

Title

Cancer Imaging Phenomics Toolkit (CaPTk):

A Quantitative Imaging Analysis Tool for Feature Extraction and Predictive Modelling of Clinical Outcomes

Contact information

Presenter 1: [Spyridon Bakas, Ph.D. \(sbakas@upenn.edu\)](mailto:sbakas@upenn.edu) Primary Scientific Coordinator of CaPTk
Presenter 2: [Sarthak Pati, M.Sc. \(sarthak.pati@uphs.upenn.edu\)](mailto:sarthak.pati@uphs.upenn.edu) Lead Developer of CaPTk
Affiliation: Center for Biomedical Image Computing and Analytics (CBICA), Perelman School of Medicine, University of Pennsylvania
Address: Richards Medical Research Laboratories, Floor 7, 3700 Hamilton Walk, Philadelphia, PA 19104
Website: www.cbica.upenn.edu/captk

Outline

Flash drives with the software, sample data, documentation & notes, will be provided to all attendees

CaPTk for General Purpose Quantitative Image Phenotypic (QIP) Analysis (75 minutes)

- Introduction to QIP Analysis: Radiomics, Radiogenomics, Machine Learning (15 minutes – S.Bakas)
- Walk-through graphical and command-line interfaces (15 minutes – S.Bakas)
- Pre-processing algorithms and their necessity (15 minutes – S.Bakas)
 - Bias correction, denoising, co-registration, skull-stripping, intensity normalization.
- Segmentation: (30 minutes – S.Pati)
 - Semi-Automatic (ITK-SNAP)
 - Interactive Learning (based on SVM and Geodesic distance transform)
 - Fully Automatic (enumerating applications, as these will be covered in the next part)
 - Pretrained deep learning models for multi-label Glioblastoma Tumor Segmentation
 - Pretrained deep learning model for skull-stripping
 - Breast Tissue (fatty and dense) Segmentation (LIBRA)
 - Lung Nodule segmentation (SBRT)

CaPTk for Specific Cancer Types (55 minutes – S.Bakas)

- Glioblastoma: (30 minutes)
 - Radiogenomic signatures
 - Predicting regions of tumor recurrence
 - Predicting patient survival in 6- and 12-month periods
- Breast Cancer: (15 minutes)
 - Fully automated estimation and visualization of percentage density, which is a clinically relevant measure for breast cancer prediction [4]
- Lung Cancer (10 minutes)
 - Prediction of survival and nodal failure risks [9]

Contributing to (and Extending) CaPTk (50 minutes – S.Pati)

- Incorporating Standalone algorithms/apps written in any language, e.g. MATLAB, Python.
- Source code integration of algorithms written in C++
- Contribution to Feature Extraction

Relevance to ISBI Audience

We would like to provide attendees with the opportunity for a hands-on tutorial using clear and focused material on a lightweight, yet powerful, software platform. The Cancer Imaging Phenomics Toolkit (CaPTk, <http://www.cbica.upenn.edu/captk>), is an open-source toolkit that is freely distributed and incorporates algorithms essential for quantitative biomedical image analysis towards covering emerging topics in biomedical imaging. Aligned with ISBI's scope to foster the exchange of knowledge among different imaging

communities and contribute to a unifying approach to biomedical imaging. CaPTk targets both computational imaging scientists (i.e., the ISBI audience) and non-computational experts (e.g. clinicians, neuroscientists). Computational scientists and technical personnel of all levels (e.g., data analysts, graduate students, post-docs) can i) construct their own end-to-end pipelines for constructing holistic studies based on the algorithmic building blocks provided by CaPTk (pre-processing, radiomic feature extraction, machine learning) for use in high throughput imaging data (batch processing), as well as ii) incorporate their algorithms into CaPTk's Graphical User Interface (GUI) that is based on Qt, ITK, VTK, and OpenCV and therefore takes advantage of the easy-to-use interface and CaPTk's user base to gain visibility/citations of their algorithms. Furthermore, CaPTk targets the non-computational experts (e.g., clinical researchers) by enabling immediate use of complex algorithms for clinically relevant studies through a user-friendly, platform-independent interface.

