The Cancer imaging Phenomics Toolkit (CaPTk)

Christos Davatzikos, on behalf of the team
Participating PIs

Computational Cancer Imaging R01 (NCI; Kontos, Conant, Schnall, Weinstein)

Computational Neuro-oncology R01 (NINDS; Davatzikos and O'Rourke)

Image Segmentation and Analysis R01 (NBIB; Yushkevich)

Kinetic Modeling P01 (NCI; Mankoff, Doot, Pryma, Fan)

Brain Connectomics R01 (NINDS; Verma and Brem)

Big Data Imaging Biostatistics R01 (NINDS; Shinohara)
Two Major Goals

• To leverage a rich family of advanced image computing algorithms

• To leverage extensive and long-standing collaborations with clinical teams who have provided input in the development of the algorithms, as well as data for training and validation of models
Feature Synthesis and Integration via Machine Learning

First Level

Image Analysis Algorithms

Segmentation:
Regions of Interest (ROIs)

Registration:
- Measure change with time:
A \rightarrow B

Image Harmonization:
Before \rightarrow After

Feature Extraction:
- Texture, histogram, dynamics, spatial pattern
- Wavelet-CP1

Image Operations:
- DICOM access
- Format conversion
- Intensity normalization
- Co-registration
- Noise Reduction
- ROI annotation

Second Level

Output Modules and Outcomes

Precision Diagnosis, Risk Estimation:
Breast Density Factor

Predictive models:
Kaplan-Meier Estimator

Personalized Treatment:
Connectomics, Infiltration

Radiogenomics
Imaging signatures of molecular characteristics:
- Breast MRI Phenotypes vs. Oncotype DX
- Imaging Signatures of GBM mutations

CaPTk Radiomic Panel

Open-CV

ITK
Feature Synthesis and Integration via Machine Learning

First Level

- Segmentation: Regions of Interest (ROIs)
- Registration: Measure change with time
- Image Harmonization: Before vs. After
- Feature Extraction: Texture, histogram, dynamics, spatial pattern
- Image Operations:
  - DICOM access
  - Format conversion
  - Intensity normalization
  - Co-registration
  - Noise Reduction
  - ROI annotation

Second Level

- CaPTk Radiomic Panel
- Open-CV

Output Modules and Outcomes

- Precision Diagnosis, Risk Estimation: Breast Density Factor, Kaplan-Meier Estimator
- Predictive models: EGFRvIII+, EGFRvIII-
- Personalized Treatment: Radiogenomics
- Imaging signatures of molecular characteristics: Breast MRI Phenotypes vs. Oncotype DX, Imaging Signatures of GBM mutations

Image Analysis Algorithms

Input Images
ITK-SNAP

• Support for color, multi-channel, and time-variant images

• Segmentation done using Random Forest and Level Sets algorithms

• Transfer of data between ITK-SNAP and CaPTk is seamless – giving users the option to use the former’s segmentation and user interaction functionality with latter’s computation capabilities

Screenshot of ITK-SNAP’s interface

Breast Segmentation

- Fully-automated segmentation of the breast area and the dense breast tissue, estimation of **Percent Density** (PD%)

- Adaptive fuzzy-c-mean (FCM) clustering based on intensity histogram and acquisition parameters

Initialize seed points for the MICCAI BRATS 2015 award-winning GLISTRboost method using CaPTk Console

Image Processing Portal
(ipp.cbica.upenn.edu)

- Generative segmentation
- Discriminative population based refinements
- Bayesian personalized refinements
Web Portal for Compute-heavy CaPTk Functions: CBICA cluster via high-end NIH instrumentation grant:

- **General-purpose compute nodes**
  - 175 Servers
  - 3500 CPU-cores
  - 44 TB RAM

- **Storage & Network**
  - 250TB Disk Array
  - 10&40Gb-s Ethernet

- **High RAM nodes**
  - 4 Servers
  - 96 CPU-cores
  - 16 TB-RAM

- **GPU nodes**
  - 125 Servers
  - 2500 CPU-cores
  - 31TB TB RAM
  - 125 GPUs
  - 5K GPU-cores/server
Advances in tractography enhance neurosurgical planning, but are limited by edema, mass effect, and tract infiltration.

Confetti facilitates automated identification of all tracts (including eloquent tracts) without manual drawing of ROIs, making the tracts robust and replicable.

Confetti enhances the primary objective of neurosurgery: maximal safe resection in the presence of tumor edema and infiltration.

Confetti interface is dedicated for fiber tract visualization and it is integrated with CaPTk Console.
Feature Synthesis and Integration via Machine Learning

Second Level

Radiogenomics: Imaging signatures of molecular characteristics
- Imaging Signatures of GBM mutations
- Breast MRI Phenotypes vs. Oncotype DX
- Imaging Signatures of GBM mutations

Output Modules and Outcomes

Precision Diagnosis, Risk Estimation:
- Breast Density Factor
- Kaplan-Meier Estimator

Predictive models:
- Connectomics vs. Oncotype DX
- Infiltration

Personalized Treatment:
- Radiogenomics

Input Images

Image Analysis Algorithms

Segmentation:
- Regions of Interest (ROIs)

Registration:
- Measure change with time:
  - A
  - B
  - B → A

Image Harmonization:
- Texture, histogram, dynamics, spatial pattern

Feature Extraction:
- Wavelet-DP1

Image Operations:
- DICOM access
- Format conversion
- Intensity normalization
- Co-registration
- Noise Reduction
- ROI annotation

