Big Data for Medical Analytics

BIG MediLytics

Oncology

Breast cancer

Radiomics for Breast Cancer

S. Rabinovici-Cohen [1], A. Abutbul [1], K. Antila[3], E. Hexter [1], O. Hijano-Cubelos[2], A. Khateeb [1], J. Pajula[3], S. Perek [1]
[1]IBM Research – Haifa, [2]Institute Curie, [3]VTT Technical Research Centre

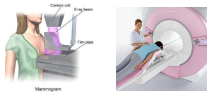
1. Pilot Overview

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.

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

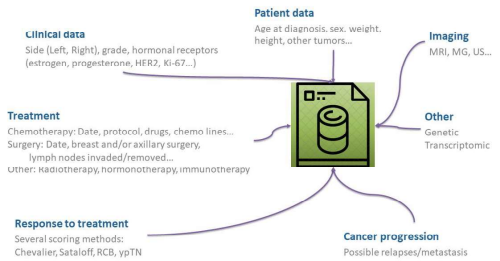
2. Radiomics for NACT Prediction

- 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



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

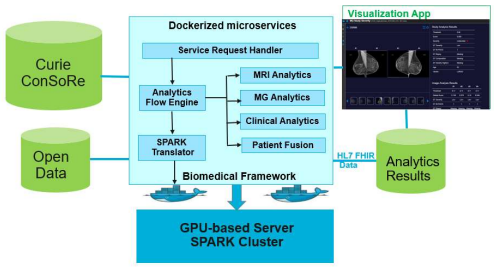
1. Age at diagnosis
2. BMI
3. KI67 percent
4. HER2 positive
5. Progesterone status
6. EE grade

MRI Imaging Types

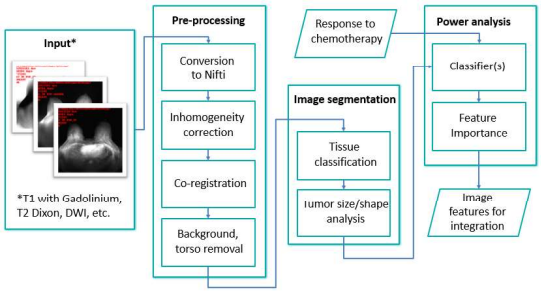
Type of Scan	Patients Number
T1 (any variant)	335
T2 (any variant)	335
Diffusion weighted	335
T1 with Gadolinium	46
Any contrast agent	284
T2 with Dixon method	151
Substraction	304

Total number of MRI scans: 12174
Total number of MRI substraction scans: 1055
Total number of MG scans: 2836
Total number of US scans: 1086

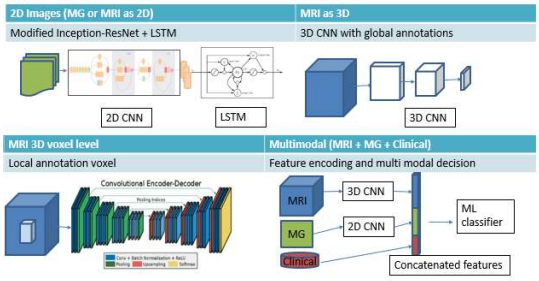
6. Pilot Architecture



7. VTT Image Analysis Pipeline



8. IBM Algorithmic Building Blocks



BIG DATA VALUE PUBLIC-PRIVATE PARTNERSHIP
This project is part of BDV PPP

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780495.

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