Short description of the material to be covered

CaPTk is a software platform to facilitate radiographic cancer image analysis, with a current focus on brain, breast, and lung cancer. CaPTk integrates advanced, validated tools and practices [8] to perform various aspects of biomedical image analysis, that have been developed in the context of active clinical research studies and collaborations that are geared towards addressing real clinical needs. With emphasis given in its use as a lightweight viewer, and with no substantial prerequisites to requiring a computational background, CaPTk aims to enable the rapid translation of cutting-edge computational algorithms into a routine clinical decision making and reporting workflow. Its long-term goal is providing widely used technology that leverages the value of advanced imaging analytics in cancer prediction, diagnosis and prognosis, as well as in better understanding the biological mechanisms of cancer development.

We intend to have a two-part tutorial; 1) starting from the use of CaPTk for general purpose QIP analysis, 2) focusing into currently implemented algorithms in CaPTk for specific cancer types (brain, breast, and lung), and provide details on the technical backend and how a user can incorporate their own algorithm into CaPTk front end. The 1st part, presented by Spyridon Bakas, the Lead Scientific Coordinator of CaPTk, shall involve participants getting familiar with the provided sample data and specific case studies in order to guide them towards using the incorporated algorithms through both the Graphical User (GUI) and the Command-line (CLI) Interfaces. The sample data will comprise of anonymized i) multimodal Magnetic Resonance Imaging (MRI) volumes of patients diagnosed with glioblastoma and ii) full-field digital mammography (FFDM) images of breast cancer patients. We will attempt to offer a comprehensive analysis pipeline, starting immediately after downloading multi-file DICOM files from a medical PACS and covering i) NIfTI conversion, while keeping the DICOM header information, ii) pre-processing algorithms, including careful correction for magnetic field inhomogeneities, denoising, co-registration for concurrently assessing voxels of multimodal scans, and skull-stripping, iii) tissue segmentation, including segmentation of the various heterogeneous glioma sub-regions (through ITK-SNAP [1], GLISTRboost [2], DeepMedic [3] and Geodesic Distance based algorithms), and breast tissue segmentation and density estimation (through LIBRA [4]), iv) (radiomic) feature extraction and parameterisation, including morphology, intensity, histogram-based, and texture, v) train and apply customized machine learning models using the extracted features, as well as vi) specialized predictive modelling tools for glioblastoma, including prediction of survival [5] and recurrence [6], and a radiogenomic biomarker of EGFRvIII [7]. To showcase the potential clinical value of specialized predictive modelling, an example case study is shown below (Fig.1) highlighting the accurate prediction of tumor recurrence [6] that can potentially directly influence radiotherapy by refined personalized dose escalation planning on radiation regimens.

During the last part the Lead Developer of CaPTk (Sarthak Pati - Presenter 2) will present the various ways that computational imaging scientists (i.e., the ISBI audience) can incorporate their existing algorithms (available as either C++ classes, or standalone applications written in MATLAB/Python) into the CaPTk's GUI, allowing for direct use of the easy-to-use and lightweight interface, enabling use of their advanced algorithm by non-computational experts, e.g. clinicians, thereby leading to improved visibility/citations of their tools from others. In addition, possible ways to contribute to the feature extraction pipeline will be discussed.

Example case of Predictive Modelling of Recurrence that can potentially influence clinical practice

