoders (Bank et al., 2020 preprint), a type of unsupervised DL method, have been increasingly adopted. During training, autoencoders compress the original data (i.e. images) into the latent space. Then, using only the information from the latent space (i.e. point pattern), autoencoders reconstruct it back into an image that should 'look' exactly like the original one. This approach ensures that the essential information about the data used to train the model is retained, as the reconstruction process forces the model to learn how to retrieve the original data from the low-dimensional latent space. This opens doors for using these models to interpret cell biology data (i.e. identifying underlying cellular processes and/or dynamics that could not be directly observable from acquired data) and infer novel information. For example, latent spaces generated by autoencoders have been used to infer cell cycle phases (Ulicna et al., 2023 preprint), cell lineage information (Yang et al., 2020) and cell outcome after drug treatment (Umarov et al., 2021).

The approach of learning from latent space representations of complex data, also known as representation learning, has shown immense potential in microscopy and beyond (Blaschke et al., 2018; Gómez-Bombarelli et al., 2018; Gunawan et al., 2023; Pyzer-Knapp et al., 2022). Importantly, the scientific community is making strides in democratising these tools. User-friendly platforms like CellProfiler Analyst 3.0 (Stirling et al., 2021) are making advanced ML methods for data exploration and quantification accessible to a broader range of researchers. Additionally, standardised workflows for learning morphological representations of microscopy images are being developed (Caicedo et al., 2017).

After using techniques for data exploration, it is common to perform data prediction. This is because the insights gained from exploring datasets inform the selection and engineering of both features as well as output classes, which is crucial for accurate predictions.

Data prediction – advancing cell biology and disease research

Data prediction with ML refers to forecasting outcomes based on the training data. Data prediction typically consists of using supervised ML models for classification (Fig. 3C) or regression tasks. Common applications of ML-driven classification are cell state (e.g. healthy and diseased cells), cell type or cell cycle phase. For these tasks, software like Advanced Cell Classifier (Piccinini et al., 2017), containing user-friendly access to ML models for classification, have been made available. Models for regression have also been applied in microscopy research, for tasks like estimating cell counting (Xue and Ray, 2017 preprint) and cell detection (Xue and Ray, 2017 preprint). Moreover, data prediction can be done directly in latent space representations produced for exploration (Lafarge et al., 2019; Palma et al., 2023 preprint; Yang et al., 2024; Chow et al., 2022; Lu et al., 2019). The examples mentioned in the data exploration section have all been used for prediction, namely, dendritic cell differentiation pathways (See et al., 2017), myofibroblast cell subgroups (Hillsley et al., 2022) and types of neurological synapses (Unterauer et al., 2024), among others.

Data explanation – understanding microscopy data with ML

Most of the models mentioned for data curation, exploration and prediction are based on DL. These models are typically termed as 'black boxes', given that their decision-making processes are not understandable by humans. However, if we could understand how or why these models made a prediction from the input dataset, we cannot only understand more about how the model works but also about the data itself. This lies within the field of 'explainable and interpretable' ML (Rudin, 2019). In the context of microscopy, this translates to asking, 'what visual features in the image contributed the most for the outcome of the model' (Fig. 3D). Researchers have started to develop models that are inherently interpretable for microscopy images, by coupling classifier models with interpretable latent space representations (for an overview, see Soelistyo and Lowe, 2024 preprint). For example, from a live-cell movie, a model was developed that predicted each cell outcome (i.e. apoptosis or mitosis). From this model, what the most important biophysical properties of cells that drove the specific outcome could be determined – the most important feature was related to cellular environment (i.e. cellular density) (Soelistyo et al., 2022). In another study, an interpretable model was built to understand what drives the classification of embryo cells into good or bad quality for *in vitro* fertilisation. It was found that the most important cellular features were the blastocyst size and the quality of trophectoderm (a layer surrounding the blastocyst) (Rotem et al., 2024). These are just a few examples of an emerging field of AI that potentially allow us to learn from what these models have learned (Sadafi et al., 2023; Xu et al., 2021). In the future, we see that questions that can be addressed range from what biological marker contributes the most for the cell outcome, or, in the case of live-cell movies, when did a given outcome start being predictable.

In conclusion, the integration of ML techniques in microscopy workflows – from data curation and exploration to prediction and explanation – offers a powerful method that not only enhances the quality and utility of microscopic data but also provides a gateway to uncovering novel insights into cellular mechanisms.

Challenges and risks

Although relatively new in microscopy, ML has been used for decades in fields such as computer vision (Voulodimos et al., 2018) and natural language processing (Nadkarni et al., 2011). A number of pitfalls and challenges associated with ML use have been identified in this context and partially addressed. In this section, we anticipate the impact of such limitations for ML in microscopy, alongside proposed solutions to tackle them.

