Data Curation Vital in Applying Predictive Analytics to Patient Care

Successful AI and machine-learning capacities within health systems require a foundational attention to data quality and the broader care environment. Adam Patterson
Thu, 06/11/2020 - 16:17

To achieve the potential of predictive diagnostics through the use of artificial intelligence in health care systems, organizations must focus on data quality and curation, said Jonathan Weiner, co-director of the Johns Hopkins Center for Population Health IT.

Weiner noted the importance of building machine-learning capacities atop the input of well-structured data for private and federal health care networks.
“High quality data in, high quality data out,” Weiner said. “Eighty-five percent of the issues in applying data to improving quality of care rest on effective data management and curation.”

While AI/ML capacities have shown demonstrable promise in areas such as predictive diagnostics, one of the most substantial challenges in applying patient information within these fields rests on finding means of standardizing health data, a quandary Weiner recognized is compounded by the fact that “most electronic health records are something like 50 to 60% free text.”

The push toward EHR standardization remains an ongoing project across all American health care networks, with Weiner saying “there is no health care organization in the world that has completely solved the problem of EHR standardization.”

The application of machine-learning capacities to data standardization also requires the analytic methods and data sets used in their development be fundamentally sound.

“When analytics are good, adding machine learning on top of that will give you a modest benefit,” he said. “But if you don’t have clean or well-curated data and just throw machine learning at it without those other steps, you’ll have a mess.”

Using machine-learning capacities to standardize data without these essential inputs results in a fundamentally ineffective process. “Only if you apply higher-order machine learning with a team that has already used those other necessary techniques will you start to see serious benefits,” Weiner said.

In terms of broader structural development necessary to streamline the delivery of health care, Weiner said the incorporation of AI and machine-learning techniques is merely one element of any comprehensive modernization program that can be used to improve health care delivery, rather than a process that wholesale replaces the necessary human inputs.
“The essential purpose of any automating function is to save human input for tasks it would be best used for,” he said. “Most of the time when people say, ‘AI is better,’ what they mean is I can diagnose patients better, identify veterans who have poor continuity better. I have not found many studies where AI or machine learning is a significant improvement simply on its own.”

AI and machine-learning techniques are, therefore, best understood as an accessory to health care modernization and continuity of care. “ML and AI have to be taken within the broader data application and human-machine interface,” Weiner said.

He said this approach has been validated by research and analysis he has overseen at the Johns Hopkins Center for Population Health IT, with the ultimate value of digital modernization measured by its ability to provide improvements commensurate with changes in the broader healthcare environment.

“We need to remember that medical care has to be integrated within the broader health environment, including the social environment and the mental health sphere. Integrating medical records is one thing, but there are a whole array of human considerations that have to be incorporated as well,” he said.