

mSphere of Influence: the Rise of Artificial Intelligence in Infection Biology

Artur Yakimovich^a

^aMRC-Laboratory for Molecular Cell Biology, University College London, London, United Kingdom

ABSTRACT Artur Yakimovich works in the field of computational virology and applies machine learning algorithms to study host-pathogen interactions. In this mSphere of Influence article, he reflects on two papers “Holographic Deep Learning for Rapid Optical Screening of Anthrax Spores” by Jo et al. (Y. Jo, S. Park, J. Jung, J. Yoon, et al., *Sci Adv* 3:e1700606, 2017, <https://doi.org/10.1126/sciadv.1700606>) and “Bacterial Colony Counting with Convolutional Neural Networks in Digital Microbiology Imaging” by Ferrari and colleagues (A. Ferrari, S. Lombardi, and A. Signoroni, *Pattern Recognition* 61:629–640, 2017, <https://doi.org/10.1016/j.patcog.2016.07.016>). Here he discusses how these papers made an impact on him by showcasing that artificial intelligence algorithms can be equally applicable to both classical infection biology techniques and cutting-edge label-free imaging of pathogens.

KEYWORDS anthrax, artificial intelligence, bioimage analysis, computer vision, convolutional neural networks, deep learning, label-free imaging, machine learning

Modern day microscopy creates an incredible opportunity for infection biology. Pathogens that required complex visualization techniques just a decade ago can now be readily visualized using benchtop equipment (1–4). Notably, label-free techniques may in some cases lower, or even eliminate, the need for pathogen-specific reagents (5–7). Inevitably, the availability and abundance of pathogen and host-pathogen microscopy data now necessitate unbiased quantification and highly parallel automation to approach the depth of information contained in these images. Recent advances in machine learning (ML) and deep learning (DL) offer the possibility of novel image analysis, quantification, and classification methods (8). Dubbed artificial intelligence (AI) algorithms, according to the name of the computer science discipline, ML/DL allows for unprecedented ways of detection through pattern recognition. Two articles that highlight the power of AI for infection biology, “Holographic Deep Learning for Rapid Optical Screening of Anthrax Spores” (9) and “Bacterial Colony Counting with Convolutional Neural Networks in Digital Microbiology Imaging” (10) applied novel algorithms and label-free holographic image acquisition to bacterial spores and colonies, respectively. These two early examples showcased the flexibility of AI to tackle image analysis problems across divergent size scales.

For optical screening of anthrax spores, Jo and colleagues (9) demonstrate how a DL algorithm learns subtle patterns in order to accurately distinguish spores from *Bacillus anthracis* and four related species. Notably, images used in the study were obtained by label-free holographic microscopy, which does not rely on any biochemical or immunohistochemical labeling. While this imaging technique posed an additional challenge for the DL algorithm, it served to extend the applicability of the methodology. The trained model, called HoloConvNet, is comprised of an artificial neural network (ANN) accompanied by the list of weights and biases (i.e., parameters) obtained in the training. The HoloConvNet architecture, a so-called convolutional ANN (CNN), is inspired by the mammal visual cortex. CNNs learn a magnitude of subtle image features

Citation Yakimovich A. 2019. mSphere of Influence: the rise of artificial intelligence in infection biology. *mSphere* 4:e00315-19. <https://doi.org/10.1128/mSphere.00315-19>.

Copyright © 2019 Yakimovich. This is an open-access article distributed under the terms of the *Creative Commons Attribution 4.0 International license*.

Address correspondence to a.yakimovich@ucl.ac.uk.

The views expressed in this article do not necessarily reflect the views of the journal or of ASM.

 The Rise of Artificial Intelligence in Infection Biology: a short commentary on how researchers around the world revolutionise the field of Infection Biology. @ayakimovich

