The Center for Dynamic Data Analytics

cdda
center for dynamic data analytics

An NSF I/UCRC

Dimitris Metaxas
Director, CDDA
Rutgers University

"From Chaos to Knowledge"

What is the CDDA?

- The CDDA is an NSF-sponsored Industry/University Cooperative Research Center (I/UCRC)

- NSF I/UCRC Centers foster **pre-competitive** research collaborations between non-academic organizations and academia

- The CDDA is based on Rutgers University, Stony Brook University and 9 Industrial center members

Project, Member and Partner Summary

Security Domain and Troubleshooting Ontologies

Medical Imaging Analysis/Segmentation

3D Compression

3D Medical Imaging Analysis

Image Segmentation

Medical Imaging Analysis/Segmentation

Anomaly Detection

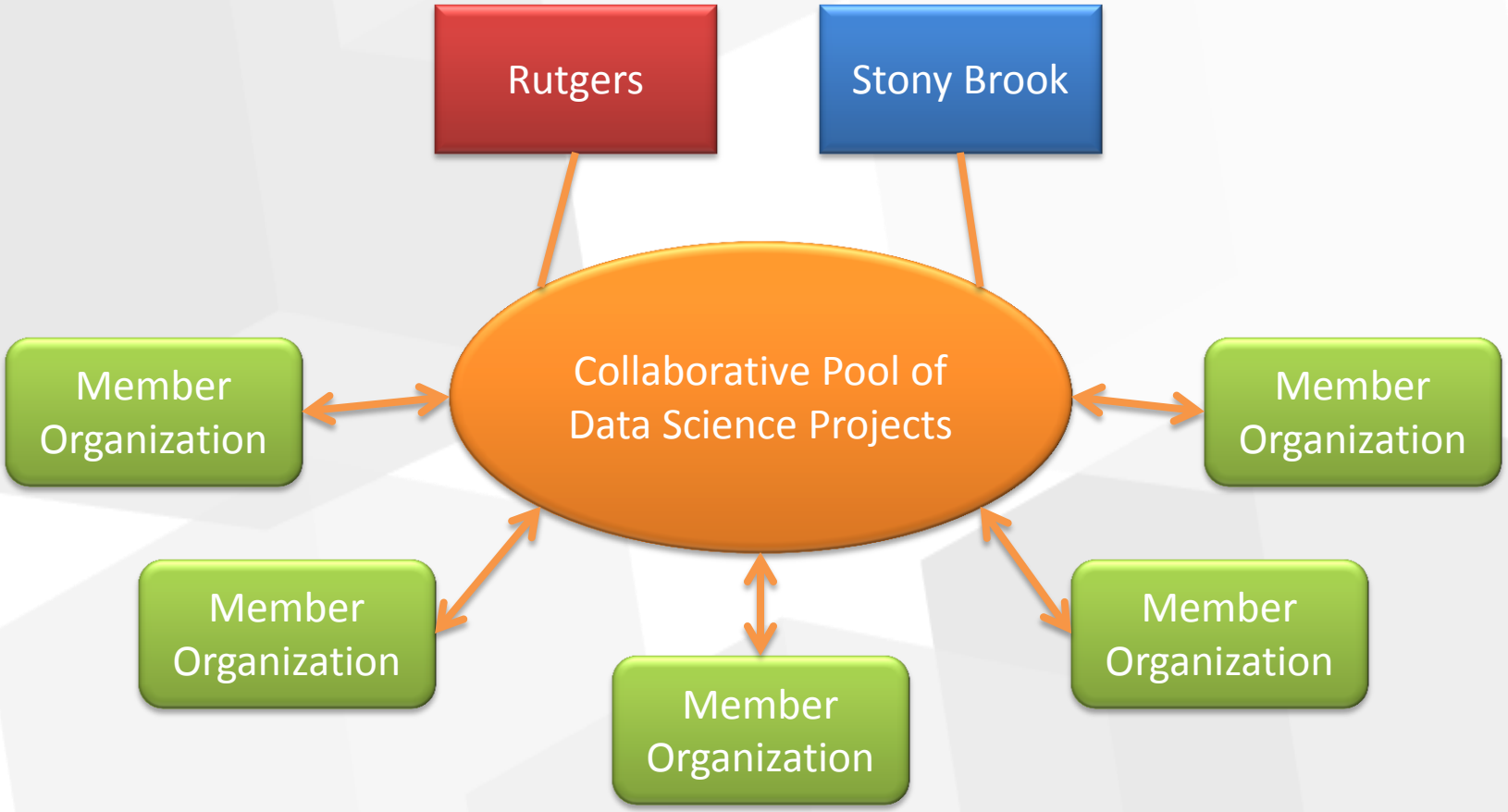
Electronic Medical Records Analysis

White House Big Data Initiative

Image Analytics

How does it work?

