HIR
Healthcare Informatics Research

Managing Unstructured Big Data in Healthcare System

Hyoun-Joong Kong

Editorial Taskforce Member of Healthcare Informatics Research, Chungnam National University, Daejeon, Korea

The digital transformation or the 4th industrial revolution which are very recent information technology (IT) agenda make many countries expect the big data to be a source of new economic value that will determine the success and failure of those governments in the future. Due to this trend, the big data industry in the healthcare field has been growing rapidly in recent years and several global IT companies in the United States and Europe are reporting big data use cases in the medical field.

Medical big data refers to large-scale data that is difficult to handle with existing database management systems in a digitalized healthcare environment including medical centers, wearable devices, and social medias. The medical data, which are exploding exponentially, also include large volume of structured and unstructured data as other domains [1].

The big problem of healthcare fields is that about 80% of medical data remains unstructured and untapped after it is created (e.g., text, image, signal, etc.) [2]. Since it is hard to handle this type of data for Electronic Medical Record or most hospital information system, it tends to be ignored, unsaved, or abandoned in most medical centers for a long time [3]. Although a lot of data are still created in many hospitals, it is hard to be connected with medical big data research and artificial intelligence industry in healthcare. Therefore, we need to manage those unmanaged unstructured big data in healthcare systems before mentioning development of medical artificial intelligence which is currently based on machine learning technology.

In many hospitals, time series data are most unmanaged out of many types of unstructured medical data owing to its huge file size despite of the great value of their application. Typical unstructured big data in hospital are as following. The first type of data is medical video data that are recently created explosively from new types of medical imaging devices (e.g., endoscope, laparoscope, surgery robot, capsule endoscope, emergency video camera, thoracoscope, etc.). The second one is biosignal data that have been displayed on screen of patient monitor in operating rooms or intensive care units and wearable health monitoring devices. The third one is audio data that are verbally or nonverbally created from patients pathophysiologically and medical staffs for efficient communication in clinical procedures.

For enhancing the use of these unstructured medical big data, we need to establish the data collection, anonymization, and quality assurance processes. And meta data for each types of unstructured medical data need to be defined, standardized, extracted, and visualized automatically. Then open platform for integration and utilization of the unstructured clinical data should be developed while reflecting these concepts.

Even if machine learning technologies with high accuracy were developed, it would be useless without quality controlled, standardized and structured data for the unstructured medical big data. Besides, field-oriented education programs for nurturing multidisciplinary specialist who are able to interpret, analyze and utilize the unstructured medical big data should be discussed altogether with related healthcare industry-side.

References

1. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA 2014;311(24):2479-80.
2. HIT Consultant. Why unstructured data holds the key to intelligent healthcare systems [Internet]. Atlanta (GA): HIT Consultant; 2015 [cited at 2019 Jan 15]. Available from: https://hitconsultant.net/2015/03/31/tapping-unstructured-data-healthcares-biggest-hurdle-realized/#.XFvZ1lwvOUk.
3. Pak HS. Unstructured data in healthcare [Internet]. Fremont (CA): Healthcare Tech Outlook; c2018 [cited at 2019 Jan 15]. Available from: https://artificial-intelligence.healthcaretechoutlook.com/cxoinsights/unstructured-data-in-healthcare-nid-506.html.