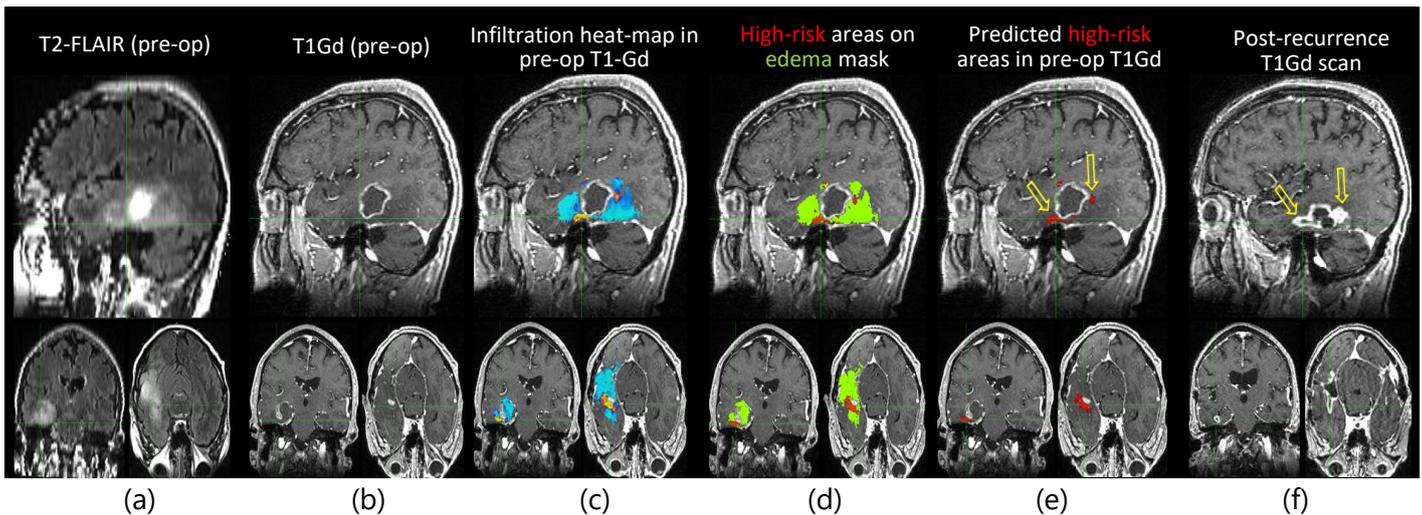


Fig.1. (a) and (b) show the pre-operative scans. (c) shows the infiltration heatmap of the predicted probability of recurrence overlaid on the pre-operative post-contrast (T1Gd) scan. Blue areas suggest lower risk of recurrence, while yellow and red areas suggest higher risk. (d) shows the "edema" in green representing the T2-FLAIR envelope overlaid on the T1Gd scan and the high-risk regions in red as a subset of the T2-FLAIR envelope. (e) shows the high-risk areas on the pre-operative T1Gd scan. (f) shows nodular enhancement (in the areas predicted as high-risk in (e)) around the cavity several months later. These nodular enhancements were resected and pathologically-proven to be actual tumor recurrence.

References

1. P.A.Yushkevich, *et al.* "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability". *Neuroimage* 2006 Jul 1;31(3):1116-28
2. S.Bakas, *et al.* "GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation", *Brainlesion, Springer LNCS* 2016;9556:144-55. DOI: 10.1007/978-3-319-30858-6_13
3. K.Kamnitsas, *et al.* "Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation", *Medical Image Analysis* 2016;36:61-78
4. B.M.Keller, *et al.* "Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation," *Medical Physics* 2012;39(8):4903-4917
5. L.Macyszyn, *et al.* "Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques", *Neuro Oncology* 2016;18(3):417-25
6. H.Akbari, *et al.* "Imaging Surrogates of Infiltration Obtained Via Multiparametric Imaging Pattern Analysis Predict Subsequent Location of Recurrence of Glioblastoma", *Neurosurgery* 2016;78(4):572-580
7. S.Bakas, *et al.* "In vivo detection of EGFRvIII in glioblastoma via perfusion magnetic resonance imaging signature consistent with deep peritumoral infiltration: the ϕ -index", *Clinical Cancer Research* 2017;23(16):4724-34.
8. A.Zwanenburg, *et al.* "Image Biomarker Standardisation Initiative", arXiv:1612.07003
9. H.Li, *et al.* "Predicting treatment response and survival of early-stage non-small cell lung cancer patients treated with stereotactic body radiation therapy using unsupervised two-way clustering of radiomic features", *Int. Workshop on Pulmonary Imaging*, 2017

Description of the expected audience, especially prerequisite knowledge & skills

The proposed hands-on tutorial will be tailored to target the ISBI audience. Computational imaging scientists of all levels (e.g., graduate students, post-docs), interested in taking advantage of an existing platform-independent, easy-to-use, light-weight GUI, towards having algorithms used by people without computational background (e.g., clinicians), as well as by computational analysts towards processing large batches of data.

Prerequisites: Intermediate knowledge of programming and command line usage experience. Use of CMake and ITK will be beneficial.

Description of the course packs (including slides and other tutorial material) that will be provided to the audience in hardcopy and/or electronic format.

The organizers will provide all the slides, sample data, source code, and binaries beforehand through a webpage dedicated to the material for the proposed tutorial. Furthermore, the organizers will provide flash drives with all this material during the tutorial. Note that all the provided material will be free from copyright issues.

Detailed Description of provided material:

- Source code of CaPTK's current release with Git repository information
- Installer packages (including dependencies) for Windows, Linux and macOS
- Sample data of 3 glioblastoma patients, including multi-parametric MRI data, overall survival information, and predicted areas of recurrence.
- Sample data of 3 breast cancer patients, including digital mammography images.
- Sample data of an extracted feature set of sample images
- Slides of the complete hands-on tutorial (i.e., Overview, Part 1, Part 2, Part 3) including the discussion notes.