University, Project and Membership Relations



How does it work?

Center Operations

Fall

Spring

- There are semi-annual IAB Meetings (Spring and Fall) where the IAB, the Research Team and potential members gather to discuss ideas for new projects, the status of existing projects and the results of completed projects.


- Each University has a Center Director (Rutgers - Dimitris Metaxas , Stony Brook - Ari Kaufman)


- The Managing Directors (James Mielke and Rong Zhao) manage center operations and outreach.

- Each I/UCRC has a Center Evaluator to provide guidance and to ensure compliance with NSF guidelines

- Membership fees are applied to project costs (Students, Software, Equipment, etc.)

Several Benefits of Conducting Data Research through the CDDA

- PhD level research for your organization at typically a fraction of the cost of a full-time employee

- Direct collaboration with center faculty, post-doctoral researchers, graduate students and other center members

- Project reviews, continuous informal interaction and timely access to reports, papers and intellectual property generated by the center

- Professional networking with other CDDA members
- Access to center equipment, facilities, and other CDDA infrastructure

- Recruitment opportunities with center graduate students, among some of whom may already be aligned with initiatives at your organization

- Leveraging of the research and projects of all CDDA members

Q: *What type of analytics do your consumers want? By what mechanisms do you determine this?*

A: Analytic interests vary by industry, but our strengths are in medical imaging, general imaging and data mining.

Their interests are summarized by industry (generally) and specifically determined during partnership meetings

Q: *What makes this type of data "big" (e.g., volume, velocity, variety, veracity)?*

A: Generally volume and variety, specifically as it relates to medical imaging.

Velocity becomes important in developing near-real-time analytics for trauma type applications.

Q: *Do you confront issues of data standards and interoperability?*

A: In an industry/university collaboration, these issues are typically managed by the industry-side team.

Q: *Do you confront issues of privacy/security/ethics?*

A: In an industry/university collaboration with pooled IP, data remains confidential per member, however algorithms are shared with all members.

Q: *Do you confront issues of limited current capacity in the data sciences?*

A: Yes, students with strength in mathematics and an interest in data science are uncommon.

Q: *Does your project involve partnerships or other types of sustained organizational relationships?*

A: Yes, see The Center for Dynamic Data Analytics, an NSF Industry/University Cooperative Research Center!

Q: *How has your big data work changed your field?*

A: The CDDA has helped exemplify the need for a collaborative approach to data science. The nature of the field itself is multi-disciplinary.

We also just launched a new MSc program on Biomedical Imaging and Pharmaceutical Science

Q: *What advice do you have for others running big data projects?*

A: Collaborate, collaborate, collaborate!

What Constitutes a Meaningful Dataset?

Important Issues

- How do you integrate experimental and non-experimental data sets?
- Methodologies for normalising , integrating, unifying and building big data sets from disparate sources.
  - How Big Data Analytics/Data Science can help answer these questions?
- What are the keys to quality control?

Important Issues

- How do you integrate experimental and non-experimental data sets?
- What has been the status so far?
  - Pharma typically uses “clean data” that has been thoroughly checked and reviewed.
- Challenges:
  - Integrate new data that may not be as “clean”. Compare the “clean” and “unclean” data sets.
  - Integrate data from Heterogeneous Databases

Possible Solutions

- Methodologies for normalising , integrating, unifying and building big data sets from disparate sources.
  - Requires analytics methods from the relative new field of Data Science
  - Collaboration among Clinicians, Biologists, Computer Scientists, and Statisticians
  - Reduce Costs Significantly
  - Impossible task by Humans