Dataset shift

Dataset shift describes when there are significant differences between the dataset used for training and the dataset used for inference (Uhlmann et al., 2022), and results in less accurate or biased inference on the investigated dataset. Artefacts associated with data shift are common for models with very well-defined tasks that cannot easily be generalisable across cell types or conditions (Hallou et al., 2021). Concept shift, more specifically, consists of using different input data types between training and inference. In microscopy, that typically translates to training a model on a given imaging modality while using the trained model to infer information on another type of imaging modality. For example, if a segmentation model is trained on brightfield microscopy images and later used on fluorescence microscopy images, the mismatch in imaging modalities can introduce a concept shift, potentially leading to worsened segmentation accuracy. Some models for which the tasks are not specific to either the imaging modality or the biological question being asked can bypass data shift by being trained on broad scope datasets. For example, the Segment Anything Model (Kirillov et al., 2023 preprint) has proven to be extremely generalisable for segmentation of any object, given that it was trained on millions of images, with hundreds of millions of parameters optimised during training.

Novelty hindering

Strikingly, generalisable models are far from a fit for all applications. The risk with very general models designed for simple repetitive tasks is that when applied to a well-posed biological question, they hinder any information in the inputted distribution that deviates from that simple task. This leads to hindering novelty, overlooking statistically low-occurring representations in the data. For instance, in image recolouring (Wang et al., 2022), training data contains rare or unique cellular structures or biological features that have limited occurrence. The generalisation of the model is therefore likely to result in less accurate and distinctive colorizations for these specific elements. Interestingly, this issue has partially been alleviated in fields like natural language processing, where selective sampling strategies (Figueroa et al., 2012; Magar and Farimani, 2023; Weinstein et al., 2019 preprint) are incorporated to ensure sufficient representation of rare or unique instances in the training dataset. This involves using

active learning methods that focus on collecting data points that are most informative or challenging for the model, such as sentences that contain ambiguous content. Because there is a tradeoff between dataset shift and novelty when it comes to what type of data should be used for training, an important prerequisite for successful analyses is to determine whether the question being asked requires a general or specific model.

Hallucinations

Another possible challenge related to using generative ML models is the phenomenon known as 'computer hallucinations', wherein ML models generate misleading or fake information with seemingly great confidence, posing a threat to their reliability. Although such artefacts, e.g. artificial sharpening (Marsh et al., 2018), have also been observed in conventional image processing, it is a challenge that needs to be addressed if ML is to become a broadly used tool for image reconstruction and synthetic image generation. A commonly used safeguard is to retain some user feedback in the loop with reinforcement learning (RL) (Cao et al., 2023 preprint). RL consists of a trial-and-error-based learning strategy, very similar to the human learning process; the user is presented with a small subset of the proposals generated by the trained model and then asked to discriminate between 'likely' and 'unlikely' proposals, which are associated with positive rewards and penalties, respectively. This partial annotation systematically assesses hallucinations as unlikely. The model then learns to adjust its parameters to maximise rewards and minimise penalties, leading to a decrease of unlikely outputs, which includes hallucinations as well as unrealistic proposals. This approach has been successfully applied for drug design, for which ML is very well established. For instance, by incorporating human preferences into the ML models, such as known biological properties of compounds, the accuracy and usability of the proposed model is vastly improved in novel drug design (Liu et al., 2023). In the case of microscopy, this framework translates to providing the user with a small subset of the images or output generated by the trained model to annotate as likely or unlikely based on *a priori* knowledge of the biological process studied. More recently, DL-based architectures incorporating physics-informed components have been successfully implemented in order to not only mitigate hallucinations but also to improve model accuracy (Burns and Liu, 2022 preprint; Li et al., 2022). For example, the Richardson–Lucy DL network (RLN) was developed for the deconvolution of fluorescence microscopy images (Li et al., 2022). RLN adds the image formation process as a model constraint, allowing more accurate image reconstructions. This approach has been used in super-resolution microscopy reconstruction, achieving better performance than widely used DL methods, such as content-aware image restoration (CARE) (Weigert et al., 2018).

Data privacy

Data privacy has emerged as a key challenge in the context of health data analysis or more broadly data-driven life science over the past decade (Holub et al., 2023; Jadon and Kumar, 2023; Norori et al., 2021). ML-based pipelines are far from immune to privacy leakage. The community often considers that when working with mostly *in vitro* systems for microscopy, this issue can be overlooked – but this is only partially accurate. With the rise of smart, high-throughput computational microscopy strategies for diagnosis and personalised medicine, privacy will inevitably become a parameter to mitigate with. Overall, data privacy itself remains very much an active field of research. Although methods such as differential privacy, where noise is added to a sample from an individual in order to protect its privacy, but keep the global distribution (Abadi et al., 2016) are relatively well-established and provide privacy guarantees, they tend to impact negatively on the usability of the output. More tailored research into how microscopy images with sensitive patient information can be handled, will need to be undertaken in the coming years.

Bias

Because ML relies on data for training, it is particularly susceptible to biases (Chandak and Tatonetti, 2020; Grossmann et al., 2023; Lee et al., 2024; Norori et al., 2021; Takan et al., 2023). One salient example in life sciences comes from an investigation of a commercial algorithm for prioritising care for patients, which was found to underestimate illness severity in black patients relative to white patients due to imbalanced data from those populations (Obermeyer et al., 2019). Within microscopy, analogous representation issues could emerge if models are trained on datasets that undersample certain experimental conditions, cell types, rare phenotypes or imaging modalities. By examining the training datasets and their distributions, evaluating model performance across subgroups, and proactively addressing observed skew through pre-processing methods, data augmentation and weighting schemes, we can work to maximise discovery potential and reliability in microscopy data analysis.

