ungeneralizable results, and pay attention to the non-ML factors (such as partnerships and user interface design) that are necessary to move towards clinical impact. ❐

Po-Hsuan Cameron Chen[1,2]*, Yun Liu[1,2] and Lily Peng[1]

[1]*Google AI Healthcare, Mountain View, CA, USA.*
[2]*These authors contributed equally: Po-Hsuan Cameron Chen, Yun Liu.*
**e-mail: cameronchen@google.com*

Published online: 18 April 2019
https://doi.org/10.1038/s41563-019-0345-0

References

1. LeCun, Y., Bengio, Y. & Hinton, G. *Nature* **521**, 436–444 (2015).
2. Gulshan, V. et al. *JAMA* **316**, 2402–2410 (2016).
3. Esteva, A. et al. *Nature* **542**, 115–118 (2017).
4. Krause, J. et al. *Ophthalmology* **125**, 1264–1272 (2018).
5. Ehteshami Bejnordi, B. et al. *JAMA* **318**, 2199–2210 (2017).
6. Poplin, R. et al. *Nat. Biomed. Eng.* **2**, 158–164 (2018).
7. Ting, D. S. W. & Wong, T. Y. *Nat. Biomed. Eng.* **2**, 140–141 (2018).
8. Xu, K. et al. Preprint at https://arxiv.org/abs/1502.03044 (2015).
9. Moher, D. et al. *BMJ* **340**, c869 (2010).
10. Japkowicz, N. & Stephen, S. *Intell. Data Anal.* **6**, 429–449 (2002).
11. Rajkomar, A. et al. *npj Digit. Med.* **1**, 18 (2018).
12. Ren, S., He, K., Girshick, R. & Sun, J. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2017).
13. Liu, Y. et al. *Arch. Pathol. Lab. Med.* https://doi.org/10.5858/arpa.2018-0147-OA (2018).
14. Steiner, D. F. et al. *Am. J. Surg. Pathol.* **42**, 1636–1646 (2018).
15. De Fauw, J. et al. *Nat. Med.* **24**, 1342–1350 (2018).
16. Sofka, M., Milletari, F., Jia, J. & Rothberg, A. in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (eds Cardoso, J. et al.) 258–266 (Springer, 2017).
17. Zoph, B., Vasudevan, V., Shlens, J. & Le, Q. V. Preprint at https://arxiv.org/abs/1707.07012 (2017).
18. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. in *IEEE Conference on Computer Vision and Pattern Recognition* 4510–4520 (IEEE, 2018).
19. Bishop, C. *Pattern Recognition and Machine Learning* (Springer, 2006).
20. Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. Preprint at https://arxiv.org/abs/1611.03530 (2016).
21. Bergstra, J. & Bengio, Y. *J. Mach. Learn. Res.* **13**, 281–305 (2012).
22. *ILSVRC* http://www.image-net.org/challenges/LSVRC/announcement-June-2-2015 (2 June 2015).
23. Alba, A. C. et al. *JAMA* **318**, 1377–1384 (2017).
24. Niculescu-Mizil, A. & Caruana, R. in *Proc. 22nd International Conference on Machine Learning* 625–632 (ACM, 2005).
25. Thabane, L. et al. *BMC Med. Res. Methodol.* **13**, 92 (2013).
26. Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G. & Thomas, R. *Indian J. Ophthalmol.* **56**, 45–50 (2008).
27. van Smeden, M., Van Calster, B. & Groenwold, R. H. H. *JAMA* **319**, 1725–1726 (2018).
28. Sayres, R. et al. *Ophthalmology* **126**, 552–564 (2018).
29. Graham, K. C. & Cvach, M. *Am. J. Crit. Care* **19**, 28–34 (2010).
30. Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N. & Folk, J. C. *npj Digit. Med.* **1**, 39 (2018).
31. Shlens, J. *Google AI Blog* https://ai.googleblog.com/2016/03/train-your-own-image-classifier-with.html (2016).