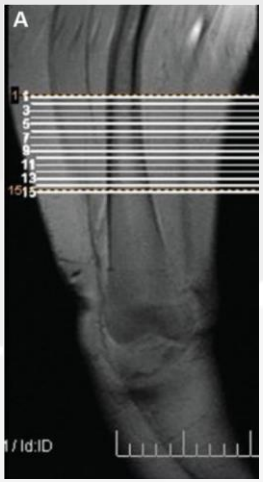
- Example Data Analytics Solutions
  - A successful collaboration between Rutgers and Bioclinica
  - Quantitative Tissue Assessment

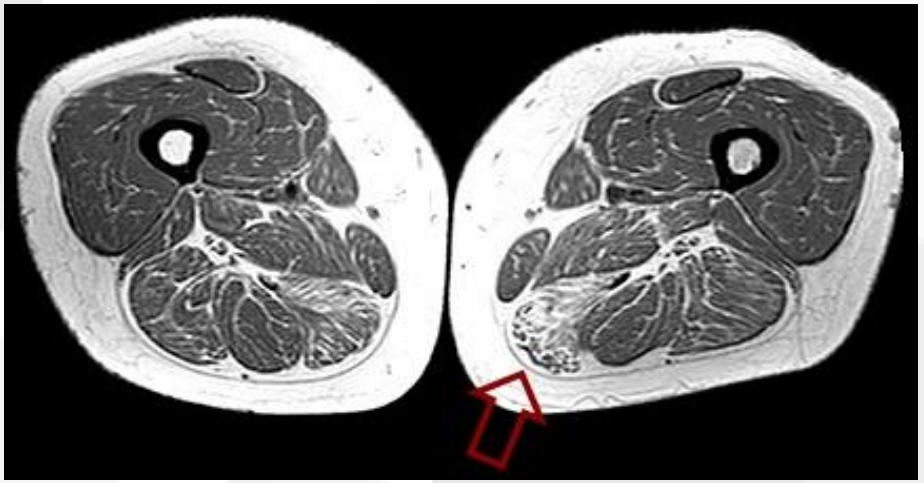
Quantitative Tissue Assessment

- Large datasets in clinical trials
  - Different imaging equipment
  - 1000+ patients (big variance, limited training)
  - 3D data (≈1e7 voxels, e.g. 400 x 400 x 200)
- Automatic tools for clinical assessment
- Quantitative data from medical images
- Muscle, liver and spleen diseases

Quantitative Tissue Assessment: Muscle
Axial T1-weighted Imaging: Acquisition

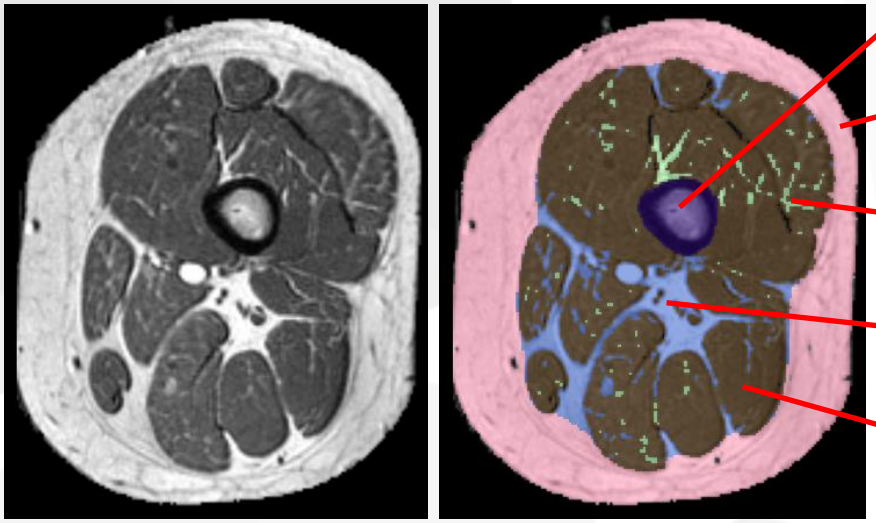
- Lower extremity scans
  - Upper leg (thigh) and/or lower leg (calf)
- 2D vs 3D acquisition and anatomical selection
- Metabolic/Fat infiltration in the thigh





Goal:
Quantitative Thigh Muscle/Fat Analysis:

- Clinical trials: large data set (big variance, no training)
- Automated quantification of different tissues
- Automated segmentation of muscle groups



Bone & marrow

Subcutaneous adipose tissue (SAT)

