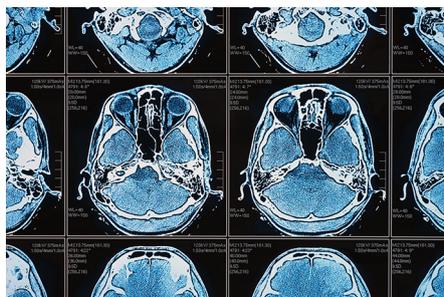


Ascent of machine learning in medicine

Machine learning is swiftly infiltrating many areas within the healthcare industry, from diagnosis and prognosis to drug development and epidemiology, with significant potential to transform the medical landscape.

The field of medicine has so far relied heavily on heuristic approaches, whereby knowledge is acquired through experience and self-learning, which is imperative in the highly variable healthcare environment. The increase in knowledge and understanding of diseases has been associated with the growth in information and data partly thanks to advances in tools that generate quantitative and qualitative measurements of physiological parameters. Such a big data field is ripe for the application of machine learning (ML). Indeed, there is a growing realization of the potential of ML as a platform that can glean information from numerous sources into an integrated system that can significantly aid decision-making processes for highly skilled workers. When ML was still in its nascency, it was conjectured that the success of an intelligent system that can learn and improve would depend on the ability to collect and store large amounts of data in a knowledge base¹. The improvements that have been made in computational resources as well as data storage and sharing in the past decade have been a significant enabler in harnessing the potential of machine learning systems in medicine.

In this issue, we invited a number of expert researchers in the field of ML to discuss the procedures, advances, challenges and future directions of this growing area for diagnosis, prognosis and drug development. In a [Comment](#), Cameron Chen and colleagues discuss the key principles necessary for developing ML models specific for diagnosis and prognosis of diseases such as diabetic retinopathy. They highlight the key steps involved in the path to medical deployment, from problem selection and data collection to model development, validation and monitoring. Minh Doan and Anne Carpenter's [Comment](#) focuses on the potential and use of ML in image-based diagnosis. While the majority of diagnostic assays rely on imaging labels to determine the presence of specific biomarkers in cells and tissue, the promise of label-free techniques is significant, especially in complex diseases where heterogeneity in disease hallmarks may affect accuracy in detection and categorization. Indeed, the combination of deep learning algorithms with label-free imaging has shown the ability to classify T-lymphocytes against colon cancer epithelial cells with high



Machine learning algorithms can be trained to detect complications on medical imaging data. Credit: JIRAROJ PRADITCHAROENKUL/Alamy Stock Photo.

accuracy². Therefore, the potential to utilize ML to aggregate large datasets would significantly accelerate the process of disease identification.

Equally exciting is the progress that has been made so far in applying computational methods such as ML in drug discovery and development. Only two decades ago, the timeframe for the development of a new drug was approximately 12 years, yielding a pre-approval cost of almost US\$1 billion³. Sean Ekins and colleagues discuss in a [Perspective](#) how ML has been exploited to expedite the process of drug development by exploring the possibility of incorporating models in all aspects of the pathway to predict the viability of a molecule for clinical use. In addition, a [Comment](#) by a panel of scientists working at the interface of ML and pharma highlights the potential and challenges of ML in pharmacokinetics, specifically in predicting absorption, distribution, metabolism, excretion and toxicology properties of new drugs. Adopting this approach could play a significant role in discovering new molecules or repurposing existing drugs for rare conditions or epidemics where urgency is key. With the increase in antibiotic resistance, exploiting ML techniques is already proving quite powerful in identifying new antibacterial agents in a faster and potentially cheaper way⁴.

Using ML in these contexts relies on the collection and analysis of large quantities of data, but with the emergence of big data comes the challenge of statistical inference from complex datasets to

recognize genuine patterns, while also limiting false classifications and making conclusive judgments on diagnosis and treatment options. Statistical bioinformatics has proven very useful in proteomic and genomic data analysis, and the adoption of ML to build predictors and classifiers has shown significant potential. In a [Comment](#), Andrew Teschendorff presents guidelines on how to avoid common pitfalls when utilizing ML with big 'omics' data. One of these pitfalls is known as the 'curse of dimensionality' and relates to the problem of the amount of sufficient training data required to make meaningful models from high-dimensional data. Indeed, naive use of ML models in omics data can easily result in overfitting, whereby random variations in the data that are not associated with real biological variations are classified as a phenotype or observable characteristic of interest.

It is evident that ML is creating a paradigm shift in medicine, from basic research to clinical applications, but it should be carefully adopted. Vulnerabilities such as security of data and adversarial attacks, where a malicious manipulation in the input can result in a complete misdiagnosis, which could be utilized for fraudulent interests, present a real threat to the technology. This was recently demonstrated by a group of researchers who showed that a carefully calculated perturbation on an image of a benign skin mole that is imperceptible to the human eye can be misclassified as a malignant mole, with 100% confidence⁵. In addition, guidelines should also be established to ensure that there are not only global standards to protect patient data but also a consideration for the society as a whole and the impact that this growing field is having on them⁶. However, collaborative efforts that include all major stakeholders will certainly add to the positive influence that ML adoption is having in research and medicine. □

Published online: 18 April 2019
<https://doi.org/10.1038/s41563-019-0360-1>

References

1. Clancey, J. W. & Shortliffe, E. H. *Readings in Medical Artificial Intelligence: The First Decade* Ch. 1 (Addison Wesley, 1984).
2. Chen, C. L. et al. *Sci Rep.* **6**, 21471 (2016).
3. DiMasi, J. A., Hansen, R. W. & Grabowski, H. G. *J. Health Econ.* **22**, 151–185 (2003).
4. Fjell, C. D. et al. *J. Med. Chem.* **52**, 2006–2015 (2009).
5. Finlayson, S. G. et al. *Science* **363**, 1287–1289 (2019).
6. Smallman, M. *Nature* **567**, 7 (2019).