Leveraging machine vision in cell-based diagnostics to do more with less

Highly quantitative, robust, single-cell analyses can help to unravel disease heterogeneity and lead to clinical insights, particularly for complex and chronic diseases. Advances in computer vision and machine learning can empower label-free cell-based diagnostics to capture subtle disease states.

Minh Doan and Anne E. Carpenter

Current diagnostic and monitoring assays are typically performed with reagents that label specific cellular and molecular hallmarks of illness (Fig. 1a). Each of these so-called biomarkers yields a single data point: the amount of the cellular constituent that has been targeted, and they often require decades of careful study to identify and validate. Furthermore, detecting biomarkers in the clinic typically requires specific reagents, special instrumentation and/or complex laboratory manipulations, which may adversely disturb the true states of the biological targets.

At the same time, the medical community recognizes that disease heterogeneity is a major challenge obscuring accurate diagnosis and effective treatment. Precision medicine, where relatively specific treatments are tailored to each patient based on their characteristics, requires new diagnostics that can classify patients by disease subtype and by response to a given treatment regime. This is particularly relevant for complex illnesses such as autoimmune diseases and cancers, and for chronic diseases that progress over time, such as diabetes and obesity. These illnesses show substantial patient-to-patient variability in terms of symptoms, genetic underpinnings and progression. Thus, a major challenge facing medicine is to develop diagnostics that reveal this heterogeneity: a given therapeutic may only be effective when it 'matches' the right patient, or even the right subclone of cells within a given patient, and response must be monitored over time.

Fortunately, cell-based diagnostics are advancing in multiple respects. Historically, cytology was limited mainly to manual microscopic examination of a biopsy specimen prepared on a slide. However, a broader range of patient phenotypes can be detected now, including multiplexed histopathology, flow cytometry, biochemistry, genetics, proteomics, immunophenotyping and more. The prospect of obtaining a more comprehensive map of disease is on the horizon.

Here we focus on several particularly promising strategies for cell-based diagnostics that identify biomarkers based on particular elements of cell morphology rather than simply the amount of a particular target molecule of the cell. They offer single-cell resolution and fine-grained classification of samples and often can be performed label-free, preserving samples' integrity and reducing the cost of reagents and instrumentation. Although not yet in widespread clinical use, we describe the advances in devices for capturing single-cell images and in machine learning that are likely to power a new generation of cell-based diagnostics, including some that are label-free.

Practical single-cell imaging platforms

Among a wide variety of single-cell analysis assays that could be employed for cell-based diagnostics, three platforms are most commonly used to detect subtle differences between individual cells using images.

Microscopy is the most common and convenient platform for imaging cells; digital cameras can capture high-resolution images of cells in multiple colorimetric and fluorescent channels, together with transmitted and phase contrast illumination. Over the centuries, many improvements in automation, illumination, photonics, optics, cameras and labelling techniques have been applied to enable image acquisition with increasing speed, resolution and specificity. Microscopy is already in widespread use for clinical cytology, where it allows microscopic examination of biopsy specimens and the inspection of the characteristic cell or tissue