First Level

CaPTk Radiomic Panel

ITK

Open-CV
CaPTk Radiomic Panel

- Segmentations: volumes and signals within ROIs
- Normalized histograms of different protocols; optimized binning
- Kinetics
- Textures
- Spatial patterns/distributions
- Parametric maps from longitudinal scans
Input Images

First Level

Image Analysis Algorithms

Segmentation:
- Regions of Interest (ROIs)

Registration:
- Measure change with time:

Image Harmonization:
- Before
- After

Feature Extraction:
- Texture, histogram, dynamics, spatial pattern

Image Operations:
- DICOM access
- Format conversion
- Intensity normalization
- Co-registration
- Noise Reduction
- ROI annotation

CaPTk Radiomic Panel

ITK

Feature Synthesis and Integration via Machine Learning

Open-CV

Second Level

Output Modules and Outcomes

Precision Diagnosis, Risk Estimation:
- Breast Density Factor
- Kaplan-Meier Estimator

Predictive models:
- Breast MRI Phenotypes vs. Oncotype DX
- Connectomics

Personalized Treatment:
- Infiltration

Radiogenomics
- Imaging signatures of molecular characteristics:

Breast MRI Phenotypes
- EGFRvIII+
- EGFRvIII-
Computational Neuro-Oncology

Imaging Signatures of Molecular Characteristics

Predictors of clinical outcome

Predicting Survival:

HR(low & high): 10.64 (95% CI 5.9-19.3, p<0.001)
HR(med & high): 3.88 (95% CI 2.3-6.6, p<0.001)
HR(low & med): 2.77 (95% CI 1.8-4.2, p<0.001)

Predicting Infiltration and Recurrence

Bakas et al., Clinical Cancer Research 2017

Akbari et al., Radiology, 2014
Akbari et al., Neurosurgery, 2016
Brain Connectomics

Peri-lesional Effects of GBMs

- Automated atlas-based tract extraction (using connectivity signatures instead of shape help address mass effect)
- Edema invariant tractography (using multicompartment models fitted to multishell imaging)
- Tumor connectome (effect of tumor on distant regions, regional vulnerability and functional rerouting)

Web-based Integrated Surgical Planning Environment

Global Effects of GBMs
Radiomic Breast Cancer Phenotypes

Breast Cancer Phenotyping via Imaging:

• Segmentation and multi-parametric feature extraction
• Identification of intrinsic phenotype patterns
• Prognostication and treatment response prediction
Software Architecture Overview

CaPTk GUI

Algorithmic Applications

- **Neuro** (predictive modelling, surgical planning, etc.)
- **Torso** (tissue density estimation, automatic MS segmentation, etc.)
- **Utilities**
  - Feature Extraction
  - Smoothing
  - Registration
  - Bias correction

Low-level functions providing common operations for all algorithms as C++ objects

Third-party

- **ITK** (I/O, segmentation, registration, etc.)
- **OpenCV** (machine learning)
- **VTK** (rendering & visualization)
- **Qt** (user interaction)

CBICA

- **Parser**
- **System Utilities**
- **CWL converter** (future)

Access algorithms and utility functions as objects

CLI/GUI access directly to applications
Installation

- Only 64-bit machines supported due to processing requirement
- Installers are designed to work as double-click and install
- Windows and Linux installers are available and macOS to be ready soon
- The entire CaPTk (console and all applications) will also be available as docker images soon

<table>
<thead>
<tr>
<th>Operating System</th>
<th>Version</th>
<th>Installer System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows</td>
<td>7+</td>
<td>MSI</td>
</tr>
<tr>
<td>macOS</td>
<td>Snow Leopard+</td>
<td>DMG</td>
</tr>
<tr>
<td>Ubuntu</td>
<td>14.04+</td>
<td>DEB</td>
</tr>
<tr>
<td>centOS</td>
<td>6+</td>
<td>RPM</td>
</tr>
</tbody>
</table>
## Dependencies

### Core Application (written in C++ for efficiency):

<table>
<thead>
<tr>
<th>Library</th>
<th>Utility</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qt</td>
<td>Graphical Framework of application</td>
<td><a href="http://www.qt.io">www.qt.io</a></td>
</tr>
<tr>
<td>Visualization Toolkit (VTK)</td>
<td>Visualization of images, charts, etc.</td>
<td><a href="http://www.vtk.org">www.vtk.org</a></td>
</tr>
<tr>
<td>Insight Toolkit (ITK)</td>
<td>I/O, image processing, etc.</td>
<td><a href="http://www.itk.org">www.itk.org</a></td>
</tr>
<tr>
<td>OpenCV</td>
<td>Machine Learning</td>
<td><a href="http://www.opencv.org">www.opencv.org</a></td>
</tr>
</tbody>
</table>

### Different Components (can be C++ or Python):

<table>
<thead>
<tr>
<th>Library</th>
<th>Utility</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>numpy</td>
<td>Higher level math operations</td>
<td>numpy.org</td>
</tr>
<tr>
<td>pyqt</td>
<td>User interface of some applications</td>
<td>wiki.python.org/moin/PyQt</td>
</tr>
<tr>
<td>dipy</td>
<td>Diffusion Imaging</td>
<td>nipy.org/dipy</td>
</tr>
</tbody>
</table>