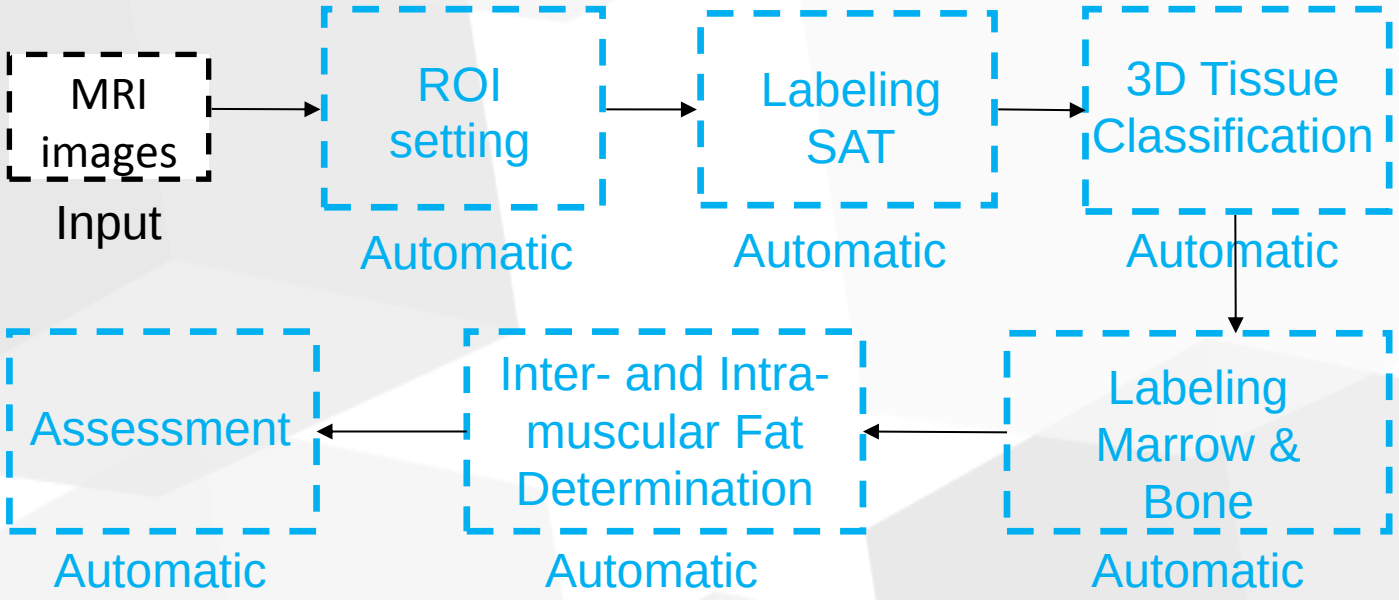
Intra-muscular fat

Inter-muscular fat

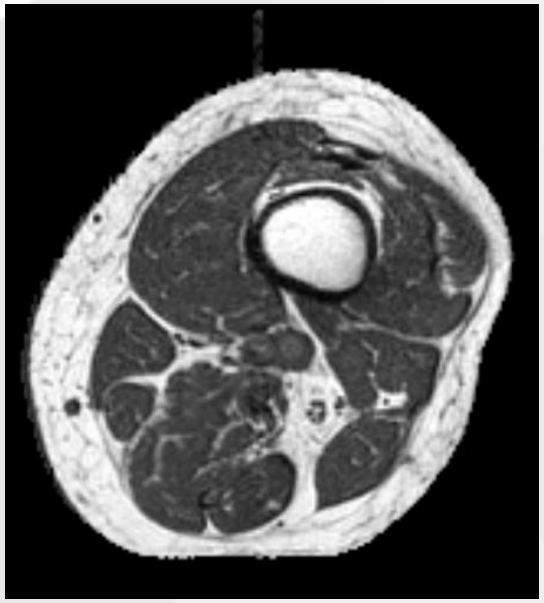
Muscle

Muscle Tissue Assessment: Quantitative Framework

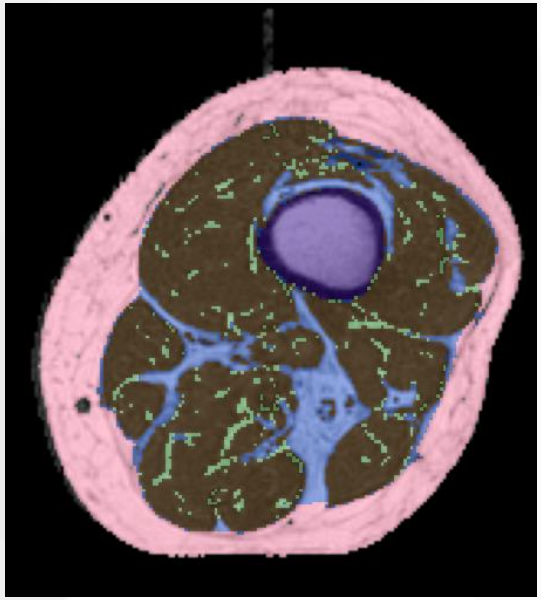
- Many different Analytics Steps: Segmentation, Labeling, Learning, Parameter Extraction, Determination, Assessment



