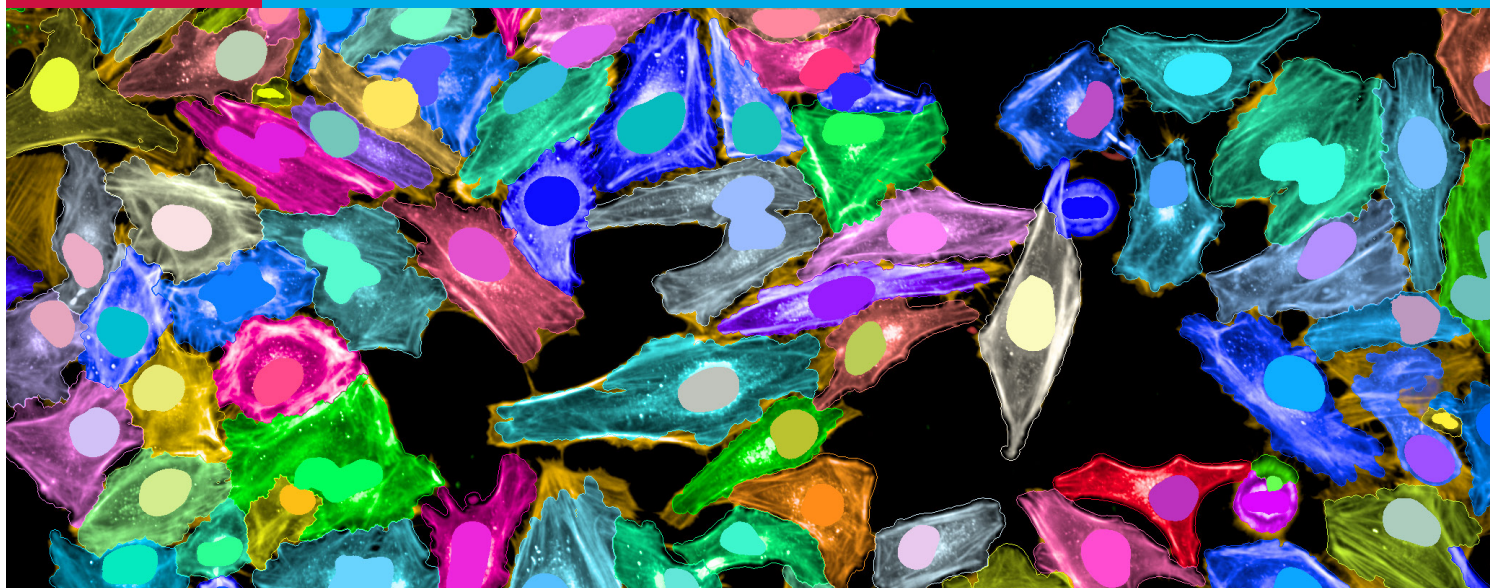




Applications of AI, ML and DL in Cellular Imaging for Improved Drug Discovery Productivity

Cellular assays and imaging technologies have substantially increased in throughput capabilities in recent years. This is good news for researchers who need to generate robust data because it enables them to obtain more data in less time while maintaining high data quality. The challenge with such large datasets is how to evaluate them in a timely manner. These advances and challenges are particularly important for drug discovery for which actionable data and timely results are paramount.

An answer to this challenge is being found in the application of artificial intelligence-based technologies to imaging analysis. In particular, machine learning and deep learning applications are being developed and used broadly for drug discovery workflows. Here we consider the ways in which artificial intelligence is being used in cellular image analysis, beginning with what these data analytics terms mean.

Artificial Intelligence, Machine Learning, and Deep Learning

The initial concept of artificial intelligence (AI) is credited to British mathematician Alan Turing who, in the 1930s, began asking if it was possible to develop a machine that could “think.” The term “artificial intelligence” was first coined in the mid-1950s by John

McCarthy, an American mathematician and computer scientist. McCarthy was referring to the development of machines that could think autonomously and “do things that, when done by people, are considered to involve intelligence.”¹ That same basic supposition still applies today: Artificial Intelligence is known as the science of developing computers that can use data to make human-like decisions and take action.

