Radiomics 2012
Meeting Report

RJ Gillies, HLM
robert.gillies@moffitt.org

On Nov 1-2, 2012, a select group of researchers met in Clearwater, FL to discuss progress, challenges and future directions for Radiomics, which is the large scale conversion of medical images into mineable data. This was an international meeting, with investigators coming from as far away as China and the Netherlands. Although this meeting was held under the auspices of a State of Florida Grant and an NIH U01 grant associated with the Quantitative Imaging Network, it is notable that most investigators attended using their own funds.

The meeting was primarily focused on describing the decision support systems that can be developed from these data. Although the majority of the meeting was focused on lung cancer screening and diagnosis, the application of these approaches to other cancers was sometimes explicitly discussed. Related to this, a few significant hurdles were identified in the heterogeneity of standard of care images and the need to increase the size and availability of curated data sets.

There has been remarkable progress in the 18 months since the previous meeting held in April of 2011. Primarily among the progress were reports form a number of investigators who have used radiomics features to predict outcome or the presence of cancer. In addition, current limitations of image acquisition/processing, segmentation, feature validation, and data sharing were better delineated with considerable effort spent on defining actionable solutions.

The NCI/QIN was represented by **Larry Clarke** and **Robert Nordstrom**, who led the meeting with a discussion of the landscape of radiomics research within the NIH. Larry discussed increasing the collaborations with ECOG/ACRIN by providing advanced software tools for analysis of the data collected through their clinical trials. In addition, Larry and Bob described the new informatics initiatives announced by the NCI, and the need for them to have a biological connection.

The first session focused on lung cancer screening with low dose CT. This session was led by **Tony Reeves** (Cornell) who outlined some of the challenges and solutions to high throughput screening. Currently, the most validated predictor for the presence of cancer is growth of a nodule between subsequent scans. Growth can manifest as an increase in volume or density of a lesion. Development of appropriate risk assessment models will require data to be analyzed from thousands of patients and this requires robust automated segmentation routines. **Yuhua Gu** (Moffitt) and **Xiuli Li** (Beijing) presented their latest promising results on autosegmentation of lung fields and parenchymal texture analysis for nodule detection, combined with autosegmentation of nodules. Discussion of this paper also pointed out the need to compare algorithms across common data sets, such as LIDC and NLST. Lack of access to NLST remains an issue. **Ron Walker** (Vanderbilt) represented the EDRN's efforts to move lung cancer screening to the post-NLST level. He is coordinating a multi site trial with data acquired using standardized protocols. Data to date suggest that lesion doubling times can be assessed and that data may converge to recommend a fixed interval for follow up repeat scans to assess presence of growth. However, challenges still remain with respect to automated

size determination of GGOs. **Larry Hall** (Moffitt/USF) showed benign/malignant classification data from the LIDC, wherein he was able to obtain ~85% accuracy with J48 decision trees or support vector machines (SVM). Four features were able to predict 2-year survival to 70% and this was higher if a more homogeneous input data set was used.

A number of investigators discussed the use of Radiomics in prediction of radiation response. As pointed out by **Eduardo Moros** (Moffitt), image-guided radiation therapy involves acquisition of daily cone beam CTs using rigidly fixed acquisition conditions. These data coupled provide a rich and technically homogeneous data set for radiomics evaluation. **Tom Dilling** (Moffitt) described a highly curated data set of radiation therapy data with extensive metadata and outcomes. **Hugo Aerts** (Dana Farber) described an impressive body of work that developed a training set using radiotherapy planning CTs from the MAASTRO clinic (Maastricht), which was then prospectively tested against a similar large data set from Radboud (Nijmegen). He identified 12/422 features that were predictive of survival from this training and test set. Although there is further work to be done to validate these features, this was an impressive finding.

Apart from imaging in radiation oncology, the acquisition parameters become much more heterogeneous. **Steve Eschrich** (Moffitt) presented data from a retrospective analysis of CT scans from 285 NSCLC patients present in the PACS database at Moffitt. These data were heavily weighted towards patients who had microarray gene expression profiles, so were primarily stage I/II. Of these, only 208 patients were scanned at Moffitt, and the remainder of the patients were from consortium sites. Consequently, there were 19 different makes and models of scanners used, 25 different convolution kernels were used, X-ray tube voltages varied between 120-140 kVp, and reconstructed slice thicknesses varied between 1-10 mm. Not surprisingly, image features were significantly different between scanners, acquisition conditions, reconstruction algorithms and manufacturers. There was much discussion of these data with the realization that, while data in clinical trials and radiation therapy planning are relatively homogeneous, standard of care diagnostic CTs are not. . Heterogeneity of data acquisition, particularly regarding reconstruction filters and slice thicknesses, remain a source of variability in tumor characterization.

