Demystifying the Jargon: The Bridge between Ophthalmology and Artificial Intelligence

Aaron S. Coyner, BS
J. Peter Campbell, MD, MPH
Michael F. Chiang, MD, MA

Publications related to artificial intelligence (AI) and machine learning have risen exponentially in the past 5 years in the medical literature, including a number of articles involving retinal disease. The mathematical theories beneath machine learning methods have been around for decades but in most cases were too computationally intense to implement by hand. Recent advances in computer central processing units and graphics processing units have enabled the application of these models to solve real-world problems. This has led to rapid advances in the fields of AI and machine learning—and specifically deep learning—and to a growing number of medical and ophthalmologic applications. As with the rapid evolution of any new technologies, there can be confusion about new terminologies, and AI is no exception. Though they are often used interchangeably, the terms “artificial intelligence,” “machine learning,” “deep learning,” and “neural networks” may be confusing for ophthalmologists to distinguish, and they are not synonymous. Below, we attempt to define these terms in a manner accessible to both ophthalmologists and vision researchers (Figure 1).

Definitions

Artificial Intelligence

AI is a broad category for all of the methods described herein. In general, there are two main types of AI. General AI describes algorithms, which are typically implanted in machines, that have the ability to “replicate human thought, emotion, and reason (and remain, for now, in the realm of science fiction).” The focus of this article is narrow AI, which comprises algorithms “that perform specific tasks as well as, or better than, humans” and are not necessarily implanted in robotlike beings. Within AI, there are several broad approaches, a few of which are described below; each approach has advantages, disadvantages, and ideal applications, which are beyond the scope of this brief overview.

Machine Learning

Non-AI diagnostic scoring systems and risk models have existed for many years. Often based on empirical data, scales such as the Early Treatment for Diabetic Retinopathy Study scale assign weights to predetermined disease features that algebraically sum to a final score of some prognostic significance. Machine learning is a category within AI, of algorithms in which the computer (rather than the human) “learns” the complex relationships and patterns from data that are then used to generate some “output,” which may be a diagnosis, severity score, or treatment recommendation. These algorithms include very simple to extremely complex models, such as linear and logistic regression, naïve Bayes classifiers, support vector machines, decision trees and random forests, and deep convolutional neural networks. Though the relative weights of features are determined by the computer (rather than the human) in machine learning algorithms, they are defined (and, therefore, somewhat interpretable) and based only on features that are provided to the computer a priori.

Deep Learning

Deep learning is a specific category of machine learning in which features are not assumed or input by the programmer. Rather, deep learning algorithms use data sets to create their own abstract features, which are then used to perform classification or regression. For example, Gulshan et al. demonstrated that a deep learning network could classify diabetic retinopathy, in agreement with the Early Treatment for Diabetic Retinopathy Study scale, using only retinal fundus images as input and the consensus diagnoses of multiple clinicians as the “class labels,” without being told that features such as microaneurysms, intraretinal hemorrhages, or neovascularization were important. Although these methods are powerful, a significant caveat is that their generated features remain opaque and uninterpretable. Thus, these methods are often referred to as “black boxes.” However, it has now become clear that, with enough labeled data, deep learning methods can approach or exceed human performance for a variety of tasks, including medical image assessment.

Artificial and Convolutional Neural Networks

Artificial neural networks (ANNs) are a specific type of machine learning algorithm that were modeled after biological neural networks to take advantage of Hebbian theory. Much like biological neural networks, ANNs contain many nodes (analogous to cell bodies) that communicate with other nerve cells via weights (analogous to axons and dendrites). This type of feed-forward network takes advantage of the “nerves that fire together, wire together” postulate, in which synapses between neurons are strengthened when their
neurons have correlated outputs. A special case of the ANN is the convolutional neural network (CNN). In a simple ANN architecture, all nodes in a layer are connected to all other nodes in the next layer. For image recognition tasks, this is not ideal; instead, a CNN feeds patches of an image to specific nodes in the next layer of nodes, thereby preserving the spatial context from which a feature was extracted. In this case, filters are trained to extract specific features from images (e.g., vertical lines, triangles) and mark their location on a feature map. A deep CNN then uses the feature map as input for the next layer, which uses new filters to create a new feature map. As this process continues, extracted features become highly abstract but highly useful for prediction. For this reason, deep CNNs have become particularly common for image-based diagnostic systems in ophthalmology.

**Generative Adversarial Network**

As opposed to discriminative algorithms (discussed above), generative algorithms predict—or generate—new data from class labels. A generative adversarial network is a learning procedure that pits a generative CNN against a discriminative CNN. For example, imagine we’d like to train a CNN to produce realistic retinal fundus images. To do so, we train a CNN to discriminate (recognize) retinas and another CNN to, initially, generate (create) random images. The discriminator alerts the generator as to how similar its images are to those of an actual retina. As training progresses, the generator learns key features of the retina and begins producing true-to-life retinal fundus images.

**Black Box**

Deep learning methods are often referred to as “black boxes” because an input goes into the box, and an output emerges, but it is not exactly clear what happens in between. In contrast, simple linear algorithms which, while not always as powerful as deep learning methods, are highly interpretable: one can examine the coefficients by which each feature is multiplied and discover important predictors for the task at hand. With deep learning, a complex series of matrix multiplication and abstract image filters makes the interpretability of this task significantly harder.

**Heatmap**

Generating heatmaps, also known as “class activation maps,” is a method that attempts to address the “black box” dilemma. This simple technique highlights areas on images that contributed to their classification labels. In the case of a retinal image detector, one might expect that a heatmap would show the optic nerve, macula, and retinal vasculature (or some combination thereof) to be highlighted, assuming that these areas contain the most critical information within the image. Improved methods of explaining the inner workings of deep learning networks using heatmaps are an active area of research.

**Conclusions**

Ophthalmological diagnosis is heavily dependent on the interpretation of images, which is often subjective and qualitative. AI has potential to disrupt this paradigm by providing objective data to assist with diagnosis. As such, it will be important for clinicians to recognize the potential benefits of AI for improving the quality and delivery of ophthalmic care, as well as its limitations and possible pitfalls. Understanding the terminology in this article is an important first step.

**References**


Footnotes and Financial Disclosures

1 Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon.
2 Department of Ophthalmology, Casey Eye Institute, Oregon Health and Science University, Portland, Oregon.

Financial Disclosures: The authors made the following disclosures: M.F.C.: Scientific advisory board — Clarity Medical Systems; Consultant — Novartis; Initial member — Inteleretina, LLC.

Supported by grants T15LM007088, R01EY19474, P30EY010572, and K12EY027720 from the National Institutes of Health (Bethesda, Maryland) and by unrestricted departmental funding from Research to Prevent Blindness (New York, New York). No funding organizations had any role in the design or conduct of this research.

Correspondence:
Michael F. Chiang, MD, Casey Eye Institute, Oregon Health and Science University, 3375 SW Terwilliger Boulevard, Portland, OR 97239. E-mail: chiangmi@ohsu.edu.