Algorithm Validation Toolkit (AVT)

David Clunie
RadPharm, Inc.

RSNA 2008
December 4th 2008
Acknowledgements

• Robert Schwanke, Siemens Corporate Research
• John Pearson, Siemens Corporate Research
• Chenyang Xu, Siemens Corporate Research
• Lawrence Tarbox, Washington University
The Problem to be Solved

- Automated and semi-automated image analysis algorithms are being developed and promulgated to support lesion
  - detection and characterization
  - size quantification and change over time
- Also image analysis for supporting functions
  - registration and spatial transformation
- For a broad range of applications
  - including clinical use and for clinical trials
- How does one “validate” their correctness?
  - adequately
  - efficiently
• Many ongoing and new efforts to collect data, establish “truth” and define “methodology”
• LIDC
  • Lung Image Database Consortium
• RIDER
  • Reference Image Database to Evaluate Response
  • phantom images (FDA)
• NIST Biochange 2008
  • change measurement algorithm evaluation
• MICCAI
  • Segmentation challenge (liver, brain, coronary)
• QIBA
  • Quantitative Imaging Biomarkers Alliance (vol CT, DCE-MR, PET)
• …
AVT – One Component of a Solution

- A “toolkit” of software components, for
  - Accessing a collection of images
  - Gathering the “truth”
  - Gathering measurements, e.g., of “change”
  - Applying statistical methods to assessment of accuracy, precision, repeatability

- Persisting and making accessible data, measurements and results
- Applicable to human, semi-automated and automated measurement techniques
AVT – Example Use Case

- Validation of measurement of change in lesion size over time
- Ground truth at multiple time points (manual + consensus panel)
- Automated detection, measurement of size, computation of change
- MVT analysis of detection (ROC), consistency (Bland-Altman)
- Successive iterations of experiment with improvement of algorithm

Baseline
Max Diameter 36.2mm
Volume 6.1cm³

Baseline +20 weeks
Max Diameter 32.6mm
Volume 9.48cm³ 55% increase
AVT – Supports Broad Range of Use Cases

• Essentially limited only by
  • Data types
    • images, annotations, vocabulary
  • Source of truth
    • Image Analysis component
    • alternative source (convert to AIM and import)
  • Interfaces to algorithm under test
    • persistence method and form of output
    • orchestration of execution
  • Range of statistical methods provided by R in MVT
  • AVT support for orchestration of experimental paradigm
AVT – Use Case Variants

- **Modality**
  - single (CT), multiple (e.g., CT/PET, multispectral MR)

- **Acquisition**
  - One, multiple ("coffee break")

- **Pre-processing**
  - registration, segmentation, deformation, detection, propagation

- **Annotation**
  - manual, semi-automated, automated

- **Readers**
  - one, multiple, adjudicated, inter- and intra- variability

- **Measurement**
  - size (RECIST, volume), function (SUV)
  - change (two time points, multiple time points)
AVT – Role of First Users

- Strong emphasis on feedback from users throughout project
- Evaluation and testing of use-case model
  - walk through of use-cases
- Evaluation and testing of early deliverables
  - various modalities, measurements and paradigms
  - existing and new public image data sets
  - existing and new “ground truth” annotations
- Users involved
  - ACRIN, RadPharm, OHSU, UCLA, FDA CDRH, U Chicago, Cornell, U Michigan
AVT – Extend Existing Infrastructure

- Requirement to re-use and integrate existing tools
  - XIP (eXtensible Imaging Platform)
  - caGRID accessibility (grid data service)
  - National Cancer Image Archive (NCIA)
  - DICOM images
  - AIM annotations
  - caBIG vocabulary (Common Data Elements (CDE))
  - R statistical package
- Open source
  - components
  - deliverables
AVT – Initial Concept of Operations

IA: Image Analysis
• IA used to perform measurements and encode as AIM

AD: Assessment Database
• AD captures AIM & additional image & provenance data

MVT: Measurement Variability Tool
• MVT computes measurement statistics and correlations between metrics and outcomes
• **Image Analysis (IA):** An image analysis component that displays images and permits features to be identified, measured and marked and the results exported in the form of AIM data structures;

• **Assessment Database (AD):** An assessment database schema for storing the AIM objects and results produced by the first component (or equivalent functions);

• **Measurement Variability Tool (MVT):** Tools to extract measurements placed in the assessment database and compute their variability as a function of such variables as intra- and inter-rater, scanner, exam type, processing, time, software used.
AVT – Architecture Overview

IA -- Image Annotation Tool
- Algorithm Plug-ins
- Cached objects: AIM/DICOM, etc
- AD Wrapper

AE -- Algorithm Execution Tool
- Algorithm Plug-ins
- Cached objects: AIM/DICOM, etc
- AD Wrapper

MVT Measurement Variability Tool
- R LIB
- Cached objects: AIM/DICOM, etc
- AD Wrapper

XIP/AVT Workstation

Non-Grid Data Sources

Web Client

Images, Patient info Annotations, Collections Experiments

Modified XIP Host

AD Client

AD Server

AD as a caGrid data service (Phase III)

caGRID data services
AVT – Data Flow

AVT Phase II is a caGRID client ONLY

Assessment Database

Algorithm under test

Image Annotation

Algorithm Execution

Measurement Variability Tool

DICOM

AIM

DICOM

DICOM

AIM

AIM

XIP Host

Statistics

DICOM DS e.g. NCIA

AIM DS e.g. NCIA
Image Annotation Tool records seed point and estimates tumor boundary.
Dependency of the Prototype on Lesion Size & Slice Thickness

- Linear regression analysis on performance dependency on lesion size

The algorithm is relatively robust to lesion size, demonstrated by the small coefficient of determination in the linear regression analysis.

- Non-parametric analysis on slice thickness dependency

Due to the non-Gaussian distribution and variance inhomogeneity, Kruskal-Wallis test was applied to study the performance variation among different slice thickness.

Kruskal-Wallis analysis (non parametric testing) showed the algorithm is relatively robust to slice thickness.
Algorithm improvement over successive versions

Note:
- Red stands for the comparison between auto results and ground truth #1.
- Green stands for the comparison between auto results and ground truth #2.
- Purple stands for the comparison between ground truth #1 and #2.
XIP Applications designed as Scene Graphs
Can replace components with your own in visual editor

Statistics components written in R
AVT – Where are we?

- Initial RFP for Phase I Gap Analysis of requirements versus existing tools
  - 2007/04 – drafted
  - 2007/12 – completed
- Phase 2 Delivery of IA, AD and MVT components and framework
  - 2008/07 – awarded
  - In progress
  - 2009/06 - delivery
AVT – What it means to you now

- Success of all caBIG deliverables is predicated on wide spread adoption
- AVT is based on existing infrastructure
  - XIP, AIM and CDE
  - NCIA
  - caGRID tools and services
  - R statistics
- Evaluate and adopt these for
  - communication and persistence
  - visualization and annotation
  - analysis
- Then your project or algorithm will “plug in” to AVT!