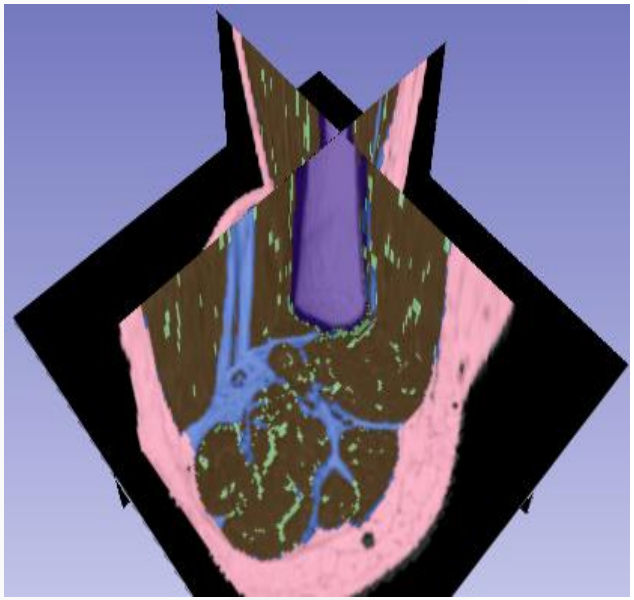
Quantitative Tissue Assessment: Muscle



A slice in original MR data



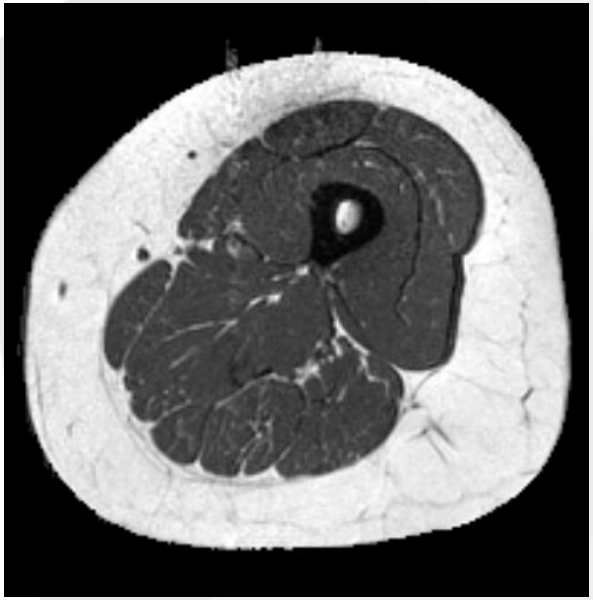
Multi-labels of the tissues



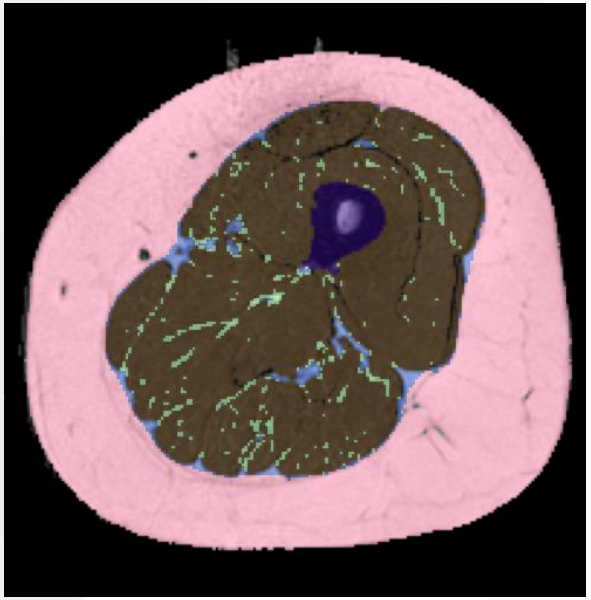
3D view of multi-labels

72-year-old male. Tissues quant (volume, cm^3):
1. Muscle: 622.21; 2. Subcutaneous fat (SAT): 445.63; 3. Inter-muscular fat (inter-MAT): 120.04; 4. Intra-muscular fat (intra-MAT): 64.32; 5. Marrow and Bone: 63.13.