Conclusions and future outlooks

As ML becomes an essential tool in life science, we reflect in this Review on its integration into microscopy-based research. With users in mind, it is imperative that we truly understand the technical requirements of ML-based pipelines so we can fully benefit from its uses in microscopy image processing and analysis. Similar to microscopy acquisition, ML-based analytics are bound by a fundamental trade-off: what kind of datasets do you have and what do you want to get out of it? This comes down to three important dataset characteristics for ML (quantity, transferability and content), how they can be prioritised during acquisition, and how they impact on the trained models. Three potential uses of ML in microscopy include data curation, exploration, prediction and explanation. These span from more task-driven strategies, with more 'human-guided' objectives, such as cell segmentation, to open-ended exploration of visual patterns within images in a more data-driven manner. Importantly, lessons learned from fields with more extensive ML experience can inform strategies for mitigating challenges and risks associated with ML for microscopy. By leveraging these insights, researchers can navigate complexities such as dataset shift, hallucinations, novelty hindering, model biases and privacy concerns more effectively, fostering greater reliability and reproducibility in microscopy research.

In conclusion, the intersection of ML and microscopy holds immense promise for advancing our understanding of life and disease, and unlocking novel insights into complex biological systems. Importantly, we should pursue a symbiotic path where ML aids in data analysis and hypothesis generation, while biologists continue to provide essential domain knowledge and validate discoveries through experimental interventions. Rather than viewing AI as a replacement for human expertise, we suggest it is very much a complementary tool that enhances our ability to decipher the complexities of biological systems.

Competing interests

The authors declare no competing or financial interests.

Funding
Our work in this area is supported by the Knut and Alice Wallenberg foundation (KAW 2020.0239) and a Data-Driven Life Science (DDLS) fellowship (I.C., E.L.S.B. and J.G.).

Special Issue
This article is part of the Special Issue 'Imaging Cell Architecture and Dynamics', guest edited by Lucy Collinson and Guillaume Jacquemet. See related articles at https://journals.biologists.com/jcs/issue/137/20.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K. and Zhang, L. (2016). Deep Learning with Differential Privacy. In: Proceedings of the 2016, ACM SIGSAC Conference on Computer and Communications Security, pp. 308-318.

Abraham, V. (2004). High content screening applied to large-scale cell biology. *Trends. Biotechnol.* **22**, 15-22. doi:10.1016/j.tibtech.2003.10.012

Alvelid, J., Damenti, M., Sgattoni, C. and Testa, I. (2022). Event-triggered STED imaging. *Nat. Methods* **19**, 1268-1275. doi:10.1038/s41592-022-01588-y

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M. and Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **8**, 1-74. doi:10.1186/s40537-021-00444-8

André, O., Kumra Ahnlide, J., Norlin, N., Swaminathan, V. and Nordenfelt, P. (2023). Data-driven microscopy allows for automated context-specific acquisition of high-fidelity image data. *Cell Reports Methods* **3**, 100419. doi:10.1016/j.crmeth.2023.100419

Archit, A., Nair, S., Khalid, N., Hilt, P., Rajashekar, V., Freitag, M., Gupta, S., Dengel, A., Ahmed, S. and Pape, C. (2023). Segment anything for microscopy. *bioRxiv* 2023.08.21.554208. doi:10.1101/2023.08.21.554208

Bagheri, N., Carpenter, A. E., Lundberg, E., Plant, A. L. and Horwitz, R. (2022). The new era of quantitative cell imaging—challenges and opportunities. *Mol. Cell* **82**, 241-247. doi:10.1016/j.molcel.2021.12.024

Bai, J., Cao, L., Mosbach, S., Akroyd, J., Lapkin, A. A. and Kraft, M. (2022). From platform to knowledge graph: evolution of laboratory automation. *JACS Au* **2**, 292-309. doi:10.1021/jacsau.1c00438

Bank, D., Koenigstein, N. and Giryes, R. (2020). Autoencoders. *arXiv* arXiv.2003.05991. doi:10.48550/arXiv.2003.05991

Beghin, A., Kechkar, A., Butler, C., Levet, F., Cabillic, M., Rossier, O., Giannone, G., Galland, R., Choquet, D. and Sibarita, J.-B. (2017). Localization-based super-resolution imaging meets high-content screening. *Nat. Methods* 14, 1184-1190. doi:10.1038/nmeth.4486

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J. and Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645. doi:10.1126/science.1127344

Blaschke, T., Olivecrona, M., Engkvist, O., Bajorath, J. and Chen, H. (2018). Application of generative autoencoder in de novo molecular design. *Mol. Inform.* **37**, 1700123. doi:10.1002/minf.201700123