Binsheng Zhao (Columbia) arrived late because of Hurricane Sandy, but was nonetheless able to add some rigor to the debate over harmonized acquisition and reconstruction parameters. She showed, by Bland-Altman analyses of test-retest data, that higher resolution reconstruction (e.g., 1.25 mm and 2.5 mm) using lung-specific reconstruction filters *is always preferred* for quantitative analysis of lung lesions. Regarding slice thickness, optimum data were obtained if the slice thickness was at least half of lesion diameter. The lesion size in her data ranged from 1 cm to 8 cm, with an average of 4 cm. Data were not significantly degraded with minimal slice reconstructions at 1.25 mm. Of course, reconstruction slice thickness has to be balanced against radiologists' preferences, but an outcome of these data would suggest that *image reconstructions should be performed at least twice*: once for quantitative (computer) analysis and again at lower resolution for qualitative (radiologist) analysis.

Nonetheless, with unfiltered data, significant relationships between image features and gene expression or outcome were identified. **Olivier Gaevert** and **Sandy Napel** (Stanford) showed results from local and public domain data using advanced clustering to

identify CT features that were significantly related to survival and gene expression, specifically hypoxia- and ras-related gene clusters. Furthermore, they analyzed 17 advanced features from PET images of NSCLC to infer relationships to cell cycle, ECM and NFkB-related gene signatures. These data are consistent with the hypothesis that image feature data can represent underlying genomic drivers. **Matthew Schabath** (Moffitt) presented a highly curated data set of NSCLC patients with microarray gene expression data and sequencing data for 4 known driver mutations for *EGFR*, *KRAS*, *STK11/LKB1*, and *p53*. These data have been used to inform a prognostic model and the hypothesis/challenge was put forth that addition of radiomics data to this data set may improve prognostic power. Notably, patients for this study were not chosen based on the quality of their CT data, so a highly heterogeneous group of acquisition and reconstruction algorithms is expected. **Olya Grove** (Moffitt) presented some interesting data using an unfiltered data set wherein she developed new quantitative features derived from radiologically relevant semantic features related to density and spiculation. With these new features, even using unfiltered data, she was able to predict 2-year survival in adenocarcinomas with $p<0.005$. Further reductions in p-value are possible and these data remain to be tested against a prospective set, but they are nonetheless compelling.

With regard to feature selection, **Yoga Balagurunathan** (Moffitt) presented an analysis of test-retest data that was able to prioritize features according to their reproducibility, their biological dynamic range and their independence. These analyses used concordance correlation coefficients, CCC, and linear dependency matrixes (measured by coefficient of determination). These were analyzed on images that were either manually or automatically segmented. Depending on the cutoffs used, 219 features could be reduced to anywhere from 6 to 79 that had, respectively, extreme or acceptable levels of reproducibility and independence. These image features were then used to discriminate the population based on clinical and histological priors (~16 different ways). Accuracy in survival prediction was close to 82 % using best three image features, and this improved to 84% using gene expression. A similar Test-Retest analysis was presented by **Emmanuel Rios** (MAASTRO) who performed this pre-clinically in rat tumors acquired at 0.6 and 1 mm slice thicknesses. Analyses and results were similar to those from human data presented by Moffitt. With stringent (interclass correlation >0.85) to acceptable (ICC > 0.7) statistical cutoffs for reproducibility, the feature list was reduced to between 33% and 50%, respectively, of its original number. Furthermore, in these studies, analyses of features were conducted in hypoxic and hyperoxic rats, and showed that some textural feature changes could be highly correlated to oxygenations status.

Anant Madabhushi (Case Western) was new to the radiomics group and was a special invitee who provided an overview of the efforts in his center at correlating and quantitatively integrating multi-modal, multi-scale imaging and molecular information for disease prognosis and diagnosis. In these studies, radiological images (macro scale) were spatially registered and correlated to gross (meso) and histo (micro) pathology, and underlying gene or protein (nano) expression. A combination of machine learning and pattern recognition approaches have been developed to test the hypothesis that the radiologically and pathologically visibly "phenome/anatome" reflects underlying molecular biology. This hypothesis is central to the radiomics enterprise which seeks to

infer underlying molecular causes using profound analyses of radiological features. A large array of work was presented, including prostate cancer, from which gross specimens are available following prostatectomy. Computer extracted multi-parametric MRI features can be used to inform grading and the presence of highly malignant disease. In breast cancer, dynamic radiological (MRI) image features alone were able to predict the underlying receptor status of the lesion (e.g. ER, PR, Her2) and sophisticated computational analysis of standard H&E histopathologic specimens of ER+ breast cancers was able to predict "Oncotype" score; a prognostic measure of gene expression patterns used clinically to determine whether a patient with ER+ breast cancer is a candidate for adjuvant chemotherapy or hormonal therapy alone. These focused analyses can benefit from input of more data sets that could become available through data sharing.