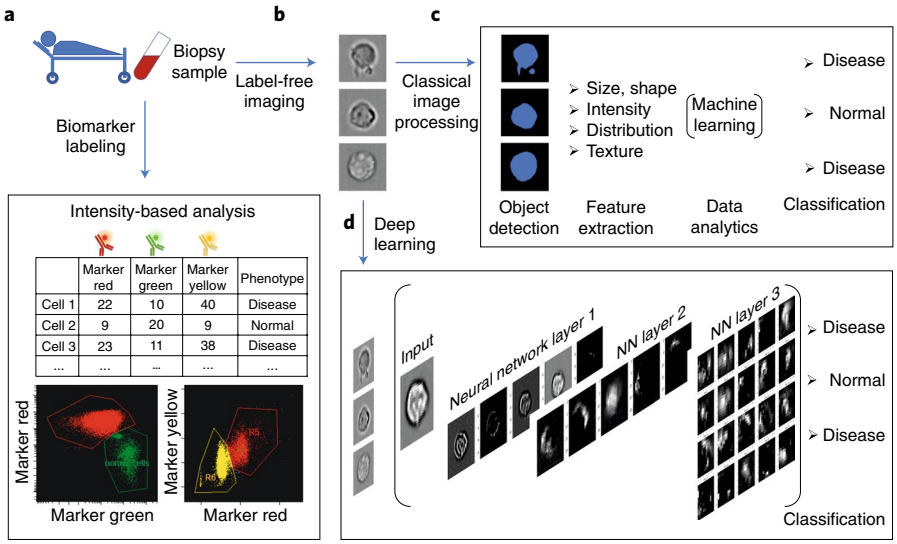

Fig. 1 | Valuable information is hidden in label-free images. High-throughput cell-based diagnostics allow samples to be analysed at single-cell resolution and across multiple channels. **a**, Conventional cell-based diagnostics often rely on specific biomarkers to identify disease status. The readouts are mainly intensity signals of the labelled targets. For multiplexed assays that involve several biological targets, such intensity-based analysis typically requires manual pairwise comparisons for the relevant markers. **b**, Recent research[9–11,13,14] indicates that label-free channels of images (such as brightfield and darkfield) can contain equivalent information, potentially replacing fluorescent markers. To accomplish this, however, it requires sophisticated extraction of information from images. **c**, In classical image processing pipelines, designed features (such as shape, intensity, texture) are helpful inputs for a machine classifier to learn the characteristic pattern of the phenotypes. However, feature engineering requires image analysis expertise and is limited in its maximum accuracy. **d**, In contrast, deep neural networks are generally more accurate and also more flexible: they identify features on their own by learning relevant patterns from a large number of examples (training dataset, not shown). One caveat is the loss of direct interpretability of the discovered features due to the hierarchy of abstract representation, as shown here on three hidden layers of a simple neural network.

hallmarks of disease. Screening cytology, such as Pap smears for abnormal cervical cells, enables life-saving early discovery of disease in the absence of clinical symptoms. As we will discuss, bringing machine learning to the analysis of microscopy/histology images offers tremendous potential for cell diagnostics with a greater ability to discern among patient subtypes.

For cells in suspension, such as in blood samples, flow-based systems are more favourable. Although flow cytometers are routinely used for clinical diagnostics, particularly in hematological disorders, they collect only a single measurement per biomarker per cell: whole-cell fluorescence intensity, using one fluorescence-conjugated marker per colour channel. Imaging flow cytometry[1] is a new generation of flow cytometers, combining the fast fluidic sampling of a flow-based system with the spatial resolution of microscopy. Equipped with optics and cameras, imaging flow cytometry captures fluorescent, brightfield and darkfield high-content spatial information in the form of images at a throughput of several hundreds to thousands of objects per second. Signals from unwanted events, such as debris, can be more easily detected and ignored than in conventional flow cytometry. Currently, few cell-based diagnostics are in clinical use that rely on imaging flow cytometry, but as we will discuss, this is likely to change: the spatial information (that is, images) that imaging flow cytometry brings may soon reduce or eliminate the need for the specific biomarkers that are required for conventional flow cytometry (Fig. 1a). The fact that each image captures a distinct single cell is also well-matched for deep learning algorithms, as we discuss below.

Microfluidic chips are an increasingly popular tool for bioimaging, compatible with both suspension and adherent cells. They are not yet used widely in clinical diagnostics, but are promising. Such chips are composed of channels smaller than one millimetre in at least one dimension, allowing small quantities of fluid to flow to stationary adherent cells and/or letting suspended cells to flow in liquids. This design can be highly modular and customizable, thus enabling parallelization and microcontrol of multiple functions in a single compact device, such as mixing, particle manipulation, imaging, tracking and other automated assays[2]. It can be used to study living cells together with their associated extracellular materials in the supernatant[3,4], which is much less feasible by microscopy or flow cytometry.

Trends in cell image analysis

A dramatic revolution in computer vision has suddenly made new technology available for image analysis that, when combined with the image-capturing devices just described, could yield a crop of novel cell diagnostics.