Boutros, M., Heigwer, F. and Laufer, C. (2015). Microscopy-based high-content screening. *Cell* **163**, 1314-1325. doi:10.1016/j.cell.2015.11.007

Bray, M.-A., Singh, S., Han, H., Davis, C. T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S. M., Gibson, C. C. and Carpenter, A. E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757-1774. doi:10.1038/nprot.2016.105

Burgess, J., Nirschl, J. J., Zanellati, M. C., Lozano, A., Cohen, S. and Yeung-Levy, S. (2024). Orientation-invariant autoencoders learn robust representations for shape profiling of cells and organelles. *Nat. Commun.* **15**, 1022. doi:10.1038/s41467-024-45362-4

Burns, Z. and Liu, Z. (2022). Untrained, physics-informed neural networks for structured illumination microscopy. *arXiv* arXiv.2207.07705. doi:10.48550/arXiv.2207.07705

Caicedo, J. C., Cooper, S., Heigwer, F., Warchal, S., Qiu, P., Molnar, C., Vasilevich, A. S., Barry, J. D., Bansal, H. S., Kraus, O. et al. (2017). Data-analysis strategies for image-based cell profiling. *Nat. Methods* **14**, 849-863. doi:10.1038/nmeth.4397

Caicedo, J. C., Singh, S. and Carpenter, A. E. (2016). Applications in image-based profiling of perturbations. *Curr. Opin. Biotechnol.* **39**, 134-142. doi:10.1016/j.copbio.2016.04.003

Cao, Y., Sheng, Q. Z., McAuley, J. and Yao, L. (2023). Reinforcement learning for generative AI: a survey. *arXiv* arXiv.2308.14328. doi:10.48550/arXiv.2308.14328

Chai, B., Efstathiou, C., Yue, H. and Draviam, V. M. (2023). Opportunities and challenges for deep learning in cell dynamics research. *Trends Cell Biol.*, S0962-8924(23)00228-3. doi:10.1016/j.tcb.2023.10.010

Chamier, L., Laine, R. F., Jukkala, J., Spahn, C., Krentzel, D., Nehme, E., Lerche, M., HernÃ¡ndez-Pérez, S., Mattila, P. K., Karinou, E. et al. (2021). Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nat. Commun.* **12**, 2276. doi:10.1038/s41467-021-22518-0

Chandak, P. and Tatonetti, N. P. (2020). Using machine learning to identify adverse drug effects posing increased risk to women. *Patterns* **1**, 100108. doi:10.1016/j.patter.2020.100108

Chen, B.-C., Legant, W. R., Wang, K., Shao, L., Milkie, D. E., Davidson, M. W., Janetopoulos, C., Wu, X. S., Hammer, J. A., Liu, Z. et al. (2014). Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 6208. doi:10.1126/science.1257998

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, 6233.

Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y. L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S. et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316-320. doi:10.1038/nature09531

Chow, Y. L., Singh, S., Carpenter, A. E. and Way, G. P. (2022). Predicting drug polypharmacology from cell morphology readouts using variational autoencoder latent space arithmetic. *PLoS Comput. Biol.* **18**, e1009888. doi:10.1371/journal.pcbi.1009888

Christiansen, E. M., Yang, S. J., Ando, D. M., Javaherian, A., Skibinski, G., Lipnick, S., Mount, E., O'neil, A., Shah, K., Lee, A. K. et al. (2018). In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792-803. doi:10.1016/j.cell.2018.03.040

De Angelis, A., Ferrara, M. A., Coppola, G., Di Matteo, L., Siani, L., Dale, B., Coppola, G. and De Luca, A. C. (2019). Combined Raman and polarization sensitive holographic imaging for a multimodal label-free assessment of human sperm function. *Sci. Rep.* **9**, 4823. doi:10.1038/s41598-019-41400-0

Diaz-Papkovich, A., Anderson-Trocmé, L. and Gravel, S. (2021). A review of UMAP in population genetics. *J. Hum. Genet.* **66**, 85-91. doi:10.1038/s10038-020-00851-4

Dunn, J., Kidzinski, L., Runge, R., Witt, D., Hicks, J. L., Schüssler-Fiorenza Rose, S. M., Li, X., Bahmani, A., Delp, S. L., Hastie, T. et al. (2021). Wearable sensors enable personalized predictions of clinical laboratory measurements. *Nat. Med.* **27**, 1105-1112. doi:10.1038/s41591-021-01339-0

Ellenberg, J., Swedlow, J. R., Barlow, M., Cook, C. E., Sarkans, U., Patwardhan, A., Brazma, A. and Birney, E. (2018). A call for public archives for biological image data. *Nat. Methods* **15**, 849-854. doi:10.1038/s41592-018-0195-8

Fernandes, V., Carvalho, G., Pereira, V. and Bernardino, J. (2024). Analyzing data reduction techniques: an experimental perspective. *Appl. Sci.* **14**, 3436. doi:10.3390/app14083436

Figueroa, R. L., Zeng-Treitler, Q., Ngo, L. H., Goryachev, S. and Wiechmann, E. P. (2012). Active learning for clinical text classification: is it better than random sampling? *J. Am. Med. Inform. Assoc.* **19**, 809-816. doi:10.1136/amiajnl-2011-000648

