Big Data Analysis in Medicine

Dr. Ayman El-Baz
JB Speed School of Engineering
University of Louisville
Department of Bioengineering

The term “Big Data” refers to the amalgamation and processing of huge data sets that are composed of different data types (e.g., clinical, genomic, imaging, pathological, etc.) and have rapidly become more massive and complex, particularly with the advent of new technologies. Big Data within the context of biomedical research is a major problem that needs to be solved due to substantial increases in the amount of medical data routinely generated and collected by healthcare providers over the last two decades. A recent PubMed search for the term “big data” yields 1470 entries, with the earliest occurring in 2003. A breakdown by year shows the majority of publications are from 2012 or later. In 2011, the McKinsey Global Institute issued a 156-page report titled “Big data: The next frontier for innovation, competition, and productivity”. This report indicated $300 billion in potential annual value in Big Data to healthcare in the U.S., with a shortage of 140,000 to 190,000 individuals with the required deep analytical skills, indicating a need for programs to train the next generation of scientists with the necessary skill set to deal with all aspects of Big Data.