Machine learning (ML) is a branch of AI focused on how a machine learns from data. ML data analysis automates model building through iterative analytical computations. It does so by using human-designed algorithms and training data to evaluate new information, make decisions, and even adjust its own models. Machine learning is being used in self-driving cars, fraud detection, and shopping recommendations, just to name a few.²

Deep learning (DL) is a type of ML that teaches the computer to perform human-like tasks such as recognizing speech and images or predicting outcomes. It does so using multiple algorithms that may interpret the data in different ways. The multiple algorithms are interconnected to form an artificial neural network (ANN) that can evaluate immense datasets and look for deep patterns. DL is one of the most promising areas of artificial intelligence for cellular image analysis.²

Neural Networks

- ▶ A neural network is a software architecture with interconnected nodes that work similarly to neurons in the human brain. Neural networks use algorithms to recognize patterns and correlations in datasets, cluster and classify the data, and continuously learn and improve.³

History of ML and DL in Cellular Imaging and Drug Discovery

The use of ML and DL in cellular imaging and drug discovery began in earnest in the 1990s. One of the early applications of DL in biomedical imaging was using a neural network to build a model for the detection of cancerous nodules in lung X-rays. That ANN contained two layers, compared to today's neural networks that can have more than 100 layers. Since then, the application of ML and DL in imaging and drug discovery has steadily expanded. The last decade in particular has been prolific for the development of new ML and DL applications for different drug discovery activities, such as:

- *De novo* molecule design
- Predicting the 3-dimensional shape of a protein
- Predicting a molecule's structure-activity relationship
- Predicting drug-target interactions
- Predicting reactions in retrosynthetic analysis
- Benchmark studies in preclinical and clinical trials
- Drug repurposing studies

Current Applications and Challenges

High throughput workflow advances enable the generation of larger and larger datasets. Similar ongoing advances in computing power and the ML/DL it supports enable faster and deeper evaluation of large datasets. These technologies are being used for high-dimensional, image-based profiling to improve the drug discovery process.

The drug discovery process has a few perennial challenges that ML/DL is helping address. Here we look at four of the most commonly cited challenges along with examples of how ML/DL is helping researchers overcome them.

Reduce the Time and Costs for Drug Development

It is commonly known that drug discovery takes an extraordinary amount of time and money to complete. The use of ML/DL in high content screening of thousands of potential drug candidates reduces the time, personnel, and costs needed to complete the screening and data analysis.

One emerging example of this is cell painting which uses up to six different dyes to stain cellular compartments.⁴ Cellular imaging is used to extract different features from the "painted" cell. Hundreds, or even thousands, of such features are extracted from individual images or cells to produce a valid dataset. If completed manually, this type of image analysis would require days to weeks of multiple scientists' time (and the associated costs).

ML/DL completes image analysis and data extraction in minutes, sometimes even seconds, for a field or well with one scientist overseeing the process. Using a technology like PerkinElmer's Harmony® software, thousands of parameters can be derived, the information classified into disease phenotypes by technology such as PerkinElmer's Signals™ Screening analysis and visualization software, and all within a significantly shorter timeframe for even the largest of datasets. This use of ML/DL enables the rapid identification of drugs with high efficacy that should continue to next steps, and concurrently identifies those that should be dropped from further consideration. In addition to saving time and costs for assay completion, it also saves future time and costs that would be incurred should ineffective drugs continue to the next steps of evaluation.

Gain Novel Insights Early

Insights gained from earlier studies are used to design downstream studies. Therefore, the robustness and accuracy of earlier studies is crucial for subsequent steps. DL is being used to uncover deep insights from cellular imaging that are not obtainable by traditional image analysis.

