AI, Machine Learning, and Novel Statistical Methods in Biomedical Research

Yu Shyr, Ph.D.
Harold L. Moses Chair in Cancer Research
Chair, Department of Biostatistics
Director, Vanderbilt Center for Quantitative Sciences
Director, Vanderbilt Technologies for Advanced Genomics Analysis and Research Design
Professor of Biostatistics, Biomedical Informatics, and Health Policy
Vanderbilt University Medical Center
Nashville, TN, USA

For physicians, advanced data science methods such as artificial intelligence (AI), machine learning and novel statistical methods are no longer just a headline in the news. Use of advanced data science methods are gaining acceptance in the field of medicine, as researchers push the limits of rapidly progressing technologies to assist in delivering excellence in health care.

With the growth in use of advanced data science methods, however, we must bear in mind the limitations of these technologies. AI—specifically, machine learning (ML) of the deep-learning type—is only as good as the training dataset from which the algorithm learns. Deep learning in the clinical setting is most highly developed for image recognition; a training dataset for this purpose consists of a large number of case and control images. The algorithm then uses these images to teach itself how to differentiate, for example, malignant lesions from benign tumors or normal tissue. If the training dataset consists only of unambiguous cases and controls—rather than representing real-world variation and ambiguity—the algorithm will fail to function as desired with real-world patients. Thus, well-considered experimental design remains essential to realize the promise of deep learning.

This one-hour lecture will provide an overview of the state of the art in AI, machine learning, novel statistical methods, and deep learning algorithm for clinical application, including examples from the research literature as well as FDA-approved devices that use advanced data science methods. We will then address the pitfalls of deep learning, including the need for clear and rigorous standards for regulatory approval of devices and software. We close with thoughts on the future of advanced data science methods in medicine.