Radiomics for Breast Cancer

1. Pilot Overview

Goals: Improve treatment response for breast cancer by using AI to analyze mammograms, US, and MRI images along with structured clinical data. Reduce costs by tailoring treatment for the individual patient.

- Project lead
- Image analytics
- Clinical analytics
- Deep learning
- IBM
- Patient images and clinical data
- Clinical knowledge
- Data hosting
- VTT
- Image analysis
- Interpretable features

2. Radiomics for NACT Prediction

- Neoadjuvant Chemotherapy Treatment (NACT) option
  - Decision today is made based on clinical variables only
  - Less than half of treated patients achieve pathological complete response with no evidence of residual disease
  - Failed treatment worsens the patient prognosis
  - Failed treatment increases the cost

- Radiomics can improve NACT response prediction
  - Extract large amount of features from multi-modal medical images
  - Apply deep learning and computer vision algorithms for precision medicine

3. Heterogeneous Data Collection

- Patient data
  - Image analysis
  - Healthcare
  - Tumor

- Treatment
  - Chemotherapy, hormone therapy, radiation therapy

- Response to treatment
  - Treatment response: Complete, partial, stable, progressive

4. Curation and Anonymization

- Clinical Data
  - A cohort of ~1700 patients
  - Women with breast cancer who received NACT between 2012 - 2018
  - Use NLP algorithms to extract data from various reports
  - Anonymize PHI as age and dates

- Imaging Data
  - Multi-modal imaging
  - mammograms
  - Magnetic Resonance Images (MRIs)
  - Ultrasound (US) images

- Images are very valuable and include intrinsic information, but only ~400 patients have imaging

- All the images are anonymized

5. Data Statistics

- Important Clinical Features
  1. Age at diagnosis
  2. BMI
  3. Ki67 percent
  4. HER2 positive
  5. Progesterone status
  6. ER grade

- MRI Imaging Types
  - Type of scan
    - T1 (pre-contrast)
    - T2 (pre-contrast)
    - DCE
    - Diffusion weighted
    - Apparent diffusion coefficient
    - Percent signal reduction
    - T2 with DCE method
    - T1 with DCE method

- Total number of MRI scans: 12714
- Total number of MRI subtractions scans: 1055
- Total number of US scans: 2836
- Total number of US scans: 1086

6. Pilot Architecture

- Open Data
- Decentralized microservices
- Spoke/Hub Platform
- Image Analysis Framework
- Big Data Analytics Framework
- Clinical Analytics Framework
- Patient Feedback

- GPU-based Server
- SPARK Cluster

7. VTT Image Analysis Pipeline

- Pre-processing
  - Image registration
  - Image segmentation
  - Tissue classification
  - Cancer biomarkers

- Feature extraction
  - Image analysis
  - Power analysis
  - Calibration

8. IBM Algorithmic Building Blocks

- 2D Inception v4
- 2D CNN
- LSTM
- 3D CNN

- Multi-view CNN + LSTM + Global

- Local attention used
- Feature encoding and multi-modal decision

Contact: simona@il.ibm.com, efrahimn@il.ibm.com