It is first helpful to understand existing approaches for analysing cell images for diagnostic purposes. Of course, the most widespread is the visual assessment of phenotypes by pathologists. This raises challenges: trained experts are expensive and cannot analyse enormous datasets efficiently, as in whole-slide scans of a tissue biopsy, for example. Furthermore, discrepancies among pathologists' judgment are well-documented[5], and it is possible that patterns exist in cell morphology that the human visual system is simply not equipped to perceive[6].

Image analysis software can overcome many of these challenges. In classical image processing (Fig. 1c) a researcher designs algorithms to identify each cell, its borders and any relevant subcellular compartments (for example, nuclei or other organelles) in the images so that many different kinds of measurements of these identified regions of the image can then be taken. These so-called morphological features include pixel intensities, size, shapes, textures, correlations and relationships among neighbour cells and subcellular components; these can be used directly as a diagnostic feature. Features can also be combined to detect more complex phenotypes that manifest in multiple features simultaneously using machine learning, where the algorithm learns to classify a cell as having a phenotype based on a specific combination of values for these metrics.

Deep learning[7], or deep neural networks (Fig. 1d), is a type of machine learning that has often proved to be more powerful than the carefully engineered image analysis workflows of the past. It takes the pixels of images as its input and multiple layers of the neural network examine and extract patterns in the presented pixels, while spanning a huge range of potential patterns beyond what humans might look for. With multiple layers (hence the name 'deep') and multiple

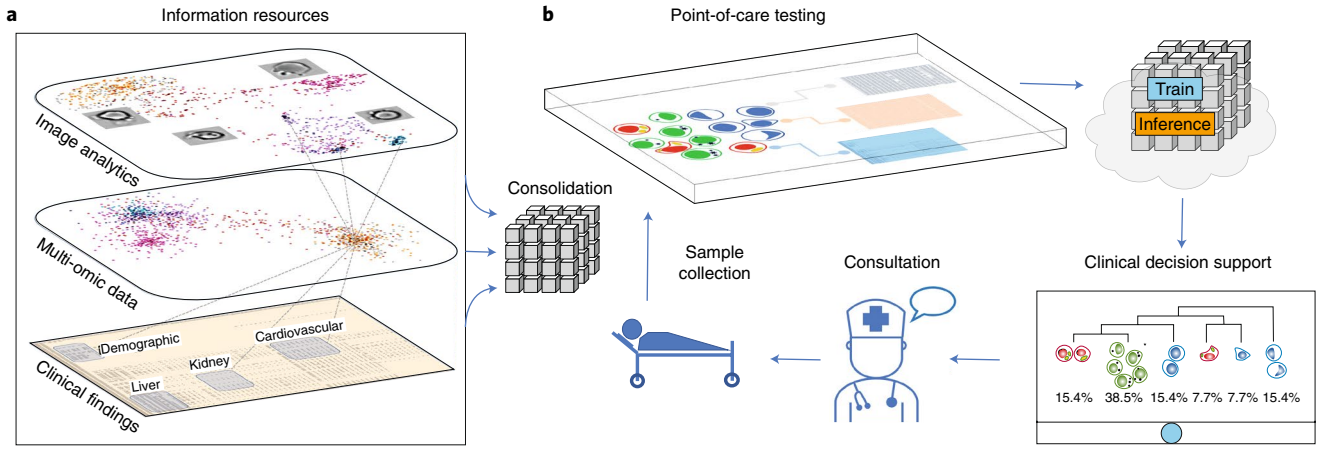

Fig. 2 | AI-powered point-of-care diagnostics usher in the era of precision medicine. Diverse and plentiful information can be efficiently integrated by machine learning to assist clinical decisions. **a**, Incorporation of cell-based bioimaging data with several other types of readouts (omic datasets, patient demographic characteristics and clinical tests) can provide a large amount of information for each patient. For example, an AI-based system can navigate the large data space on its own and consolidate information from label-free identification of a hallmark disease-bearing cell phenotype (first layer), a gene expression profile (second layer) and various entries in the clinical record (last layer), to eventually deliver a summarized report to doctors. The integrated data across layers can also be used as inputs for network analysis or machine learning algorithms to identify relevant connections (grey lines) between the disjoint pieces of information and unravel the precise diagnosis. **b**, An envisioned portable point-of-care label-free cell diagnostic chip is shown, powered by microfluidics, integrated data, machine learning, cloud computing and smart devices. Pre-trained machine learning models allow feature extraction, data integration and inference to take place in real time on brightfield cell images and can be operated on handheld devices.

