Editorial: Big Data for Health

DEVELOPMENT in the fields of biomedical and health informatics are driving major expansion in big data, not only because of the sheer volume of information generated, but also due to the complexity, diversity, and the rich context of the data that encompasses discoveries from basic sciences to clinical translation, to health systems and large-scale population studies on determinants of health. This general trend also brings socio–legal implications. While technological advances are overcoming many of the traditional barriers for transmitting, storing, and sharing information securely, health data are growing faster than healthcare organizations can consume it. Managing, extracting, analyzing, integrating, visualizing, and communicating useful information from the myriad of data generated continuously in real time in addition to existing biomedical and health information represent major challenges in big-data research. A further important challenge relates to translation of analytical outputs to useful intelligence for more effective clinical decision making and for policy formulation. New data analytic tools to facilitate scalable, accessible, and sustainable data infrastructure for the effective management of large, multiscale, multimodal, distributed and heterogeneous datasets and convert data into knowledge for support cost-effective decision aids, disease management, and care delivery need to be developed.

In promoting big data as a source of innovation in healthcare and accelerating the translational pathways from the laboratory bench to the patient's bedside, this special issue includes a collection of papers addressing some of medical and health informatics challenges related to big data. Papers submitted cover a range of topics from bioinformatics, imaging informatics, sensor informatics, and medical informatics to public health informatics. They also include initiatives that enable use of big-data analytics in health systems for improved clinical decision making, enhanced efficiency of care provision, policy development, and policy implementation.

The position paper by Andreu-Perez *et al.* provides an overview of recent developments in big data in the context of biomedical and health informatics. It outlines the key characteristics of big data and how medical and health informatics, translational bioinformatics, sensor informatics, and imaging informatics will benefit from an integrated approach by addressing different aspects of personalized information from a diverse range of data sources. Viceconti *et al.* further discussed the opportunities offered by big data for virtual physiological human, aimed at developing robust and effective *in silico* human models for medicine. By using multiple data sources, including those nontraditional information, such as Twitter feeds, Google search interests, and environmental sensor data, Ram *et al.* demonstrated their method for improving national asthma disease surveillance and predicting asthma-related emergency department visits in a specific area. Their preliminary findings show that the model can achieve about 70% precision based on environmental and social media data. By using regression decision tree analysis along with hospital admissions and insurance claim data from nearly a quarter of million individuals over three years, Xie *et al.* discussed their research in the context of big data for predicting the number of days in hospital and some of the specific issues for general population and subpopulation analysis. Fan *et al.* developed a hierarchical learning algorithm for classifying large-scale patient records for automatic treatment stratification. It has been shown that their method can leverage tree structure to train more discriminative max-margin classifiers for high-level nodes and control interlevel error propagation, thus achieving log-linear computational complexity suitable for large-scale analysis.

To demonstrate how data from social media can be used for the real-time analysis of expressed mood for mental health research, Larsen *et al.* describe their "We Feel" platform for analyzing global and regional variations in emotional expression, with a snapshot of 2.73×10^9 emotional tweets over a 12-week period annotated for emotion, geographic location, and gender. Such analysis may complement the traditional surveys and research on predisposition to mental health problems, shedding light on the fluctuations and regulation of emotions and behaviors in population groups.

The use of cloud computing in the context of big data is discussed by Lillo-Castellano *et al.* for automatic classification of intracardiac electrograms, aimed as a national level cloud-based big data service for implantable cardioverter defibrillators. A new compression-based similarity measure is created for low computational burden, allowing scale up to large populations.

With the availability of large-scale genomic profiling via NGS, it is now economically feasible to profile the whole transcriptome and genome of individual patients. This would be valuable for assessing drug-specific sensitivity as determined from the changes in genomic profiles of individual cell lines in response to a specific drug. In this context, Sheng *et al.* discussed a computational algorithm to predict the drug response of individual patients based on personal genomic profiles, as well as pharmacogenomic and drug sensitivity data. This study represents an encouraging step toward the practical realization of precision medicine for oncological care in the future.

Included in this special issue, we have also selected two imaging related papers to outline the role of imaging informatics for big data. The work by Mikhno *et al.* discussed the use of pharmacokinetic input function model and constraints derived from machine learning applied to electronic health records

Digital Object Identifier 10.1109/JBHI.2015.2442392

2168-2194 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

for the quantitative analysis of positron emission tomography brain imaging. Suinesiaputra *et al.* highlighted the importance of increased data sharing and the value of large cardiovascular epidemiological studies that enable collaborative efforts to facilitate algorithms benchmarking, disease modeling, and statistical atlases. Their effort toward big heart data demonstrates the strength of a synergistic approach by the research and clinical communities in understanding of cardiovascular function, disease progression, and therapeutics.

Organizing a special issue as such involves a huge amount of work, and we would like to thank all the reviewers for their hard work and cooperation in making this happen. Most important of all, we would like to thank all the authors in submitting their latest work to this special issue. These papers cover important practical experience of handling ever complex and heterogeneous health informatics, ranging from disease prediction, treatment stratification, personalized drug optimization to imaging and data sharing. Due to time and space constraints, we are only able to select a small collection of papers submitted to the special issue; many high-quality papers were not included due to volume and time constraints. We hope with further work, they will appear in future issues of this journal.

Last but not the least, we would like to thank the IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS editorial office, particularly H. Rafii-Tari and K. Vyas for all their editorial support throughout the preparation of this special issue.

R. ATUN, *Guest Editor*
Harvard University
Cambridge, MA 02138 USA
ratun@hsph.harvard.edu

Y. LUSSIER, *Guest Editor*
The University of Arizona
Tucson, AZ 85721 USA
yves@email.arizona.edu

C. POON, *Editor*
The Chinese University of Hong Kong
Shatin, Hong Kong
cpoon@surgery.cuhk.edu.hk

S. T. C. WONG, *Guest Editor*
Houston Methodist Research Institute
Weill Cornell Medical College
New York, NY 10065 USA
STWong@houstonmethodist.org

G.-Z. YANG, *Editor*
Imperial College London
London SW7 2AZ, U.K.
g.z.yang@imperial.ac.uk