Guest Editorial

Machine learning in clinical care: Quo vadis?

The history of artificial intelligence (AI) and machine learning (ML) is deeply intertwined with the history of computer science and the advent of computers over 60 years ago. It was not long after the birth of AI that the potential of AI to transform clinical care was recognized. An article in the *New England Journal of Medicine*[1] ventured to predict that computers "augmenting and, in some cases, largely replacing the intellectual functions of the physician." Designers of early AI systems attempted to distill the knowledge of experts into a collection of rules that the machine could use to perform the desired task, for example, diagnosing cancer or recommending treatment for hypertension. However, such rule-based systems failed to live up to the hope that many had placed in AI to transform medicine. There were many reasons for this failure: Experts, however good they were at their task, were often terrible at explaining how they arrived at their decisions. Fields like medicine were found to be "so broad and complex that it was difficult, if not impossible, to capture the relevant information in rules."[2] The resulting disillusion with AI beginning in the late 1970s ushered in a period of pessimism in the future of AI resulting in a severe cutback in funding, followed by the end of serious research for over a decade.

Fast forward to 2019. AI and ML are being heralded as panacea for addressing societal challenges, be it improving health, alleviating hunger, protecting the environment, or transforming education. So what changed? For one, starting in the mid 1980s, there was a shift in focus from rule-based systems to ML systems trained on specific tasks using large amounts of training data. An ML algorithm is provided with examples in the form of inputs (represented by features) and outputs (desired labels). For instance, digitized images of the retina read by ophthalmologists are represented by features (pixels) and labels (e.g., whether the image includes evidence of macular degeneration). ML algorithms construct from the training data, a predictive model that maps from features to labels. A key aspect of effective ML algorithms is the ability to generalize beyond the training data to correctly label new inputs (e.g., retinal images that have not been read by an ophthalmologist). In one sense, this process is similar to that of traditional statistical models, for example, regression: there are covariates, an outcome, and a statistical function linking the covariates to the outcome. But where ML shines is in handling enormous numbers of features or predictors, with modern methods capable of handling far more predictors than data samples, heterogeneous data types (from images, to genomics and clinical test results, and longitudinal observations from electronic health records), and combining them in complex, nonlinear ways to accurately predict, for each individual patient, his or her health status, health risks, and likely treatment outcomes. This has opened up the possibility of harnessing new kinds of data, including data that were not gathered specifically in a healthcare setting (e.g., diet, physical activity, lifestyle), the so-called "big data" to personalize treatments and improve health outcomes.

There is some evidence that ML could improve prognosis by integrating fine-grained data from multiple organs, infections, uncontrolled symptoms, and much more. ML is likely to dramatically change, if not displace, much of the work of radiologists and anatomical pathologists who focus on reading digitized images. There is some evidence that ML, either on its own or working in concert with expert physicians, can improve diagnosis. There is a growing interest in the promise and potential of ML to improve clinical care.[3]

In ophthalmology, ML has been applied to fundus photographs, optical coherence tomography, and visual fields, achieving good performance in the detection of diabetic retinopathy,[4] glaucoma, and age-related macular degeneration.[5,6] ML-trained predictive models in ocular imaging may be used in conjunction with telemedicine to screen, diagnose, and monitor major eye diseases for patients in primary care settings.

Is it responsible practice to let predictive models trained using ML perform diagnosis and dictate treatment in life-changing high stakes scenarios? How can we be sure that such models can be trusted? It is possible for ML to "overfit" predictions to noise in the data, especially when the number of training examples is small relative to the number of features, leading to overly optimistic estimates of the model's performance. This underscores the need for rigorous testing and validation of the trained models on truly independent validation data sets, sampled from a population that is different from that which provided the training data. Validation must consider class imbalance in the data, relative costs of false positives versus false negatives, and the inevitable tradeoffs between them.

Modern ML algorithms, for example, deep learning methods that are all the rage today, are extremely data-hungry, often requiring tens of thousands of training examples to attain acceptable performance. Is all that is needed is a large quantity of data and a powerful computer to process it? There is a popular saying in computing: "Garbage in, garbage out." ML is no exception. Biases in data collection can substantially affect generalizability of the trained model beyond the population that it was trained on. For example, a model trained on a database of

Caucasian patients may be worthless when used to make predictions on Asian patients.

ML algorithms are essentially tools for discovering complex correlations between the features or predictors and the outcomes or labels. Correlation does not imply causation. However good a model trained using ML may be at predicting health status, recommending treatments, and so on, it is important to remember that it is not a causal model. Predictive features are not causes of the predicted outcomes.