DATA SHARING

A significant amount of time was devoted to discussing data sharing issues. Relatively unfettered access to large data sets against which algorithms can be tested has great potential to stimulate and catalyze progress in this field. Three inter-related activities were described that are in various stages of implementation. **Jayshree Kalpathy-Cramer** (Data Farber) and **Dmitry Goldgof** (Moffitt) co-presented a paper describing the establishment of a "Lung Tumor Segmentation Challenge." which represents a significantly coordinated activity between QIN and QIBA image analysis and performance metrics working groups. In this ambitious proposal, three alpha sites (Moffitt, Columbia, Stanford) are providing clinical and phantom images and segmentation algorithms, with additional data being downloaded from the LIDC. Kalpathy-Cramer's group, in collaboration with TCIA, has developed a series of metrics and software tools (Target Contour Testing and Instructional Computing Software (TaCTICS)) that evaluate the reproducibility and performance of segmentation algorithms, and that will be used to analyze the segmentations provided by the participants. This enterprise will establish the ground rules and accessibility for segmentation competitions, using NSCLC as the initial example. In the future, it is envisioned that this can be expanded to other cancers and other imaging platforms. A near-term goal is to identify stable funding sources for this enterprise. **Anders Berglund** (Moffitt) unveiled a "Radiomics Database" that can be used to house local instances of images, image metadata (including features), and medical outcomes data. This will be provided free of charge to any interested investigators with the expectation that such a database will provide common ontologies for sharing data across networks. **Andre Dekker** (MAASTRO) has been working on data sharing issues throughout Europe and Globally for the last decade and has been instrumental in developing EuroCAT, a data and image sharing platform for Radiotherapy planning. In his experience, the barriers to data sharing (in order of height) are: Administrative (incentivizing participation and expenditure of effort); Political (perceived value in data from individual institutions); ethical (privacy concerns); and, finally, technical. Different approaches are either centralized (wherein all participants deposit data to a central server for public access) or federated (wherein each participant houses their own instance of a generally queriable database). Centralized models have a longer history, with some notable successes (e.g. protein and genomics) and some that have been less successful (e.g. caBIG). Although

the technical barriers are lowest for centralized databases, all of the other barriers are high. Federated models have a shorter history and are technically challenging, yet the other barriers to participation are lower. After a lengthy discussion, there was consensus that a finite number of data elements (i.e. images, metadata and outcomes) are amenable to centralized repositories, whereas complete data lists (including medical demographics, features, etc) may be better suited to federated systems.

RECOMMENDATIONS

- This meeting began the convergence around a set of guidelines for CT image acquisition and processing for radiomics analysis. Two acceptable methods are to reconstruct at minimum slice thickness with lung specific kernels or to store the raw intermediate data (along with software versioning). This needs to go forward as a formal proposal to QIBA.
- Classifier modeling has begun to show incremental and sometimes spectacular results and this activity needs to continue. A number of highly curated data sets remain to be mined. Making these available to the network will catalyze and stimulate progress. Although not formally discussed, providing data directly to collaborators through MTAs should involve co-authorship. Once deposited into centralized databases, authorship is not appropriate.
- Prospective screening trials based on the Vanderbilt model should be exported and radiomic analyses incorporated into the workflow.
- Prospective Integration of radiomics databasing and analysis into the workflow for cancer diagnosis is technically feasible and could be the subject of an Academic Industrial Partnership proposal.
- A number of funding opportunities for data sharing and informatics are available and should be explored. These include Academic Industrial Partnerships R01; Informatics U01, U24 or U01/R01 supplements. These should be pursued with focused and well defined proposals. In particular the informatics proposals will require specific relevance to biology.
- The QIN U01 RFA is still in effect and a number of related projects were identified that would be appropriate for this mechanism.
- Although it was not discussed in a formal setting, the RedCap (Research Electronic Data Capture) data manager was discussed as a potential partner in these endeavors. Both the lung EDRN group at Vanderbilt and Stanford are using this platform to host databases.