Deep Learned and Practical

Implementing Machine Learning Techniques in the Real World

Timor Kadir DPhil

Mirada Medical CTO
Optellum CTO
Visiting Fellow University of Oxford

Disclosures

- Mirada Medical
  - Employee
  - Shareholder
- Optellum
  - Employee
  - Shareholder

Overview

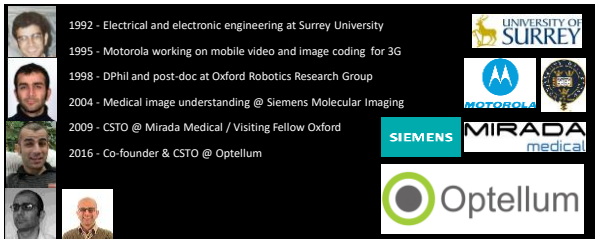
I – How do we bring new algorithms to the clinic?

II – Practical challenges

- Automatic OAR Segmentation (Mirada)
- Lung Nodule Risk Stratification (Optellum)
- Oxford SpineNet (Oxford University)

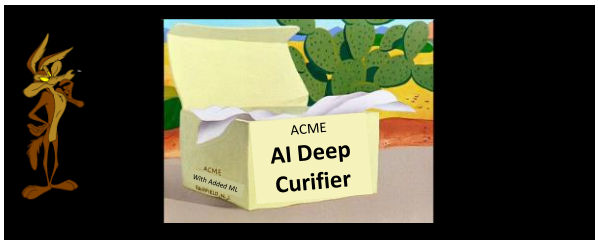
III - Perspectives: potential impact of ML in Radiation Therapy.

Timor Kadir

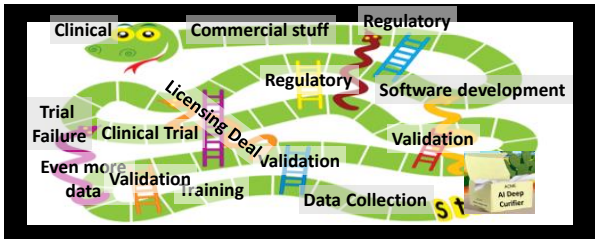


Medical Devices: From bench to bedside

Medical Devices: From bench to bedside



From bench to bedside

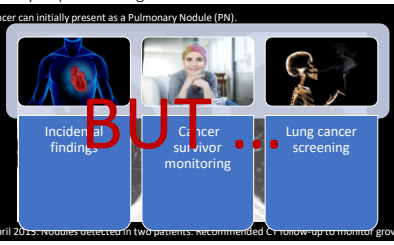


From bench to bedside

Data Collection	Lung nodule risk stratification
Training	Spinal MRI analysis
Validation	Automatic contouring

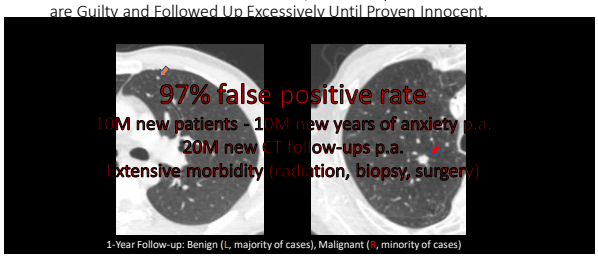
Early Detection by Chest Computed Tomography Significantly Improves Lung Cancer Survival

- Lung cancer can initially present as a Pulmonary Nodule (PN).

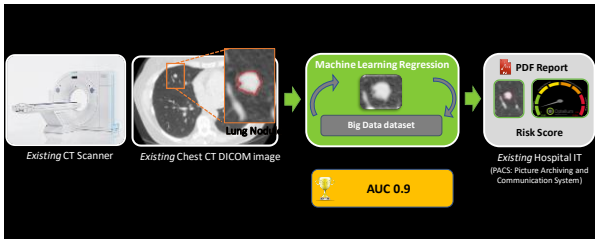


April 2015: Nodules detected in two patients. Recommended CT follow-up to monitor growth.

Problem: In Current Clinical Practice, Pulmonary Nodules are Guilty and Followed Up Excessively Until Proven Innocent.



Nodule Risk Stratification

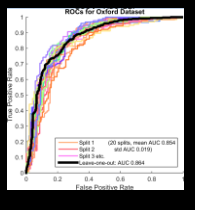


The ugly truth….

- Yes we won LungX.
- But the AUC was only ~0.65
- AUC on training data was >0.85
- Why the difference?
- Training Data

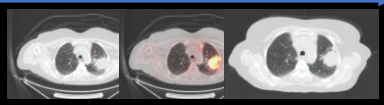
Look a little closer

- LungX
  - 10 "calibration" datasets with known ground-truth
  - 100 test set with unknown ground-truth
- We manually created a dataset from Oxford and LIDC data trying to match the characteristics of the LungX datasets.
- Bias – in LIDC most malignant lesions are large and the benign ones are small.