Published 26 June 2019

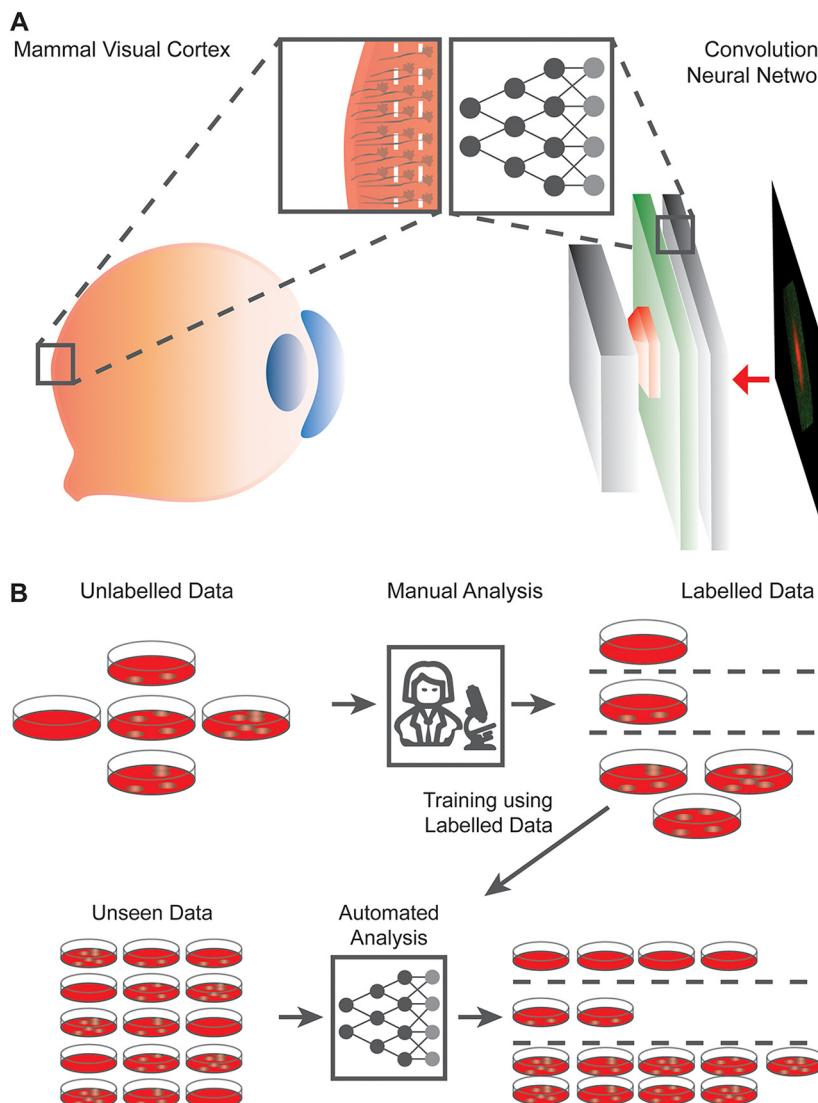


FIG 1 Simplistic illustration of an artificial intelligence algorithm and its applications in the infection biology data analysis. (A) Illustration of the Convolutional Neural Network (CNN) (right-hand side) analogy to mammalian visual cortex (left-hand side). Layers of biological neurons are known to provide an increasing amount of resolution, easing the processing of the visual information. Similarly, CNN relies on digital image information transformed through multiple convolution operations using layers of artificial neurons (depicted as colored boxes). Such processing is creating higher dimensional representations of a complex image. These representations can then be used in CNN training e.g., to distinguish between different class of images with higher precision. (B) Principle scheme of artificial intelligence algorithms application to infection biology problems improving analysis scalability and precision. This requires obtaining a digital input (e.g., image, sound recording etc.). Next such input is annotated by the lab's trained specialists to obtain the desired output. Finally, an artificial neural network algorithm is devised to map the input and output with high accuracy.

through their transformed representations (Fig. 1A). Using this methodology, the authors could distinguish *B. anthracis* from *Bacillus thuringiensis*, *B. cereus*, *B. atrophaeus*, and *B. subtilis* with greater than 96% accuracy. The holographic microscopy used in this study proved to be critical. Not only did it reveal the refractive index distribution of the various spores to be the characteristic signature on which the HoloConvoNet based its high detection precision, but by avoiding the need for specific labeling reagents, the trained DL model can now be used in various laboratories across the world.

Changing scales, Ferrari et al. (10) developed CNN-assisted bacterial colony counting as opposed to distinguishing specific pathogens. Importantly, the authors demonstrate improved detection and separation of bacterial colonies using conventional photo-

graphs of laboratory samples (10). For this, a specialist-annotated image segmentation data set was created using images of bacterial colonies from 28,500 blood agar plates. While the size of the data set used in the study illustrates the “data-hungriness” problem associated with CNNs, at 92% accuracy, the model developed is unprecedented. Remarkably, the CNN devised here is considered relatively shallow by today’s standards, suggesting that the accuracy may be further improved using the same training data set.

