FDA Regulatory Framework and Evaluation Methods for AL/ML-based Decision Tools

Nicholas Petrick, Aria Pezeshk, Berkman Sahiner
Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD

Introduction: Artificial intelligence and machine learning (AI/ML) tools are quickly becoming ubiquitous in modern society. While the introduction of digital health tools, especially those directly supporting medical decision making, have seen slower migration to the clinical setting compared to consumer AI, clinical decision support tools are increasing in availability. The expectation is that the number of digital health tools will continue to grow with the potential to actually transform the clinical decision-making process in the foreseeable future. A continued challenge for the FDA is the regulation of machine-learning-based devices in a least burdensome manner while ensuring their safety and effectiveness.

In this invited talk, I will review some of the pathways to the market for new devices and discuss several of the recently-granted De Novos for image interpretation machine learning devices that can now be used as predicates for new devices submission with similar indications and similar technologies. My presentation will go on to discuss specific approaches and methods for assessing and evaluating AI/ML tools designed to aid medical image interpretation. This latter discussion of assessment methods will include methods for assessing both an algorithm’s standalone performance and methods to assess its ability to directly aid clinical decision making. I will also outline the current state of the FDA’s Software Precertification Program, which aims at developing a more streamlined and efficient regulatory oversight of software-based medical devices developed by manufacturers who have demonstrated a robust culture of quality and organizational excellence, and who are committed to monitoring real-world performance of their products once they reach the U.S. market.

In the final part of the presentation, I will summarize some of the research conducted in our laboratory at the Division of Imaging, Diagnostics and Software Reliability related to the assessment of machine learning for medical image interpretation. This includes our research on techniques to reduce the impact from multiple use of a test data set in the evaluation of a machine learning algorithm, and how training or test data sets may be augmented using synthetic images that are generated based on principles of physics and/or image-processing.

Translational Impact: This invited talk discusses current FDA thinking on evaluation of medical devices that utilize ML/AI to assist with clinical decision making.

Disclosure Statement: The authors have no conflicts of interest.