nodes in each layer, the system is trained with example images that are known to have a particular phenotype — for example, an infected blood cell or a leukemic cell. In training, the weights within the network are adjusted until the system produces the right answer for the known cases — each layer of the deep network amplifies collections of pixels that are important to reach the ultimate goal while suppressing irrelevant variations[8]. Once trained, the network (or model) is ready to predict the correct answer for new images. Because of its extreme flexibility, it has the potential to uncover subtle morphological clues that might be missed by a human viewer or a classical image processing workflow.

Deep learning-powered cytology

The greater power and flexibility of deep learning makes it more likely that subtle disease-related cell phenotypes can be detected in images, potentially in the absence of specific labels for a particular target biomarker. For instance, whereas few biologists would claim to be able to quantify DNA content or distinguish G1, S and G2 cell cycle stages from images of unstained cells, machine learning recently accomplished this task using images from an imaging flow cytometer[9,10]. The system was trained using the known DNA content and cell cycle stage (G1, S, G2, prophase, metaphase, anaphase and telophase) as labelled by fluorescent dyes, and it learned to predict these values using only brightfield and darkfield images of cells. This initial study used the classical approach (as in Fig. 1c); a subsequent study used deep learning (Fig. 1d), in the form of convolutional neural networks, to achieve an even more accurate classification[11]. In another case, involving microscopy images of histopathology samples, a deep learning network was taught to predict tumour mutational status in non-small cell lung cancer without any specific labelling reagents, a capability no human could claim[12].

The scientific community has only just begun exploring the potential for deep learning to detect particular disease states and subtypes. One approach, in silico labelling, has been used in basic biology studies and might be fruitfully applied to create label-free cell diagnostics. Instead of a model being trained to detect a disease state or other phenotype for each cell, as we have just discussed, the model is trained to predict entire fluorescently stained images given only brightfield images. This has successfully worked for various cellular entities (nuclei, membranes, axons, dendrites), cell types and cell states (living and dead cells), and is feasible for both two-[13] and three-dimensional images[14]. It seems likely to extend to the prediction of particular diagnostic biomarkers, such that the amount and morphology of stained cell components can be predicted and measured, label-free.

Although many applications of machine learning for biological images aim to classify each cell as having a particular disease-related phenotype, for example cancer versus normal, sometimes more continuous metrics are needed. A trained deep learning network can extract cell features that contain sufficient information to order cells based on their morphological similarity, which in some cases corresponds to a physiologically meaningful chronological or developmental progression. This has so far been demonstrated for reconstructing the cell cycle without any a priori knowledge of the cell cycle being available to the algorithm[11]. Similar pseudotime analysis has been observed in deep mining of high-dimensional mass cytometric data to reconstruct the evolution of β-cell loss in type 1 diabetes progression[15].

Physically unmixing cell heterogeneity

One of the most exciting prospects is to not just identify disease-related phenotypes, but to also be able to physically sort cells based on them. This could allow additional, perhaps molecular, analyses to be carried out on subsets of cells, which would allow stratifying patients based on integrating multiple diagnostic criteria. It could also allow propagation and even re-introduction of subsets of cells to the patient.

