Adapt current tools for use with big data

To help speed up discoveries of disease *[OK?]* biomarkers and treatments, we need to work out a cheaper and faster way to process, store and use the huge medical data sets that are rapidly becoming available (see, for example, *Nature* **506,** 144–145; 2014).

By 2015, it is likely that a typical hospital will create 665 terabytes of data (for comparison, the web archive of the US Library of Congress contains less than 500 terabytes of data). This mass of information can be used to study and analyse treatments for tuberculosis and strokes, for example, and to reduce health-care costs.

But to adapt classical information-processing tools for use with modern medical data, *[OK?]* we need to address several computational challenges (see also *Nature* **498,** 255–260; 2013). One is how best to manage the exceptional volume of detailed data as it becomes available, without sacrificing information. Another is that the data mostly represent physiological processes, the characteristics of which change over time.

Current methods are also inadequate for analysing collective data originating from different sensors, such as multi-dimensional descriptions from electroencephalography or magnetic resonance imaging of interactions between brain regions.

Ervin Sejdić *University of Pittsburgh, Pennsylvania, USA. esejdic@ieee.org*