These papers highlight the fact that deep ANNs are great approximators for virtually any nonlinear classification problem. What most influenced me was the potential of AI to transform data analysis and drive new discovery. As the vast majority of such problems in biology are currently solved manually, a typical strategy of leveraging new AI opportunities in the lab involves replacing tedious manual tasks performed by a specialist with automated CNN workflow (Fig. 1B). For instance, the work of Jo and colleagues (9) was focused on single pathogens as data points, rather than single images, allowing for data set multiplexing. We have employed a comparable pathogen-centric view, while building our classifier for *Toxoplasma gondii* host-pathogen interactions (11).

A vast amount of challenges currently hinder the progress of image analysis in the field of infection biology. In this sense, infection biology represents an incredibly fertile ground for a domain-specific application of AI. Given that a high-resolution image capturing a single host cell can readily contain 10- to 1,000-fold more pathogens, it is easy to envision pathogen classification as a large data set ML problem where AI algorithms can demonstrate their edge. A multitude of recent biomedical publications following this pattern suggest the immense potential of AI to improve precision, ease the burden of repetitive manual quantification, and accelerate research discovery. Given the scale, even conservative improvements in routine tasks may well signal the dawn of a new era for infection biology, the era of machines.

ACKNOWLEDGMENTS

I thank Jason Mercer for comments.

A.Y. is supported by core funding to the MRC Laboratory for Molecular Cell Biology at the University College London, European Research Council (649101-UbiProPox), UK Medical Research Council (MC_UU12018/7) awarded to Jason Mercer.

REFERENCES

- Wang I, Suomalainen M, Andriasyan V, Kilcher S, Mercer J, Neef A, Luedtke NW, Greber UF. 2013. Tracking viral genomes in host cells at single-molecule resolution. *Cell Host Microbe* 14:468–480. <https://doi.org/10.1016/j.chom.2013.09.004>.
- Gray RD, Beerli C, Pereira PM, Scherer KM, Samolej J, Bleck CKE, Mercer J, Henriques R. 2016. VirusMapper: open-source nanoscale mapping of viral architecture through super-resolution microscopy. *Sci Rep* 6:29132. <https://doi.org/10.1038/srep29132>.
- Beerli C, Yakimovich A, Kilcher S, Reynoso GV, Fläschner G, Müller DJ, Hickman HD, Mercer J. 2019. Vaccinia virus hijacks EGFR signalling to enhance virus spread through rapid and directed infected cell motility. *Nat Microbiol* 4:216–225. <https://doi.org/10.1038/s41564-018-0288-2>.
- Wang I-H, Burckhardt CJ, Yakimovich A, Morf MK, Greber UF. 2017. The nuclear export factor CRM1 controls juxta-nuclear microtubule-dependent virus transport. *J Cell Sci* 130:2185–2195. <https://doi.org/10.1242/jcs.203794>.
- Cotte Y, Toy MF, Pavillon N, Depeursinge C. 2010. Microscopy image resolution improvement by deconvolution of complex fields. *Opt Express* 18:19462–19478. <https://doi.org/10.1364/OE.18.019462>.
- Park Y, Diez-Silva M, Popescu G, Lykotrafitis G, Choi W, Feld MS, Suresh S. 2008. Refractive index maps and membrane dynamics of human red blood cells parasitized by *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 105:13730–13735. <https://doi.org/10.1073/pnas.0806100105>.
- Yakimovich A, Witte R, Andriasyan V, Georgi F, Greber U. 2018. Label-free digital holo-tomographic microscopy reveals virus-induced cytopathic effects in live cells. *mSphere* 3:e00599-18. <https://doi.org/10.1128/mSphereDirect.00599-18>.
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521:436. <https://doi.org/10.1038/nature14539>.
- Jo Y, Park S, Jung J, Yoon J, Joo H, Kim M-H, Kang S-J, Choi MC, Lee SY, Park Y. 2017. Holographic deep learning for rapid optical screening of anthrax spores. *Sci Adv* 3:e1700606. <https://doi.org/10.1126/sciadv.1700606>.
- Ferrari A, Lombardi S, Signoroni A. 2017. Bacterial colony counting with convolutional neural networks in digital microbiology imaging. *Pattern Recognition* 61:629–640. <https://doi.org/10.1016/j.patcog.2016.07.016>.
- Fisch D, Yakimovich A, Clough B, Wright J, Bunyan M, Howell M, Mercer J, Frickel E. 2019. Defining host-pathogen interactions employing an artificial intelligence workflow. *Elife* 8:e40560. <https://doi.org/10.7554/elife.40560>.