Goldberg, I. G., Allan, C., Burel, J.-M., Creager, D., Falconi, A., Hochheiser, H., Johnston, J., Mellen, J., Sorger, P. K. and Swedlow, J. R. (2005). The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biol.* **6**, R47. doi:10.1186/gb-2005-6-5-r47

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P. and Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Sci.* **4**, 268-276. doi:10.1021/acscentsci.7b00572

Gong, Z., Zhong, P. and Hu, W. (2019). Diversity in machine learning. *IEEE Access* **7**, 64323-64350. doi:10.1109/ACCESS.2019.2917620

Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E. and Cunningham, W. A. (2023). AI and the transformation of social science research. *Science* **380**, 1108-1109. doi:10.1126/science.adi1778

Guldevall, K., Brandt, L., Forslund, E., Olofsson, K., Frisk, T. W., Olofsson, P. E., Gustafsson, K., Manneberg, O., Vanherberghen, B., Brismar, H. et al. (2016). Microchip screening platform for single cell assessment of NK cell cytotoxicity. *Front. Immunol.* **7**, 119. doi:10.3389/fimmu.2016.00119

Gunawan, I., Vafaee, F., Meijering, E. and Lock, J. G. (2023). An introduction to representation learning for single-cell data analysis. *Cell Rep. Methods* **3**, 100547. doi:10.1016/j.crmeth.2023.100547

Gustafsson, M. G. L. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* **198**, 82-87. doi:10.1046/j.1365-2818.2000.00710.x

Hallou, A., Yevick, H. G., Dumitrascu, B. and Uhlmann, V. (2021). Deep learning for bioimage analysis in developmental biology. *Development* **148**, dev199616. doi:10.1242/dev.199616

Hillsley, A., Santoso, M. S., Engels, S. M., Halwachs, K. N., Contreras, L. M. and Rosales, A. M. (2022). A strategy to quantify myofibroblast activation on a continuous spectrum. *Sci. Rep.* **12**, 12239. doi:10.1038/s41598-022-16158-7

Hohlbein, J., Diederich, B., Marsikova, B., Reynaud, E. G., Holden, S., Jahr, W., Haase, R. and Prakash, K. (2022). Open microscopy in the life sciences: quo vadis? *Nat. Methods* **19**, 1020-1025. doi:10.1038/s41592-022-01602-3

Holub, P., Müller, H., Bíl, T., Pireddu, L., Plass, M., Prasser, F., Schlünder, I., Zatloukal, K., Nenutil, R. and Brázdil, T. (2023). Privacy risks of whole-slide image sharing in digital pathology. *Nat. Commun.* **14**, 2577. doi:10.1038/s41467-023-37991-y

Icha, J., Weber, M., Waters, J. C. and Norden, C. (2017). Phototoxicity in live fluorescence microscopy, and how to avoid it. *BioEssays* **39**, 1700003. doi:10.1002/bies.201700003

Jadon, A. and Kumar, S. (2023). Leveraging generative AI models for synthetic data generation in healthcare: balancing research and privacy. In: *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)*, Istanbul, Turkiye, pp. 1-4. doi:10.1109/SmartNets58706.2023.10215825

Jan, M., Spangaro, A., Lenartowicz, M. and Mattiazzi Usaj, M. (2024). From pixels to insights: Machine learning and deep learning for bioimage analysis. *BioEssays* **46**, 2300114. doi:10.1002/bies.202300114

Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **374**, 20150202.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y. et al. (2023). Segment anything. *arXiv*, arXiv:2304.02643. doi:10.48550/arXiv.2304.02643

Klar, T. A. and Hell, S. W. (1999). Subdiffraction resolution in far-field fluorescence microscopy. *Opt. Lett.* **24**, 954. doi:10.1364/OL.24.000954

Krull, A., Buchholz, T.-O. and Jug, F. (2018). Noise2Void - learning denoising from single noisy images. *arXiv*, arXiv:1811.10980. doi:10.48550/arXiv.1811.10980

Kutsuna, N., Higaki, T., Matsunaga, S., Otsuki, T., Yamaguchi, M., Fujii, H. and Hasezawa, S. (2012). Active learning framework with iterative clustering for bioimage classification. *Nat. Commun.* **3**, 1032. doi:10.1038/ncomms2030

Lafarge, M. W., Caicedo, J. C., Carpenter, A. E., Pluim, J. P. W., Singh, S. and Veta, M. (2019). Capturing Single-Cell Phenotypic Variation via Unsupervised Representation Learning. In Proceedings of The, 2n. d International Conference on Medical Imaging with Deep Learning, pp. 315-325.

Lee, R. M., Eisenman, L. R., Khuon, S., Aaron, J. S. and Chew, T.-L. (2024). Believing is seeing – the deceptive influence of bias in quantitative microscopy. *J. Cell Sci.* **137**, jcs261567. doi:10.1242/jcs.261567

Lei, S., Zhang, H., Wang, K. and Su, Z. (2019). How Training Data Affect the Accuracy and Robustness of Neural Networks for Image Classification. In: *ICLR 2019 International Conference on Learning Representations*. ICLR, Appleton, USA.