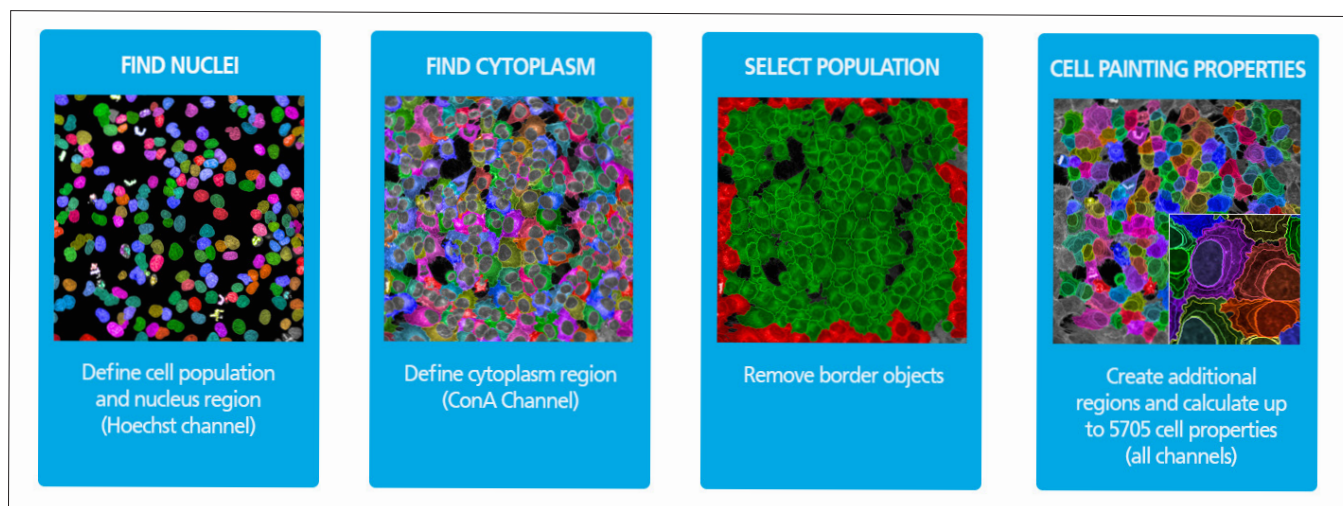


Figure 1. The image analysis sequence for cell painting in the Harmony software can extract up to 5705 properties.

For instance, complex image analysis tasks such as three-dimensional segmentation and quantification are needed to fully evaluate complex cellular and tissue models such as spheroids and organoids. That is challenging to do with traditional image analysis. ML/DL methods with their deep neural networks are able to use gathered data to automatically optimize segmentation and classification in such complex models.

A recent study explored the application of DL in image-based profiling as a means of identifying unexpected phenotypes. The researchers developed a DL framework for phenotype identification and classification that did not rely on previous knowledge of expected phenotypes. They applied the framework to several large datasets of nuclear and mitotic cell morphologies. The application was able to identify and segment several rare phenotypes that would not be identified by conventional image analysis or even ML that uses training data for expected phenotypes.⁵

Increase Productivity

As discussed earlier, ML/DL can reduce the amount of time required to conduct assays and data analysis. The time savings also contributes to increased productivity, for example, by freeing up time for more assays and allowing scientists to focus on other important aspects of their research.

One recent study looked at using DL to eliminate the need for traditional feature selection and reduction in image analysis. The researchers developed a multi-scale convolutional deep neural network (M-CNN) able to classify cellular images without the time-consuming steps of loading existing data and manual customization. In one step, the application classified cellular images into phenotypes using the images' pixel intensity values. The researchers evaluated the performance of the classification using eight benchmark datasets. The results revealed a greater classification accuracy than other standard procedures and even other CNN architectures. They were also able to use the probability outputs to quantitatively describe the phenotypes.⁶

