Assessing the Performance of Software in Measuring Tumor Change:

The Biochange Challenge Problem

Dr. Charles Fenimore

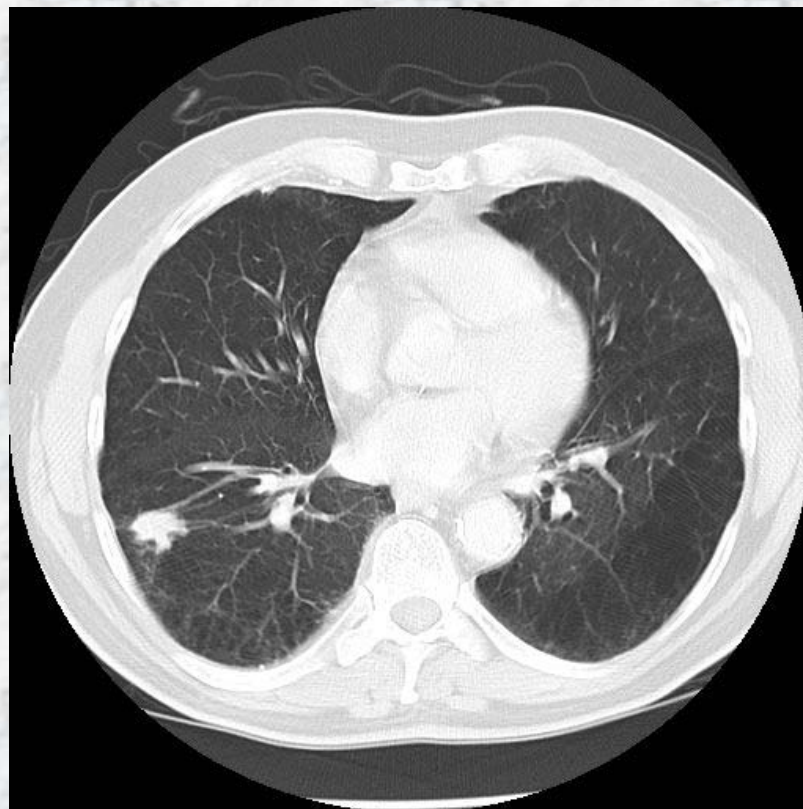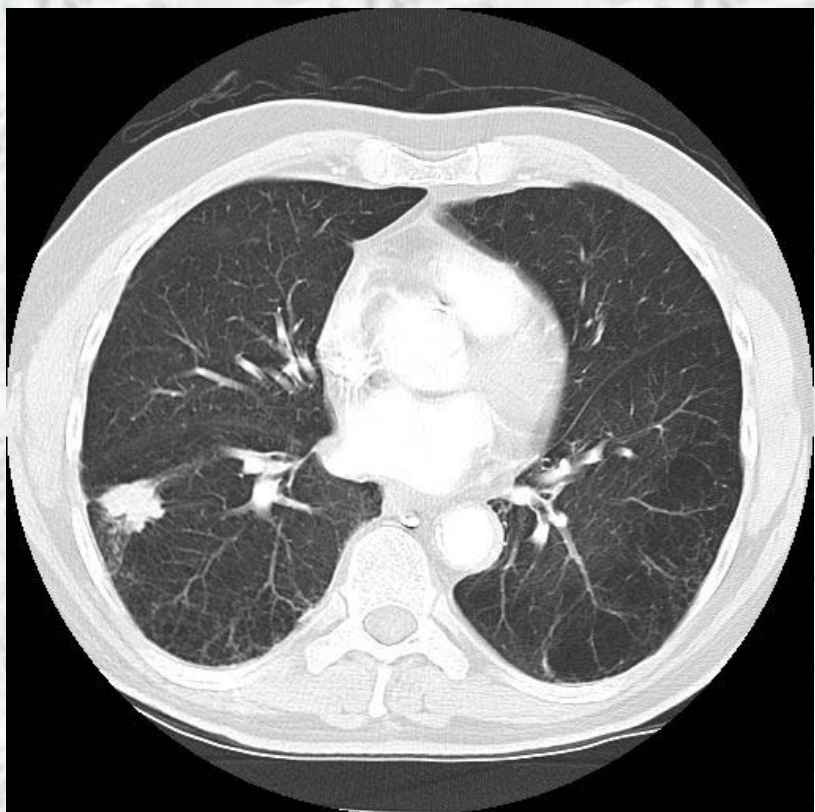Team: Brian Antonishek, Dr. John Lu, Dr. Jonathon Phillips

Biochange '08 Webpage

http://www.itl.nist.gov/iad/894.05/biochange2008/ Biochange2008-webpage.htm

NIST
National Institute of
Standards and Technology

Lesion size/change in Clinical CT

- Today: measuring change in medical images is unreliable.