Detecting and physically isolating cells based on their morphological properties has recently been demonstrated in proof-

of-principle experiments. For cells placed on a microscopic slide, an automated micromanipulation pipeline was recently developed in which image analysis and machine learning were used to guide laser capture microdissection, enabling a much higher object isolation and extraction throughput compared to manual operations[16]. In a higher-throughput approach, spatial information of cells flowing past a detector were converted into waveforms that can be directly analysed and classified by a support vector machine algorithm[17]. Using this method, the system was able to detect cells derived from human breast adenocarcinoma (MCF-7) within a mixture of peripheral blood mononuclear cells[17]. In another approach, two-dimensional images of single cells were directly analysed by deep learning to allow sorting microalgal and blood cells, where intracellular protein localization and cell–cell interaction of heterogeneous populations were detected and physically sorted[18]. When these instruments become widely available, morphology-activated label-free cell sorting of minimally perturbed living cells could enable pre-stratifying a mixture of heterogeneous cells into more homogeneous subpopulations, allowing researchers to gain a higher level of purity and enrichment of readouts in downstream analysis. Ultimately, cells sorted by morphology in label-free fashion might be cultured and reintroduced into patients, as seen in adoptive cellular immunotherapy, where clinicians collect and use patients' own immune cells to treat their cancer[19].

Intelligent label-free diagnostics

As we have seen, using advanced image analysis techniques such as deep learning can potentially extract more information out of unlabelled cell morphology. Reducing or eliminating the need for reagents, sample manipulation mechanics and energy sources can substantially improve the affordability, portability and efficiency of diagnostic tools[20]. Patient samples often need to be sent to a remote laboratory for testing, which increases costs and delays treatment. This is particularly problematic if time or resources are limiting, such as in developing countries, in areas of warfare, during infectious disease outbreaks and natural disasters, and on-board water-, air- and spacecraft. Going label-free can reduce the necessary reagents and instrumentation and thereby reduce laboratory expense, effort and error.

Moreover, eliminating labels leaves fluorescent channels available for other purposes and allows integrating other useful assays with sensitive chemistry, such as single-cell omics[21,22] for diagnosing and monitoring diseases. Although individual methodologies such as bioimaging, single-cell genomics and proteomics already offer exquisite data for deconvoluting cell heterogeneity, it is rarely possible to link these data. A lab-on-a-chip that integrates cell-based imaging, sorting and molecular omic data thus presents an opportunity to connect cell-to-cell (for instance, a given cell's image to its omic profile), or cluster-to-cluster (for instance, a cell phenotype to its genotype signature) to help maximize the ability to detect phenotype (Fig. 2a). As a result, a pathological cell as well as its extracellular components (for example, exocytosed vesicles and materials) could be described by a vast set of descriptive parameters, aiding diagnosis, particularly for rare cells. This myriad of data is a rich resource for machine learning algorithms, and more broadly for artificial intelligence (AI), to unravel disease heterogeneity and enable precision medicine.

Although computationally demanding during training, once trained, an algorithm can perform inference (that is, classification or prediction of a phenotype) with much less workload. More accessible and more affordable parallelism, such as high-performance processors and cloud computing also continue to remove barriers in hardware limitations, enabling even faster readouts. For example, a pre-trained classifier could readily perform prediction or diagnosis on new unseen patient samples and return the results to a portable device via a cloud computing server (Fig. 2b), providing information to assist clinicians in decision making.

Label-free imaging may come in different forms. In addition to the devices described, many others exist. For instance, stimulated Raman scattering as a contrast mechanism has been successfully leveraged in label-free imaging different lipid subtypes in living brain and skin tissues[23]. The technique was later bundled into a portable platform to perform histopathology examination in fresh patient specimens taken during a neurological surgery[24]. In over a hundred intraoperative cases, the technique was proven useful in not only delivering fair contrast for brain tumour detection in unprocessed images, but also giving rise to virtual staining comparable to hematoxylin and eosin. This revealed essential diagnostic cellular and histologic architectures, permitting differentiation of non-lesional and lesional tissues, facilitating human interpretation as well as automated detection by machine vision. Label-free imaging could also be accomplished via an optofluidic time-stretch microscopy system, which has seen great progress in engineering and design to overcome the trade-off between sensitivity and speed[25]. Here, the cell's spatial information is encoded in the spectrum of a time-stretched laser pulse in sub-nanoseconds. The weak signals (low number of photons) received during such short capture time and the drop in power resulting from the time stretch are compensated by Raman amplification. This technique reveals biophysical attributes of the cells including protein concentration, optical loss, intensity and phase map of the cells and other morphological features without the need for any staining. Deep learning applications based on time-stretched imaging features were demonstrated in classifying hybridoma T-lymphocytes and colon cancer epithelial cells, as well as algal cells of different lipid contents[26].