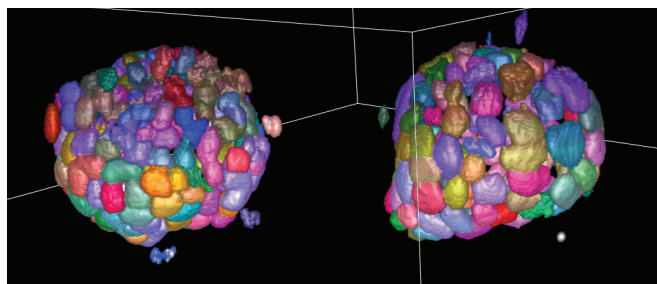


Figure 2. 3D view of MDCK cysts with 3D segmentation of nuclei analyzed in the Harmony® high-content analysis software.

Reduce Human Bias and Error

Conventional imaging analysis is prone to human bias and error, especially in the segmentation of complex images. Researchers are developing DL approaches that automate segmentation or eliminate the need for it entirely, thus minimizing bias and error in data analysis.

Recently, researchers at Charles River Labs looked at the ability of ML to accurately identify and segment ischemic stroke lesions. They used a manually delineated imaging

dataset with a DL convolutional neural network to detect lesion borders and other features. The researchers found the DL application minimized human errors and bias, and considered such an application important in increasing the consistency and quality of the data generated.⁷

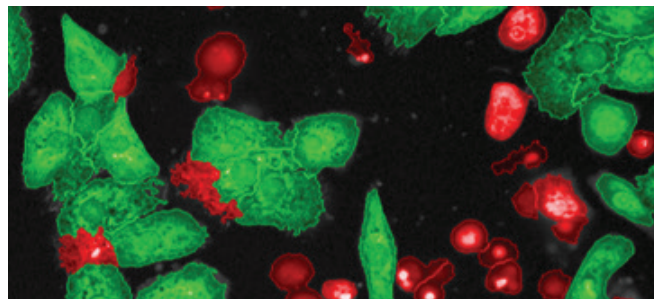


Figure 3. The PhenoLOGIC™ software plug-in for Harmony uses machine learning to enable fast and robust segmentation, as shown here with the classification of primary hepatocytes (green = healthy, red = dying).

Future Outlook

ML/DL is being used in an increasing number of drug discovery laboratories in an effort to overcome the challenges inherent in cellular imaging analysis. As they use new ML/DL approaches, they are identifying even more potential ways it can help them in their efforts, such as:

- Improving image quality by clarifying image borders and removing distortions from external causes such as dust particles
- Further improving image analysis time, computing power, and data storage
- Using precise data reduction to eliminate non-target data
- Successfully integrating and analyzing very large datasets (terabytes of data)
- Enabling label-free imaging techniques such as brightfield imaging
- Enabling assays based on more biologically relevant cell types such primary cells and stem cells

There are challenges that must be met to continue this growth of AI-enabled innovation in drug discovery. For instance, data scientists and organizations must find innovative ways to increase access to, and sharing of, databases to provide the enormous amount of data needed for deep learning. There must be an increase in the number of skilled data scientists and software engineers to design and operate AI-based platforms. Strategic and educational dialogue must continue in order to help pharmaceutical company management, scientists, and engineers overcome apprehension and skepticism about the value and potential of artificial intelligence in drug development.⁸

Despite these challenges, many pharmaceutical companies have indicated their intention to increase staff over the next few years to enable the implementation of more AI-based technologies.⁹ This heartening news, along with the achievements of the last several years, are reflected in the pharmaceutical industry market indicators. Revenue of more than \$2 billion USD is expected by 2022 thanks to AI-based solutions in the pharmaceutical sector.¹⁰

Many companies, universities, and foundations are focused on continuous improvement and advancement of ML/DL applications in biomedical fields. The Broad Institute in Cambridge, Massachusetts, is exploring a range of ML/DL applications including 3D organ modeling, antibiotic discovery, and others.