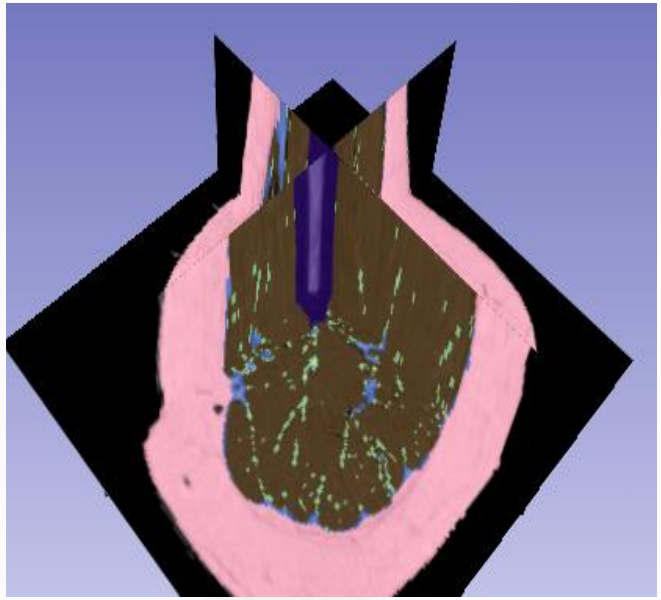
Quantitative Tissue Assessment: Muscle



A slice in original MR data



Multi-labels of the tissues

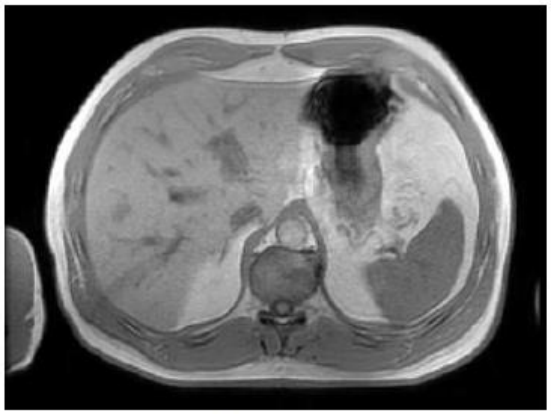


3D view of multi-labels

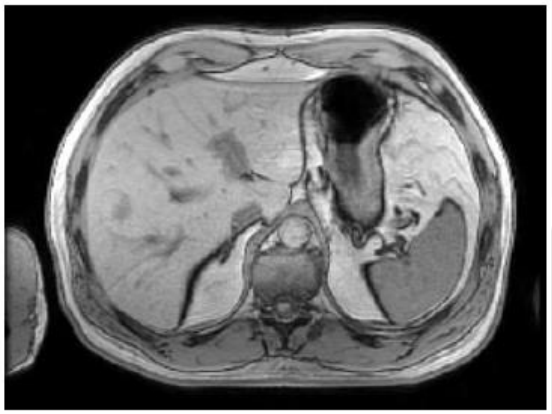
49-year-old female. Tissues quant (volume, cm^3):
1. Muscle: 667.59; 2. SAT: 868.92; 3. inter-MAT: 46.28; 4. intra-MAT: 43.03; 5. Marrow and Bone: 42.46.