Limitations of machine learning

For machine learning, including deep learning, to be sufficiently reliable for clinical use, certain limitations need to be carefully addressed. First, there are many situations in which a machine learning model can give excellent accuracy in testing but perform poorly in the real world. Typically, this failure to generalize is due to overfitting, where the machine learning model gives great accuracy, but only for images very similar to those used in training and testing, such as from a single hospital, a single instrument or even from a certain experimental batch. This problem can be solved by gathering a wide variety of images that span the range of variation expected in clinical settings, and by ensuring that the classifier is trained on a set of those images that is completely distinct from those used for testing — for example, from different hospitals, instruments and technicians. The problem can be mitigated in a complementary way by interpreting how the machine is making diagnostic decisions for the images. Unfortunately, this is often challenging for machine learning techniques and is even more so for deep learning models, which yield a 'black box' solution whose decision-making process is opaque to clinicians. This is a long-standing challenge[27], but progress has been made in recent years[28,29] to devise methods to peek into the black box and be reassured that the decision-making is grounded in morphological features that are biologically sensible.

Second, it will be critical for clinicians to consider performance metrics of an

algorithm in the real-world context. Often in pathology detection and classification studies, machine performance is compared to that of physicians as if the goal is a replacement. In practice, it is preferable to consider outputs from cell diagnostics as guidance for the clinician's final judgment rather than treating the machine learning model as a stand-alone diagnostic decision-maker, especially for diseases with a low prevalence.

Last but not least, there are certain inherent inequities in access to medical care and technical advances, such that not all demographics are equally represented in datasets. An imbalance of data entries (for example, across ethnicities, genders, socioeconomic status and so on) and the disproportionate use of computer-aided services may exacerbate biases, potentially leading to disastrous errors made by an algorithm[30].

The future

Solving these issues will be difficult but rewarding. From bedside biopsy to rapid reports, image-based machine learning-powered tests will fuel a new age of precision medicine in multiple ways: (i) for patients, less centralized tests allow on-field screening, accelerate diagnoses and identification of epidemics, facilitate preventive care, reduce costs, and can be lifesaving if a hospital is not immediately available; (ii) for doctors, these information-rich screens deliver instant rationales for more rigorous diagnosis and treatment choices, and provide closer monitoring of disease progression, therapy response, patient susceptibility and individual drug tolerance; (iii) for researchers, collected data provide cellular and molecular evidence to characterize the role of specific biomarkers, uncover hidden biological pathways, perform large-scale disease modelling and pave ways for novel and more efficacious therapies; and (iv) for clinics, the ability to quickly stratify patients can help predict key outcomes and thus use resources more efficiently, including risk estimation, predicting relapse possibility, defining criteria for discharge/readmission, forecasting mortality/prognosis, and signalling potential shock/crisis episodes. As major pillars of bioengineering innovation, imaging and machine learning provide an excellent avenue to progressively transform modern healthcare. ❐

Minh Doan* and Anne E. Carpenter
Imaging Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
**e-mail: minhdoan@broadinstitute.org*