Lelek, M., Gyparaki, M. T., Beliu, G., Schueder, F., Griffié, J., Manley, S., Jungmann, R., Sauer, M., Lakadamyali, M. and Zimmer, C. (2021). Single-molecule localization microscopy. *Nat. Rev. Methods Primers* **1**, 39. doi:10.1038/s43586-021-00038-x

Li, Y., Su, Y., Guo, M., Han, X., Liu, J., Vishwasrao, H. D., Li, X., Christensen, R., Sengupta, T., Moyle, M. W. et al. (2022). Incorporating the image formation process into deep learning improves network performance. *Nat. Methods* **19**, 1427-1437. doi:10.1038/s41592-022-01652-7

Lin, S., Schorpp, K., Rothenaigner, I. and Hadian, K. (2020). Image-based high-content screening in drug discovery. *Drug Discov. Today* **25**, 1348-1361. doi:10.1016/j.drudis.2020.06.001

Linkert, M., Rueden, C. T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., MacDonald, D. et al. (2010). Metadata matters: access to image data in the real world. *J. Cell Biol.* **189**, 777-782. doi:10.1083/jcb.201004104

Liu, Z., Jin, L., Chen, J., Fang, Q., Ablameyko, S., Yin, Z. and Xu, Y. (2021). A survey on applications of deep learning in microscopy image analysis. *Comput. Biol. Med.* **134**, 104523. doi:10.1016/j.compbiomed.2021.104523

Liu, X., Ye, K., van Vlijmen, H. W. T., IJzerman, A. P. and van Westen, G. J. P. (2023). DrugEx v3: scaffold-constrained drug design with graph transformer-based reinforcement learning. *J. Cheminform.* **15**, 24. doi:10.1186/s13321-023-00694-z

Lu, A. X., Kraus, O. Z., Cooper, S. and Moses, A. M. (2019). Learning unsupervised feature representations for single-cell microscopy images with paired cell inpainting. *PLoS Comput. Biol.* **15**, e1007348. doi:10.1371/journal.pcbi.1007348

Magar, R. and Farimani, A. B. (2023). Learning from mistakes: Sampling strategies to efficiently train machine learning models for material property prediction. *Comput. Mater. Sci.* **224**, 112167. doi:10.1016/j.commatsci.2023.112167

Mahecic, D., Testa, I., Griffié, J. and Manley, S. (2019). Strategies for increasing the throughput of super-resolution microscopies. *Curr. Opin. Chem. Biol.* **51**, 84-91. doi:10.1016/j.cbpa.2019.05.012

Mahecic, D., Stepp, W. L., Zhang, C., Griffié, J., Weigert, M. and Manley, S. (2022). Event-driven acquisition for content-enriched microscopy. *Nat. Methods* **19**, 1262-1267. doi:10.1038/s41592-022-01589-x

Marsh, R. J., Pfisterer, K., Bennett, P., Hirvonen, L. M., Gautel, M., Jones, G. E. and Cox, S. (2018). Artifact-free high-density localization microscopy analysis. *Nat. Methods* **15**, 689-692. doi:10.1038/s41592-018-0072-5

Masullo, L. A., Bodén, A., Pennacchietti, F., Coceano, G., Ratz, M. and Testa, I. (2018). Enhanced photon collection enables four-dimensional fluorescence nanoscopy of living systems. *Nat. Commun.* **9**, 3281. doi:10.1038/s41467-018-05799-w

McInnes, L., Healy, J. and Melville, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. *J. Open Source Softw.* **3**, 861. doi:10.21105/joss.00861

Montesinos López, O. A., Montesinos López, A. and Crossa, J. (2022). *Overfitting, Model Tuning, and Evaluation of Prediction Performance*, pp. 109-139. Cham: Springer International Publishing.

Moreno, X. C., Al-Kadhimi, S., Alvelid, J., Bodén, A. and Testa, I. (2021). ImSwitch: generalizing microscope control in python. *J. Open Source Softw.* **6**, 3394. doi:10.21105/joss.03394

Morgado, L., Gómez-de-Mariscal, E., Heil, H. S. and Henriques, R. (2024). The rise of data-driven microscopy powered by machine learning. *J. Microsc.* **295**, 85-92.

