Big data in healthcare

BLAKE LOUGHRAN  
Year 4 Medical Student, Sydney Medical School

EDWARD TOTH  
PhD Student, Faculty of Engineering and IT

The amount of data being generated in society is growing at an exponential rate, promoting increased efforts to consolidate and analyse it.

In health, a major source of this data is electronic medical records, already introduced in many countries. Genomic data has increased as costs involved have dramatically decreased. There is data streaming from sensing technologies, which are not currently suitable for clinical use, however have been popularised at the consumer level. Health-related data is also obtained from digital sources such as social media, blogs and Internet search results.

To be able to develop insights from the captured data, it must first be stored. While storage initially was an issue in because of the significant expense involved, the introduction of distributed and cloud computing has greatly improved feasibility of storage capacities. But the key issue for health remains the privacy and security of stored data.

A major current area of research is in analytical techniques, for example “machine learning” and “deep learning”, which are both ways of extracting meaning and insights from the data.

IMPACT ON MEDICAL DISCIPLINES

There will be many significant changes in medicine as a result of Big Data. The application of personalised medicine and clinical decision support systems will play key roles in clinical disciplines and practices, and doctors will need to develop strong technological proficiency and adapt the use of these systems in their practice.

Possible advancements in other areas of medicine include machine-aided identification in pathology and radiology. These are areas where computers currently do not have sufficient accuracy to automatically evaluate images, however it is expected that their performance will improve in the future to a level where they can provide additional assistance in image analysis. In the future, automated imaging analysis software has the potential to improve productivity, efficiency and effectiveness whilst reducing associated cost.

Epidemiological research is another area of medicine certain to be impacted. Skills as a data scientist will be at a very large premium, as it allows greater access and analysis of the large pool of available data. Across epidemiology, systematic approaches will need to be developed to manage, analyse, display and interpret large complex datasets. Epidemiologists will also need training to accommodate an increasing emphasis on collaboration and multilevel analysis.

Big data is a broad term for the use of large complex data sets, and with increasing relevance to health researchers and clinicians. In January 2015, Daniel Petre and Roger Corbett – both long term Sydney Medical School supporters – funded an international study undertaken by two Sydney students looking at developments in the world of health “big data.” The students – one medical, one engineering/IT – interviewed big data and medical specialists at universities and technology companies in United States and UK, an excerpt from their report is below.

RECOMMENDATIONS FOR MEDICAL EDUCATION

Doctors of the future will encounter a significantly different way of practising medicine compared to today. With the extensive advancements of various technologies and analytical techniques, the curriculum in medical education is bound to change.

In terms of teaching basic sciences, a more in-depth education on genetics may be required in medical curricula. This is due to the increasing availability of genetic information and the greater emphasis that has been placed on genetic research. With the application of personalised medicine, doctors of the future will benefit from advanced concepts in genetics. In particular, future doctors need to be skilled in interpreting results from genetic investigations and adept in basic genetic counselling.

It would also be advantageous for medical students to have an increased exposure to topics in Big Data, with introductory teaching in statistics and machine learning. In most medical schools, this does not necessarily require a significant change in the biostatistics courses that are offered, but merely a change in mindset of the lectures and teaching sessions. It would be optimal for courses to give students introductory exposure to the use of Big Data techniques (such as the use of extremely large cohorts or diverse data sources), ideally through problem-based learning, where experimental design techniques can be discussed.

Another aspect for education is gaining practical experience in dealing with technology and future innovations. Students will require a high degree of technological proficiency. The objective in this case is to ensure a minimum level of technological competence among graduates.