  - Leading measure of lesion size in medical imagery is the (lineal) span. Lesion change is measured by comparing the spans.

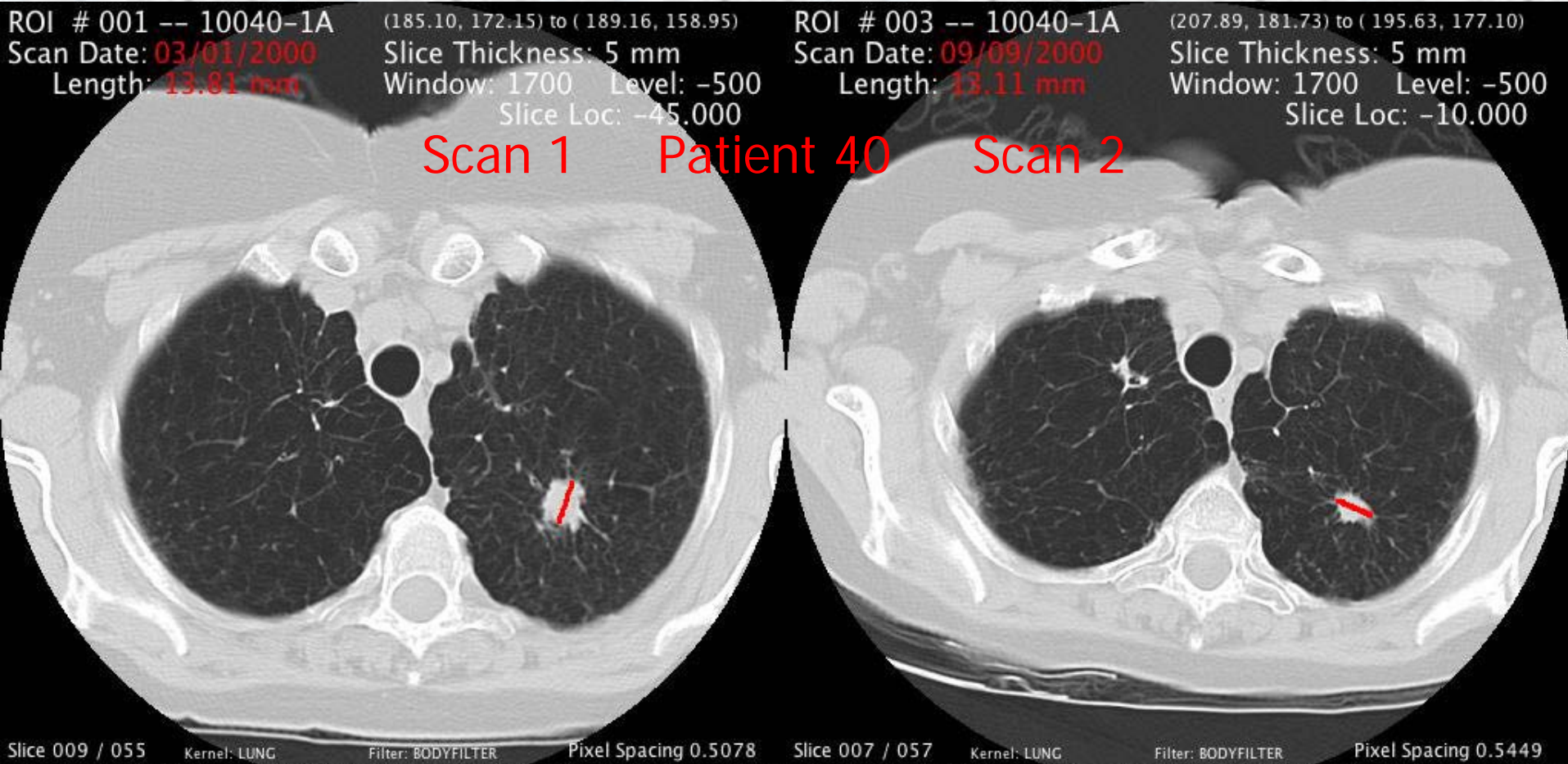- Future: measure change using volume, mass, or other biomarkers.

