Researchers unveil framework for sharing clinical data in AI era

March 24 2020

Clinical data should be treated as a public good when it is used for secondary purposes, such as research or the development of AI algorithms, according to a special report published in the journal *Radiology*.

"This means that, on one hand, clinical data should be made available to researchers and developers after it has been aggregated and all patient identifiers have been removed," said study lead author David B. Larson, M.D., M.B.A., from the Stanford University School of Medicine in Stanford, California. "On the other hand, all who interact with such data should be held to high ethical standards, including protecting patient privacy and not selling clinical data."

The rapid development of AI, coming on the heels of the widespread adoption of electronic medical records, has opened up exciting possibilities in medicine. AI can potentially streamline and improve the analysis of medical images, but first it must be trained on large troves of data from mammograms, CT scans and other imaging exams. One of the current limitations of the advancement of AI-based tools is the lack of broad consensus on an [ethical framework](#) for sharing clinical data.

"Now that we have electronic access to clinical data and the data processing tools, we can dramatically accelerate our ability to gain understanding and develop new applications that can benefit patients and populations," Dr. Larson said. "But unsettled questions regarding the ethical use of the data often preclude the sharing of that information."

To help answer those questions, Dr. Larson and his colleagues at Stanford University developed a framework for using and sharing clinical data in the development of AI applications.

Arguments regarding the sharing of clinical data traditionally have fallen into one of two camps: either the patient owns the data or the institution does. Dr. Larson and colleagues advocate for a third approach based on the idea that, when it comes to secondary use, nobody truly owns the data in the traditional sense.

"Medical data, which are simply recorded observations, are acquired for the purposes of providing patient care," Dr. Larson said. "When that care is provided, that purpose is fulfilled, so we need to find another way to think about how these recorded observations should be used for other purposes. We believe that patients, provider organizations, and algorithm developers all have ethical obligations to help ensure that these observations are used to benefit future patients, recognizing that protecting patient privacy is paramount."

The authors' framework supports the release of de-identified and aggregated clinical data for research and development, as long as those receiving the data identify themselves and act as ethical data stewards. Individual patient consent would not be required, and patients would not necessarily be able to opt out of allowing their clinical data to be used for research or AI algorithm development—so long as their privacy is protected.

"When used in this manner," the article states, "clinical data are simply a conduit to viewing fundamental aspects of the human condition. It is not the data, but rather the underlying physical properties, phenomena and behaviors that they represent, that are of primary interest."

According to the authors, it is in the best interest of future patients for

researchers to be able to look "through" the data available in [electronic medical records](#) to develop insights into anatomy, physiology and disease processes in populations, as long as they are not looking "at" the identity of the individual patients.

The framework states that it is not ethical for clinical providers to sell clinical data for profit, especially under exclusive arrangements. Corporate entities could profit from AI algorithms developed from clinical data, provided they profit from the activities that they perform rather than from the data itself. In addition, provider organizations could share [clinical data](#) with industry partners who financially support their research, if the support is for research rather than for the data.

Safeguards to protect patient privacy include stripping the data of any identifying information.

"We strongly emphasize that protection of [patient privacy](#) is paramount. The data must be de-identified," Dr. Larson said. "In fact, those who receive the data must not make any attempts to re-identify patients through identifying technology."

Additionally, if a patient's name was unintentionally made visible—for instance, on a necklace seen on a CT scan—the receiver of the information would be required to notify the party sharing the data and to discard the data as directed.

"We extend the ethical obligations of provider organizations to all who interact with the data," Dr. Larson said.

Dr. Larson and his Stanford colleagues are putting the framework into the [public domain](#) for consideration by other individuals and parties, as they navigate the ethical questions surrounding AI and medical data-sharing.

"We hope this framework will contribute to more productive dialogue, both in the field of medicine and computer science, as well as with policymakers, as we work to thoughtfully translate ethical considerations into regulatory and legal requirements," Dr. Larson said.

 More information: "Ethics of Using and Sharing Clinical Imaging Data for Artificial Intelligence: A Proposed Framework" *Radiology*, 2020.

Provided by Radiological Society of North America