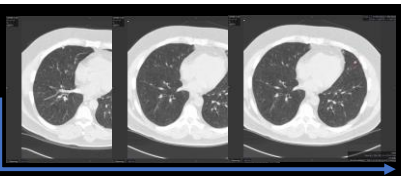
Dataset biases

Symptomatic



Biased towards large tumors, diagnostic CT, contrast, 0.625-1mm

Dataset biases



Asymptomatic

Biased towards small nodules, low dose CT, no contrast, 1-2mm

Timor's recipe #1 for ML success

- Acquire data that represents the typical distribution of disease, patient and imaging characteristics
- Test for built-in biases
- If you are going to use heterogeneous data for training, at least test with realistic data

From bench to bedside

Data Collection

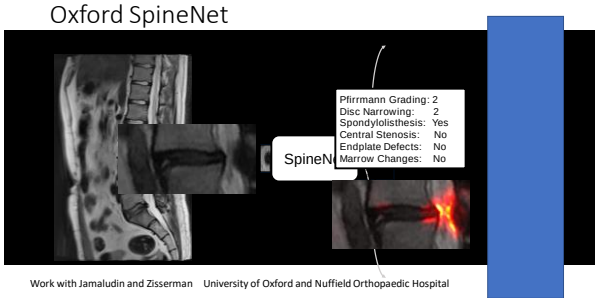
Training

Validation

From Computer Vision

- ImageNet 2017 Object Localisation Challenge
  - 1000 objects: 100k/50k training and test images
  - Winner used Res-Net with > 150 layers
- Strategies for reducing training requirements
  - Data augmentation
  - Transfer learning
  - Multi-task learning
  - Self-supervision

Oxford SpineNet



Work with Jamaludin and Zisserman University of Oxford and Nuffield Orthopaedic Hospital

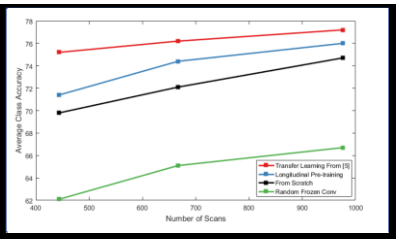
Oxford SpineNet

Data augmentation

Multi-task learning

Self-supervision

Transfer learning



Timor's recipe #2 for ML success

- Explore the utility of transfer learning, dataset augmentation, self-supervision and multi-task learning
- Other techniques are available: e.g. adversarial training
- Test robustly: if the distribution is not in the training data then it won't learn

From bench to bedside

Data Collection

Training

Validation

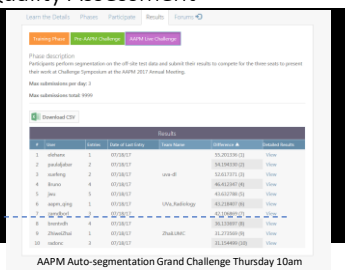
Auto-contouring Quality Assessment

Quantitative

Qualitative

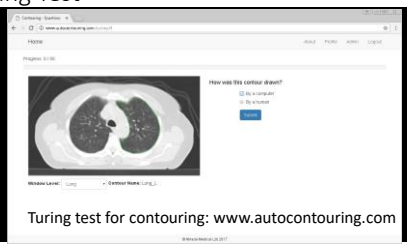
Turing Test

Atlas-based



AAPM Auto-segmentation Grand Challenge Thursday 10am

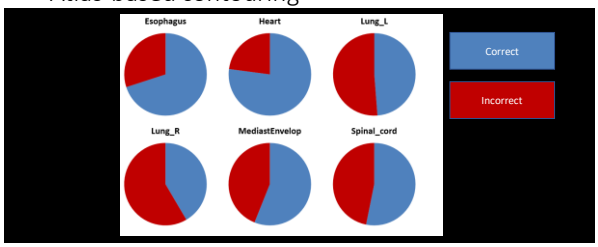
Turing Test



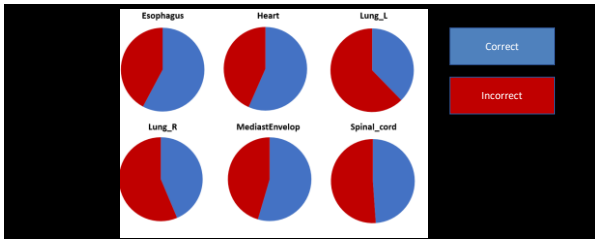
Turing test for contouring: www.autocontouring.com

Peressutti et al TU-FG-605-7

Atlas-based contouring



DLC



Timor's recipe #3 for ML success

- Conventional methods for validation should be utilized for your proposed method
- Consider Turing test type approaches
- Don't rely on one validation approach

Perspectives

Machine Learning is in its infancy in Radiation Therapy

Searched on: CNN, Machine Learning, Deep Learning, Convolutional
71 abstracts

Topics include:

Segmentation, Classification, Treatment Planning, DIR, Response assessment, Dose estimation, Image reconstruction…..

Perspectives

Data is critical but beware hidden biases

Lots of techniques now available to maximise power of datasets

Think broadly about validation protocols: Turing test