Scan 1   Patient 23   Scan 2

Lesion size/change hazards

- RECIST measure in RED. Change fraction $= L_2/L_1 - 1 = -5\%$

- Volume change (of white blob) consensus $= V_2/V_1 - 1 = -29\%$

- Disease progression or response to therapy can be masked by high levels of uncertainty.

ROI # 001 -- 10040-1A   (185.10, 172.15) to (189.16, 158.95)
Scan Date: 03/01/2000   Slice Thickness: 5 mm
Length: 13.81 mm   Window: 1700   Level: -500
                             Slice Loc: -45.000

ROI # 003 -- 10040-1A   (207.89, 181.73) to (195.63, 177.10)
Scan Date: 09/09/2000   Slice Thickness: 5 mm
Length: 13.11 mm   Window: 1700   Level: -500
                             Slice Loc: -10.000

Scan 1   Patient 40   Scan 2

Slice 009 / 055   Kernel: LUNG   Filter: BODYFILTER   Pixel Spacing 0.5078
Slice 007 / 057   Kernel: LUNG   Filter: BODYFILTER   Pixel Spacing 0.5449
1.3.6.1.4.1.9328.50.1.0040/ 1.3.6.1.4.1.9328.50.1.20329/ 1.3.6.1.4.1.9328.50.1.20366/ 1.3.6.1.4.1.9328.5 1.3.6.1.4.1.9328.50.1.0040/ 1.3.6.1.4.1.9328.50.1.20795/ 1.3.6.1.4.1.9328

Key goals of benchmarking

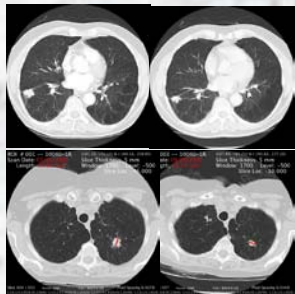Broad goal: Support the development of reliable software and methods for medical imaging measurements.

Specific problem: Assess methods for performance assessment of software – in this case for change analysis of lesions in lung CT.

NIST
National Institute of
Standards and Technology

4

Elements of Benchmarking for Change Analysis Algorithms

Evaluation Materials:

Image Collections by Academics, Government, Pharma, RSNA, etc,

Curation/metadata developed.

publicly available Image data set

Participants run their change analysis software or algorithms on data and report results to NIST

Performance Analysis

• Exploratory data analysis, analysis of sources of variation and bias: comparison between algorithms, comparison with markup.

• Big question: Is reliable ground truth available?

NIST
National Institute of Standards and Technology

Challenge Problems for Change Analysis Software & Algorithms

Benchmarking challenges: NIST & Cornell

> Biochange 2008 – Pilot study
>
> Volcano '09 – focus on lung cancer screening
>
> Biochange 2009 – response to therapy in stage 3 & 4 lung cancer.

Standards: work with algorithm developers and radiologists. NIST & other government, RSNA, ACR, . . .

Technology Applications

- Impact of high unreliability/uncertainty.

- Pharmaceutical trials are longer than they would be with accurate, reliable measures.

- Patient response to therapy is measured in units of weeks and months instead of days.

Commercial Applications

✛ Improved radiology process enhances:

- **drug trials – faster**
- **individual patient's response to therapy - measured in shorter times.**
- **more effective therapy**

Collaboration Opportunities

Download and run on the BC '08 Pilot Study data.

Participate in BC '09 when announced 10/09

- NIST provides access to CT lung data, each case has a pair of CT scans with identification of regions of interest.

- Participants run their change analysis software on the data & report results.

- NIST analyzes the performance of software / algorithms against markup.