A Dynamic Environment for Medical Image Computing Research in Canada

Amber L. Simpson[1,2,3], David Berman[4,5], Jaron Chong[6,7], Errol Colak[8,9], Aaron Fenster[5,6,7], Gabor Fichtinger[2], Masoom A. Haider[5,10], David Hocking[6], Omar Islam[11], David LeBrun[4,5], Anne Martel[3,12,13], Parvin Mousavi[2,3], Narinder Paul[6], Rossanna Pezo[14], Julia Publicover[5,13,14], William Tran[14,15], Aaron Ward[6,7], Martin J.Yaffe[5,12,13]

1. Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON
2. School of Computing, Queen's University, Kingston, ON
3. Vector Institute for Artificial Intelligence Faculty Affiliate, Toronto, ON
4. Department of Pathology, Queen's University, Kingston, ON
5. Ontario Institute for Cancer Research
6. Department of Medical Imaging, London Health Sciences Centre, London, ON
7. Department of Medical Biophysics, Western University, London, ON
8. Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, ON
9. Department of Medical Imaging, University of Toronto, Toronto, ON
10. AI, Radiomics & Oncologic Imaging Research Lab at Mount Sinai Hospital
11. Department of Radiology, Queen's University, Kingston, ON
12. Physical Sciences Platform, Sunnybrook Research Institute, Toronto, ON
13. Department of Medical Biophysics, University of Toronto, ON
14. Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, ON
15. Techna Institute, University Health Network, Toronto, ON

Historical Context and Opportunities

While radiologic and pathologic images were historically used mainly to convey anatomical information and tissue changes related to disease or injury, their scope has expanded enormously over the past 50 years. This expansion was catalyzed largely by the introduction of digital imaging methods in radiology beginning with some work in nuclear medicine and the introduction of x-ray computed tomography in 1973. In pathology, digitization of whole slide images of tissue is poised to similarly revolutionize the field. Virtually all of medical imaging is going digital, opening possibilities of more precise diagnoses, quantifying the extent of disease, planning interventions such as orthopedic procedures, radiation therapy and neurosurgery, monitoring changes in response to therapy and countless other applications.

With the development of improved imaging technology, spatial resolution, image sizes and image repetition rates have increased, and three-dimensional methods have been introduced. This has vastly increased the amount of quantitative information available from medical images, but also the volume of data that are acquired, stored, analyzed and managed. More recently, there has been a large effort internationally on developing more powerful methods for processing and analyzing medical images with a meteoric rise in activity around applications of radiomics and machine learning/artificial intelligence for these purposes.

Canada has in many ways been a leader in research in medical imaging due to the foresight of several leaders in radiology, pathology, computer science, and medical physics who developed imaging research programs at several Canadian universities supported by substantial government investments. Those seeds have grown and flourished, resulting in a vibrant, highly collaborative medical imaging research community where highly-trained personnel utilize state-of-the-art technology to develop and test applications in imaging informatics for the benefit of patients.

Some of the recognized Canadian developments include: high-resolution ultrasound systems for use both in preclinical research and drug development and for patients; digital mammography; systems and radiopharmaceuticals for positron emission imaging; image display and analysis software platforms for radiology, cardiology and pathology, robotic biopsy, surgery and intervention devices; focused ultrasound systems for therapy; and a growing enterprise of machine learning research for medical and pathology imaging at multiple institutions. As well, image-guided therapies such as the MR-Linac in radiation oncology and intraoperative magnetic resonance imaging are showing the importance of merging imaging technologies to improve treatments for cancer patients.

Current State

In addition to the development of hardware imaging systems, there is strong activity in the development of software, algorithms and platforms for the analysis and processing of images as well as in designing systems to support medical decision making. Much of this work utilizes machine learning approaches and these rely heavily on archives of previous images, coupled to correlated clinical information related to risk factors, treatments, outcomes and demographic information. **The speed and efficiency with which research and development can move forward is highly dependent on the facility with which such images and information can be obtained.**

Canada is an ideal location for developments in imaging in the health sciences, in part because of the publicly-supported health care systems. For example, in Ontario there are existing consolidated repositories of imaging data and patient information that can be leveraged for computing research such as HDIRS (imaging), ICES (non-imaging), CIHI, and Cancer Registry (cancer incidence, prevalence and outcomes) data. This environment should facilitate making medical images and data available for the benefit of Canadians and to drive research and development to strengthen Canada's high technology economy.

The requirements of Canadian medical researchers vary according to the type of work they do; however, one need that is common to many of them is for large numbers of high-quality curated medical and pathology images pertaining to many specific diseases and conditions for algorithm training and testing. In addition, many researchers require high-powered (data capacity and speed) computing resources and vast amounts of data storage. Key issues are convenient access, protection of patients' privacy, security of data and regulation of who obtains access to images and data and for what purpose.