Nadkarni, P. M., Ohno-Machado, L. and Chapman, W. W. (2011). Natural language processing: an introduction. *J. Am. Med. Inform. Assoc.* **18**, 544-551. doi:10.1136/amiajnl-2011-000464

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D. and Tzovara, A. (2021). Addressing bias in big data and AI for health care: a call for open science. *Patterns* **2**, 100347. doi:10.1016/j.patter.2021.100347

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447-453. doi:10.1126/science.aax2342

Ouyang, W., Beuttenmueller, F., Gómez-de-Mariscal, E., Pape, C., Burke, T., Garcia-López-de-Haro, C., Russell, C., Moya-Sans, L., de-la-Torre-Gutiérrez, C., Schmidt, D. et al. (2022a). BioImage model zoo: a community-driven resource for accessible deep learning in bioimage analysis. *bioRxiv*, 2022.06.07.495102. doi:10.1101/2022.06.07.495102

Ouyang, W., Bai, J., Singh, M. K., Leterrier, C., Barthelemy, P., Barnett, S. F. H., Klein, T., Sauer, M., Kanchanawong, P., Bourg, N. et al. (2022b). ShareLoc — an open platform for sharing localization microscopy data. *Nat. Methods* **19**, 1331-1333. doi:10.1038/s41592-022-01659-0

Palma, A., Theis, F. J. and Lotfollahi, M. (2023). Predicting cell morphological responses to perturbations using generative modeling. *bioRxiv*, 2023.07.17.549216. doi:10.1101/2023.07.17.549216

Piccinini, F., Balassa, T., Szkalisity, A., Molnar, C., Paavolainen, L., Kujala, K., Buzas, K., Sarazova, M., Pietiainen, V., Kutay, U. et al. (2017). Advanced cell classifier: user-friendly machine-learning-based software for discovering phenotypes in high-content imaging data. *Cell Systems* **4**, 651-655.e5. doi:10.1016/j.cels.2017.05.012

Pinkard, H., Stuurman, N., Ivanov, I. E., Anthony, N. M., Ouyang, W., Li, B., Yang, B., Tsuchida, M. A., Chhun, B., Zhang, G. et al. (2021). Pycro-Manager: open-source software for customized and reproducible microscope control. *Nat. Methods* **18**, 226-228. doi:10.1038/s41592-021-01087-6

Pyzer-Knapp, E. O., Pitera, J. W., Staar, P. W. J., Takeda, S., Laino, T., Sanders, D. P., Sexton, J., Smith, J. R. and Curioni, A. (2022). Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Computational Materials* **8**, 84. doi:10.1038/s41524-022-00765-z

Rotem, O., Schwartz, T., Maor, R., Tauber, Y., Tsarfati Shapiro, M., Meseguer, M., Gilboa, D., Seidman, D. S. and Zaritsky, A. (2024). Visual interpretability of image-based classification models by generative latent space disentanglement applied to in vitro fertilization. *Nat. Commun.* **15**, 7390. doi:10.1038/s41467-024-51136-9

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* **1**, 206-215. doi:10.1038/s42256-019-0048-x

Rust, M. J., Bates, M. and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793-796. doi:10.1038/nmeth929

Sadafi, A., Bordukova, M., Makhro, A., Navab, N., Bogdanova, A. and Marr, C. (2023). RedTell: an AI tool for interpretable analysis of red blood cell morphology. *Front. Physiol.* **14**, 1058720. doi:10.3389/fphys.2023.1058720

Schmidt, U., Weigert, M., Broaddus, C. and Myers, G. (2018). Cell detection with star-convex polygons. *arXiv*, arXiv:1806.03535. doi:10.48550/arXiv.1806.03535

Schwartz, M., Israel, U., Wang, X., Laubscher, E., Yu, C., Dilip, R., Li, Q., Mari, J., Soro, J., Yu, K. et al. (2023). Scaling biological discovery at the interface of deep learning and cellular imaging. *Nat. Methods* **20**, 956-957. doi:10.1038/s41592-023-01931-x

See, P., Dutertre, C.-A., Chen, J., Günther, P., McGovern, N., Irac, S. E., Gunawan, M., Beyer, M., HÃ¤ndler, K., Duan, K. et al. (2017). Mapping the human DC lineage through the integration of high-dimensional techniques. *Science* **356**, 6342. doi:10.1126/science.aag3009

Sekar, R. B. and Periasamy, A. (2003). Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J. Cell Biol.* **160**, 629-633. doi:10.1083/jcb.200210140

Shi, Y., Tabet, J. S., Milkie, D. E., Daugird, T. A., Yang, C. Q., Ritter. A. T., Giovannucci, A. and Legant. W. R. (2024). Smart lattice light-sheet microscopy for imaging rare and complex cellular events. *Nat. Methods* **21**, 301-310. doi:10.1038/s41592-023-02126-0

Soelistyo, C. J. and Lowe, A. R. (2024). Discovering interpretable models of scientific image data with deep learning. *arXiv* arXiv:2402.03115. doi:10.48550/arXiv.2402.03115

Soelistyo, C. J., Vallardi, G., Charras, G. and Lowe, A. R. (2022). Learning biophysical determinants of cell fate with deep neural networks. *Nat. Machine Intelligence* **4**, 636-644. doi:10.1038/s42256-022-00503-6

Speiser, A., Müller, L.-R., Hoess, P., Matti, U., Obara, C. J., Legant, W. R., Kreshuk, A., Macke, J. H., Ries, J. and Turaga, S. C. (2021). Deep learning enables fast and dense single-molecule localization with high accuracy. *Nat. Methods* **18**, 1082-1090. doi:10.1038/s41592-021-01236-x