Published online: 18 April 2019
https://doi.org/10.1038/s41563-019-0339-y

References

1. Stavrakis, S., Holzner, G., Choo, J. & deMello, A. *Curr. Opin. Biotechnol.* **55**, 36–43 (2018).
2. Shields, C. W. IV, Reyes, C. D. & López, G. P. *Lab Chip* **15**, 1230–1249 (2015).
3. Gholizadeh, S. et al. *Biosens. Bioelectron.* **91**, 588–605 (2017).
4. Reátegui, E. et al. *Nat. Commun.* **9**, 175 (2018).
5. Djuric, U., Zadeh, G., Aldape, K. & Diamandis, P. *npj Precis Oncol.* **1**, 22 (2017).
6. Murphy, R. F. *Nat. Chem. Biol.* **7**, 327–330 (2011).
7. Ching, T. et al. *J. R. Soc. Interface* **15**, 20170387 (2018).
8. LeCun, Y., Bengio, Y. & Hinton, G. *Nature* **521**, 436–444 (2015).
9. Blasi, T. et al. *Nat. Commun.* **7**, 10256 (2016).
10. Hennig, H. et al. *Methods* **112**, 201–210 (2017).
11. Eulenberg, P. et al. *Nat. Commun.* **8**, 463 (2017).
12. Coudray, N. et al. *Nat. Med.* **24**, 1559–1567 (2018).
13. Christiansen, E. M. et al. *Cell* **173**, 792–803 (2018).
14. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. *Nat. Methods* **15**, 917–920 (2018).
15. Damond, N. et al. *Cell Metab.* **29**, 755–768 (2019).
16. Brasko, C. et al. *Nat. Commun.* **9**, 226 (2018).
17. Ota, S. et al. *Science* **360**, 1246–1251 (2018).
18. Nitta, N. et al. *Cell* **175**, 266–276 (2018).
19. June, C. H., O'Connor, R. S., Kawalekar, O. U., Ghassemi, S. & Milone, M. C. *Science* **359**, 1361–1365 (2018).
20. Snodgrass, R. et al. *Nat. Biom. Eng.* **2**, 657–665 (2018).
21. Lan, F., Demaree, B., Ahmed, N. & Abate, A. R. *Nat. Biotechnol.* **35**, 640–646 (2017).
22. Ma, S., Murphy, T. W. & Lu, C. *Biomicrofluidics* **11**, 021501 (2017).
23. Freudiger, C. W. et al. *Science* **322**, 1857–1861 (2008).
24. Orringer, D. A. et al. *Nat. Biomed. Eng.* **1**, 0027 (2017).
25. Lei, C. et al. *Nat. Protoc.* **13**, 1603–1631 (2018).
26. Chen, C. L. et al. *Sci. Rep.* **6**, 21471 (2016).
27. Castelvecchi, D. *Nature* **538**, 20–23 (2016).
28. Bau, D., Zhou, B., Khosla, A., Oliva, A. & Torralba, A. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 3319–3327 (IEEE, 2017).
29. Olah, C., Mordvintsev, A. & Schubert, L. *Distill* **2**, e7 (2017).
30. Topol, E. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (Basic Books, 2019).

Opportunities and challenges using artificial intelligence in ADME/Tox

At the recent Artificial Intelligence Applications in Biopharma Summit in Boston, USA, a panel of scientists from industry who work at the interface of machine learning and pharma discussed the diverging opinions on the past, present and future role of AI for ADME/Tox in drug discovery and development.

Barun Bhhatarai, W. Patrick Walters, Cornelis E. C. A. Hop, Guido Lanza and Sean Ekins

The term artificial intelligence (AI) can have several meanings. In this context, we will be referring specifically to applications of machine learning (ML) using various methods (Table 1) in drug discovery, and more specifically to the prediction of absorption, distribution, metabolism, excretion and toxicology (ADME/Tox) properties (Table 2).

We should differentiate this approach from other computational approaches that are widely used, such as physiologically based pharmacokinetic (PBPK) and pharmacokinetic pharmacodynamic/quantitative systems pharmacology (PKPD/QSP) modelling; however, if the decision-making process is automated these could be applied in AI frameworks. In silico ADME/Tox models have made considerable progress over the past ~40 years (Table 3), and we are now squarely in the era of ML with many of the models serving their intended purpose in the industry and complementing experimental methods. The big question is whether the latest generation of computational tools such as deep neural networks (DNNs, also referred to as deep learning) will further improve the quality of the models and decision making in lead optimization.

A typical large pharma drug discovery team is generating data for between 20 and 50 different assays including biochemical,