With respect to the existing digital research infrastructure (DRI), many researchers utilize publicly-available databases; however, most of these are from international sources (e.g., The Cancer Imaging Archive, cBioPortal) and in some cases these may not be optimally relevant or adequately linked to other needed data such as clinical outcomes, disease presentations and other relevant clinical data, such as treatment and demographic data. Much of the most valuable clinical image material and data needed are locked, within institutional archives and access is often cumbersome, slow and replete with bureaucratic hurdles. Policies of archiving images vary greatly among institutions and the process of extracting needed clinical data from electronic medical records suffers from the fact that hospital systems are more orientated around billing for procedures than for organizing clinical data. Data are often not organized as synoptic elements according to well-defined lexicons, but are stored as free text, making analysis extremely labour intensive and susceptible to error. The sharing of health data is still the work of individual investigators rather than motivated by the health care system and this greatly reduces efficiency of effort. **A source of Canadian images with associated clinical data and outcomes would be extremely valuable for research and development and would better reflect the diverse make-up of the Canadian population.**

Some scientists utilize resources available through Compute Canada (Toronto, Ontario), primarily for training algorithms. However, convincing a hospital to allow upload of patient data to Compute Canada can become mired in red tape for individual researchers. Medical images are notoriously hard to anonymize due to a lack of standardized formats further complicating the transfer of data.

There are unique issues related to data curation and annotation when dealing with medical images. Data may be high dimensional and often multiparametric (3D, 4D, 5D). There is no standardization of annotation tools or of open source frameworks for analysis, although many have been developed by Canadian researchers. This leads to multiple in-house solutions that do not leverage common expertise of the wider community in a consistent fashion, leading to non-reproducible results and hindering translatability to clinical care. Annotation work is labour intensive, requiring expensive and difficult to obtain skilled human resources - e.g., diagnostic radiologists and pathologists. Efficient software tools are vital to maximizing the value of their efforts and success in assembling large, mineable imaging data sets. Beyond software, optimal leveraging of available expertise requires targeted and sustained efforts to break down interdisciplinary silos that currently exist between computational researchers, radiologists, pathologists and their respective trainees. **Such efforts should include joint undertakings in education, outreach and targeted support for trans-disciplinary research projects to ensure mutual understanding and appropriate academic recognition of contributions made by experts in diverse fields, facilitated by tri-council grant funding**.

The currently modest rate at which ideas are exchanged between computational researchers, radiologists and pathologists impedes the timely application of computational solutions to pressing clinical problems. Diagnostic challenges are encountered daily by clinicians that may represent "low hanging fruit" for computational researchers. Whereas clinicians generally lack

both the tools to address such problems and even an awareness that such tools exist, computational researchers are typically unaware of the clinical problems that are in need of solutions.

Our Vision for the (Near) Future

Canadian researchers would benefit greatly if a system were in place that facilitated access to clinical medical images, from radiology, pathology, cardiology, neurology, etc. These images would be most valuable if they were accompanied with or linked to annotations and relevant clinical data providing details on diagnosis (including molecular biomarkers), treatment, demographic information and outcomes. Tools for extracting clinical information from hospital records, for example utilizing natural language processing, are much needed to improve accuracy and practicality of conducting clinically relevant research. The development of validated software tools to remove protected health information from medical images could help to overcome regulatory barriers.

Various approaches to addressing these challenges are possible. In one, data would be federated across repositories such as Infoway, CIHI, ICES, HDIRS, and others. Integrated, uniform and accessible platforms for conducting analysis would allow multi-institutional studies to be conducted much more efficiently than at present. Models that involve bringing the data to the analysis centres, bringing the tools to the sites providing data or a hybrid of these could also be considered The National Institutes of Health Data Commons (https://commonfund.nih.gov/commons) is an example of this.

In the future DRI environment, data governance, privacy and security issues would be resolved to reduce barriers to research. A consistent framework, ideally accessible across the country, would provide access to such information for ethically approved research. Such a framework would require that a uniform code of principles on the ethical, equitable use of data originating from human subjects be developed and that potential users agree to comply with the provisions of that code. Principles of access should clearly state the conditions under which patient consent (or consent of kin) is or is not required. The structure should be based on realistic scenarios rather than total avoidance of risk, because of the enormous potential good that could come from the use of the data. If designed appropriately, **such a platform would reduce barriers to access for small research universities and health institutions to participate and contribute to research, trials and studies.**

Platforms providing well-designed and tested tools for segmentation and analysis, including a suite of tools for radiomic analysis would be of great value. The open source community has provided a start in this direction with tools such as the OHIF viewer and NVidia's image annotation server so these tools do not have to be reinvented *de novo.* Examples of current software projects in Canada include 3D Slicer, a software toolkit that supports medical imaging research supported by the CANARIE Research Platforms program, and the Pathology Image Informatics Platform funded by the National Institutes of Health. Use of a standard toolbox would contribute to consistency of measurements and the ability to reliably cross-compare data

from different laboratories. Without a common annotation and analysis platform investigators often are forced to reinvent the wheel. This is both inefficient and also may cause artifactual variability of results and impedes translation to clinical use.

Sufficient processing power would ensure that Canadian researchers could compete with private, often multinational interests - "giants" that have seemingly endless compute power. Although the economic value of research using medical images and data is unquestionable, there are other important, sometimes conflicting benefits, such as improvement in the health of Canadians. Simply allowing medical data to be "harvested" by large corporations is unlikely to maximize overall equity of benefit. Here, **the democratization of access to the data and to computing and analytical resources would contribute to create equity and spur creativity from groups too small to compete with these giants for those resources.**

SUMMARY

The usefulness of the DRI in Canada would be greatly enhanced by the development and incorporation of the following features on a consistent, national basis:

- **Governance and guiding principles for institutional research ethics boards for data and image sharing**
- **Persistent data storage that is FAIR (findable, accessible, interoperable, reproducible)**
- **Funding mechanisms that promote and encourage collaboration in the development of sharable code and tools**
- **Ability to integrate medical image data storage with high-power computing resources**
- **Democratized access to these resources**