ML algorithms often produce black box models that make it hard to understand or explain the model's predictions. Should physicians blindly trust such predictions? There is much work that is needed on enhancing the explainability of ML in general, and in medical applications of ML in particular.[7]

If clinicians turn to ML-trained predictive models for diagnosis and advice about treatments, and not simply as a support tool, the predictive models will become key players in the relationship between physicians and their patients, raising a host of legal and ethical concerns.[8] If a predictive model trained using ML makes a fatal mistake, who is accountable? Is it the physician who relied on the trained model? Is it the scientist who designed the ML algorithm? Is it the engineer who trained and validated the model? Is it the hospital that chose to deploy the model? This is largely uncharted territory from a legal and regulatory perspective.

It is imperative that physicians who use ML systems become educated about how ML algorithms function, the data sets that are used to train the models, and their limitations. This calls for fundamental changes in the training of physicians.

Finally, because ML systems are trained on labeled data, they can amplify the implicit individual or collective biases of the physicians who provided the labels. There is a risk that the predictive models trained on such data simply perpetuate or amplify undesirable bias. For example, such bias could result in patients being recommended different treatments, not based on clinical considerations, but their ability to pay. Ensuring that ML algorithms when used in clinical care settings are immune to such undesirable bias presents significant challenges. Recent work on algorithmic fairness[9,10] might have some bearing on this matter.

Acknowledgement

This work was funded in part by grants from the National Institutes of Health NCATS through the grant UL1 TR002014 and by the Edward Frymoyer Endowed Professorship at Pennsylvania State and the Sudha Murty Distinguished Visiting Chair in Neurocomputing and Data Science funded by the Pratiksha Trust at the Indian Institute of Science (both held by Vasant Honavar). The content is solely the responsibility of the author and does not necessarily represent the official views of the sponsors.

Vasant G Honavar

College of Information Sciences and
Technology, Computer Science, Informatics, Bioinformatics and Genomics, and
Neuroscience Graduate Programs,

Artificial Intelligence Research Laboratory, Center for Big Data Analytics and Discovery Informatics,
Institute for Cyberscience, Penn State Institute for Clinical and Translational Research, Huck
Institutes of the Life Sciences, University Park, PA, USA.

E-mail: vhonavar@ist.psu.edu

References

1. Schwartz WB. Medicine and the computer – The promise and problems of change. N Engl J Med 1970;283:1257-64.
2. Schwartz WB, Patil RS, Szolovits P. Artificial intelligence in medicine – Where do we stand? N Engl J Med 1987;316:685-8.
3. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380;1347-58.
4. Padhy SK, Takkar B, Chawla R, Kumar A. Artificial intelligence in diabetic retinopathy: A natural step to the future. Indian J Ophthalmol 2019;67.
5. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, *et al*. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol 2019;103:167-75.
6. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: A review. Clin Exp Ophthalmol 2019;47:128-39.
7. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? 2017 arXiv preprint arXiv: 1712.09923.
8. Danton SC, Shah NH, Magnus D. Implementing machine learning in health care – Addressing ethical challenges. N Engl J Med 2018;378:981.
9. Binns R. Fairness in machine learning: Lessons from political philosophy. Conference on Fairness, Accountability and Transparency (pp. 149-59), 2018.
10. Barocas S, Bradley E, Honavar V, Provost F. Big data, data science, and civil rights. 2017;arXiv preprint arXiv: 1706.03102.

Access this article online

Quick Response Code:

Website: www.ijo.in

DOI: 10.4103/ijo.IJO_1167_19

Cite this article as: Honavar VG. Machine learning in clinical care: Quo vadis? Indian J Ophthalmol 2019;67:985-6.

About the author

Vasant G Honavar, MS, PhD

Vasant G Honavar is a Professor of Informatics at Pennsylvania State University where he holds the Edward Frymoyer endowed chair and directs the Artificial Intelligence Research Laboratory. He also serves as the Sudha Murty Distinguished Visiting Chair in Neurocomputing and Data Science at the Indian Institute of Science. He received his PhD from the University of Wisconsin-Madison in 1990. He has published over 300 peer-reviewed articles in Artificial Intelligence, Machine Learning, Causal Inference, Bioinformatics, and Health Informatics. Some of his current work focuses on predicting health risks and health outcomes from clinical, socio-demographic, and environmental, data. Details: http://faculty.ist.psu.edu/vhonavar