The University of Tartu in Estonia is developing ML/DL applications in computational biology and neuroscience. PerkinElmer imaging and technology scientists are collaborating with their colleagues at the University to explore and advance ML/DL applications for brightfield (label-free) assays to better support live cell imaging.

Technology providers like PerkinElmer are using ML/DL to provide advanced data analysis capabilities with their imaging software. PerkinElmer scientists designed and implemented ML/DL methods in the software to automate cellular imaging analysis tasks and optimize the segmentation and classification of image data. This is especially important for complex yet more biologically-relevant models such as organoids and spheroids for which it can be difficult to obtain accurate and thorough data on morphology, texture, and phenotypic classification.

Thanks to scientists and software such as these, drug discovery researchers are making strides toward more efficient and timely workflows by pairing high-throughput imaging technology and advanced ML/DL applications for data analysis. This is great news for healthcare professionals who are anxiously awaiting better therapies for their patients facing devastating diseases and disorders. Improved drug discovery workflows are helping to make better therapies available faster.

PerkinElmer Solutions

- **Harmony® high-content analysis software** allows you to easily quantify complex cellular phenotypes for a range of applications such as live cell imaging, 3D cell models, rare cell phenotypes, and routine assays.
- **Columbus™ software for image data storage and analysis** is an instrument-agnostic image analysis and management platform that provides universal high-volume image data storage and analysis and brings access to images from a wide range of sources including all major high content screening instruments.
- **TIBCO Spotfire® based Signals™ Screening software for analysis and visualization** is an assay data analysis and management platform that enables scientists to transform raw instrument data into actionable compound ID lists in just a few clicks.

References

1. Brookings Institution. 2018. What is Artificial Intelligence? <https://www.brookings.edu/research/what-is-artificial-intelligence/>
2. SAS. Artificial Intelligence: What it is and why it matters. https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html
3. SAS. Neural Networks: What they are and why they matter. https://www.sas.com/en_us/insights/analytics/neural-networks.html
4. Bray, M.A. et al. 2016. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc. 11(9):1757-74. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5223290/>
5. Sommer, C. et al. 2017. A deep learning and novelty detection framework for rapid phenotyping in high-content screening. Molecular Biology of the Cell, Vol. 28, No. 23. <https://www.molbiolcell.org/doi/full/10.1091/mbc.e17-05-0333>
6. Godinez, W. et al. 2017. A multi-scale convolutional neural network for phenotyping high-content cellular images. Bioinformatics, Volume 33, Issue 13, Pages 2010–2019. <https://academic.oup.com/bioinformatics/article/33/13/2010/2997285>
7. Biocompare. 2019. AI Powers Advances in Preclinical Imaging Analysis. <https://www.biocompare.com/Editorial-Articles/517686-AI-Powers-Advances-in-Preclinical-Imaging-Analysis/>
8. Debleena, Paul, et al. 2020. Artificial intelligence in drug discovery and development. Drug Discovery Today, October 21. <https://doi.org/10.1016/j.drudis.2020.10.010>
9. Lamberti, M.J. et al. 2019. A Study on the Application and Use of Artificial Intelligence to Support Drug Development. Clinical Therapeutics, Volume 41, Issue 8, Pages 1414–1426. <https://www.sciencedirect.com/science/article/abs/pii/S0149291819302942>
10. Research and Markets. 2019. Growth Insight - Role of AI in the Pharmaceutical Industry, Global, 2018-2022. <https://www.researchandmarkets.com/reports/4846380/growth-insight-role-of-ai-in-the-pharmaceutical>

PerkinElmer, Inc.
940 Winter Street
Waltham, MA 02451 USA
P: (800) 762-4000 or
(+1) 203-925-4602
www.perkinelmer.com



For a complete listing of our global offices, visit www.perkinelmer.com/ContactUs

Copyright © 2021, PerkinElmer, Inc. All rights reserved. PerkinElmer® is a registered trademark of PerkinElmer, Inc. All other trademarks are the property of their respective owners.