Stirling, D. R., Carpenter, A. E. and Cimini, B. A. (2021). CellProfiler Analyst 3.0: accessible data exploration and machine learning for image analysis. *Bioinformatics* **37**, 3992-3994. doi:10.1093/bioinformatics/btab634

Stringer, C., Wang, T., Michaelos, M. and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100-106. doi:10.1038/s41592-020-01018-x

Sullivan, D. P., Winsnes, C. F., Åkesson, L., Hjelmare, M., Wiking, M., Schutten, R., Campbell, L., Leifsson, H., Rhodes, S., Nordgren, A. et al. (2018). Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nat. Biotechnol.* **36**, 820-828. doi:10.1038/nbt.4225

Takan, S., Ergün, D., Yaman, S. G. and Kılınççeker, O. (2023). Bias in human data: A feedback from social sciences. *WIREs Data Mining and Knowledge Discovery* **13**, e1498. doi:10.1002/widm.1498

Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M. et al. (2017). A subcellular map of the human proteome. *Science* **356**, eaal3321. doi:10.1126/science.aal3321

Uhlmann, V., Donati, L. and Sage, D. (2022). A practical guide to supervised deep learning for bioimage analysis: challenges and good practices. *IEEE Signal Process. Mag.* **39**, 73-86. doi:10.1109/MSP.2021.3123589

Ulicna, K., Kelkar, M., Soelistyo, C. J., Charras, G. T. and Lowe, A. R. (2023). Learning dynamic image representations for self-supervised cell cycle annotation. *bioRxiv*, 10.1101/2023.05.30.542796. doi:10.1101/2023.05.30.542796

Umarov, R., Li, Y. and Arner, E. (2021). DeepCellState: an autoencoder-based framework for predicting cell type specific transcriptional states induced by drug treatment. *PLoS Comput. Biol.* **17**, e1009465. doi:10.1371/journal.pcbi.1009465

Unterauer, E. M., Shetab Boushehri, S., Jevdokimenko, K., Masullo, L. A., Ganji, M., Sograte-Idrissi, S., Kowalewski, R., Strauss, S., Reinhardt, S. C. M., Perovic, A. et al. (2024). Spatial proteomics in neurons at single-protein resolution. *Cell* **187**, 1785-1800.e16. doi:10.1016/j.cell.2024.02.045

Voulodimos, A., Doulamis, N., Doulamis, A. and Protopapadakis, E. (2018). Deep learning for computer vision: a brief review. *Comput. Intell. Neurosci.* **2018**, 7068349.

Wang, N., Chen, G.-D. and Tian, Y. (2022). Image colorization algorithm based on deep learning. *Symmetry* **14**, 2295. doi:10.3390/sym14112295

Wassie, A. T., Zhao, Y. and Boyden, E. S. (2019). Expansion microscopy: principles and uses in biological research. *Nat. Methods* **16**, 33-41. doi:10.1038/s41592-018-0219-4

Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S. et al. (2018). Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090-1097. doi:10.1038/s41592-018-0216-7

Weinstein, B., Fine, S. and Hel-Or, Y. (2019). Selective sampling for accelerating training of deep neural networks. *arXiv* arXiv.1911.07136. doi:10.48550/arXiv.1911.06996

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018. doi:10.1038/sdata.2016.18

Williams, E., Moore, J., Li, S. W., Rustici, G., Tarkowska, A., Chessel, A., Leo, S., Antal, B., Ferguson, R. K., Sarkans, U. et al. (2017). Image Data Resource: a bioimage data integration and publication platform. *Nat. Methods* **14**, 775-781. doi:10.1038/nmeth.4326

Xu, F., Jiang, L., He, W., Huang, G., Hong, Y., Tang, F., Lv, J., Lin, Y., Qin, Y., Lan, R. et al. (2021). The clinical value of explainable deep learning for diagnosing fungal keratitis using in vivo confocal microscopy images. *Front. Med.* **14**, 797616. doi:10.3389/fmed.2021.797616

Xue, Y. and Ray, N. (2017). Cell detection in microscopy images with deep convolutional neural network and compressed sensing. *arXiv* arXiv.1708.03307. doi:10.48550/arXiv.1708.03307

Yang, K. D., Damodaran, K., Venkatachalapathy, S., Soylemezoglu, A. C., Shivashankar, G. V. and Uhler, C. (2020). Predicting cell lineages using autoencoders and optimal transport. *PLoS Comput. Biol.* **16**, e1007828. doi:10.1371/journal.pcbi.1007828

Yang, Z., Wen, J., Abdulkadir, A., Cui, Y., Erus, G., Mamourian, E., Melhem, R., Srinivasan, D., Govindarajan, S. T., Chen, J. et al. (2024). Gene-SGAN: discovering disease subtypes with imaging and genetic signatures via multi-view weakly-supervised deep clustering. *Nat. Commun.* **15**, 354. doi:10.1038/s41467-023-44271-2

Zehrer, A. C., Martin-Villalba, A., Diederich, B. and Ewers, H. (2024). An open-source, high-resolution, automated fluorescence microscope. *Elife* **12**, RP89826. doi:10.7554/eLife.89826