Name and Title: Pascal Tyrrell PhD, Director of Data Science and Assistant Professor
Department: Medical Imaging and Statistical Sciences

TITLE OF RESEARCH PROJECT: Sample Size Determination Methodologies for Machine Learning Studies in Medical Imaging Research

Number of 299Y Spots: 1 Number of 399Y Spots: 1

OBJECTIVES AND METHODOLOGY:
Following business intelligence, internet security, and natural language processing, medicine - in particular radiology - has seen an explosion of interest in machine learning. Establishing principled methods for assessing the value of machine learning algorithms in medical imaging research remains a significant challenge. Existing image classification algorithms (that identify a set of categories to which images/observations belong) continue to elude expert observer confidence in conferring or confirming a diagnosis. In order to achieve sufficiently high accuracy, machine learning algorithms rely on a finite set of “training data” (set of images), which are used by the model to ‘learn’ and optimally refine parameters. Adequate sample size of training data and testing of classifiers must be established and documented to inform academic and industry-based medical imaging research. This is a particularly poignant challenge in medical imaging, where access to enormous quantities of high-quality data is typically elusive and hence researchers often resort to experiments based on relatively small sample sizes. The required sample size for a given model applied to medical imaging remains unknown.

Objectives:
1) Systematically identify and describe relevant studies pertaining to sample size determination methods for machine learning in medical imaging
2) Propose potential applications of these methods in establishing sample size guidelines for medical machine learning.
3) Validate a sample size determination methodology testing environment by creating simulations based on medical imaging datasets.

DESCRIPTION OF STUDENT PARTICIPATION:
Overview:
Our research education program for undergraduate students consists of mentorship by a faculty member, both formal and informal seminars, discussion forums, and direct research experience at various health centers (if applicable) and will provide students with a springboard to the field of medical imaging research. Students would
be introduced to the various sub-disciplines that have a direct impact on medical imaging associated clinical care; a few of which include:

- Medical imaging sciences and technology (machine learning)
- Clinical and outcomes research
- Research methodology and biostatistics

Candidates who complete the course will gain insight into research development and clinical/social application of medical advances in the field of medical imaging research.

Educational activities typically include:
- Two (2) hours spent on continuing education per week - weekly seminars as well as at least one (1) hour a week of additional educational activity within the program including education/research rounds, journal club, and lab meetings.
- An opportunity for participation in affiliated departmental educational programs – Medical imaging Buddies and Medical imaging Volunteer Internship Program.
- Research related activities:
  I. training in the assessment of pertinent medical and health literature using web-based medical and health journal libraries and databases for the purposes of study protocol development.
  II. applying auditing practices and accurate entry of information into research databases for the purposes of relevant data analysis pertaining to their research project.
  III. training in project (study) presentation reporting/manuscript preparation for stakeholders within the scientific establishment and public in general.
  IV. Introduction to data science methodologies including biostatistics, machine learning, and image processing.

- Opportunities exist to contribute to both new and ongoing publishable research studies including systematic reviews, retrospective/prospective cohort studies, cost effectiveness studies with statistical modeling and simulation, and machine learning projects focusing on image-based pattern recognition and classification.

**MARKING SCHEME (assignments with weight and due date):**
- Participation in seminars/meetings and other educational program activities (10%).
- Comprehensive background literature review paper on the research project. Due mid-November (10%).
- Student journal (infographic format) recording research learning experiences (10%).
- Two-page interim report detailing research experience by the student (5%). Due with literature review paper.
- The Course Director will equally complete an assessment of the student’s progress and discuss it with them before the deadline for dropping courses with penalty (refer to UofT school calendar for exact dates).
- Poster presentation in Undergraduate Research Forum for ROP students (held in March) (15%).
- Final report (~10 pages double-spaced, including introduction, methods, results, discussion, and references, in standard journal format; figures and legends should be included, but do not count toward the page limit) by end of early April (50%).