Quantitative Tissue Assessment: Liver

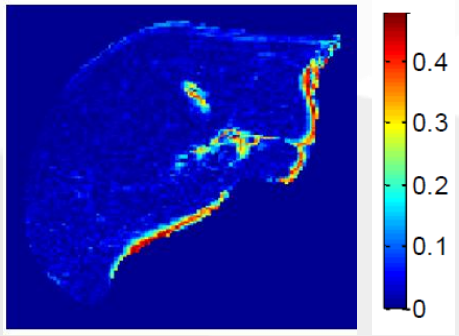
- Fatty liver diseases
- Automated 3D segmentation of liver
- Automated quantification of hepatic fat-fraction distribution



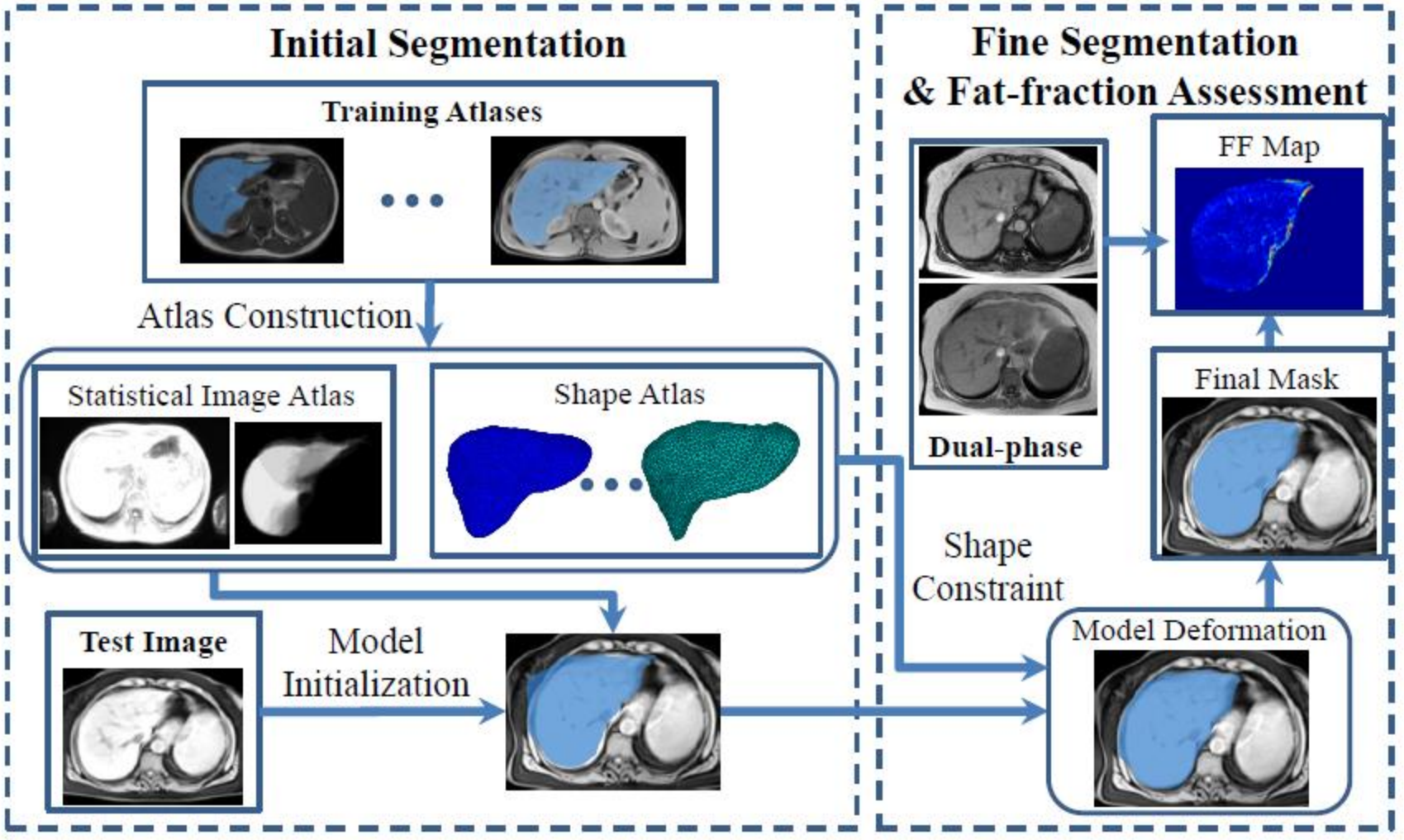
(a) In-phase



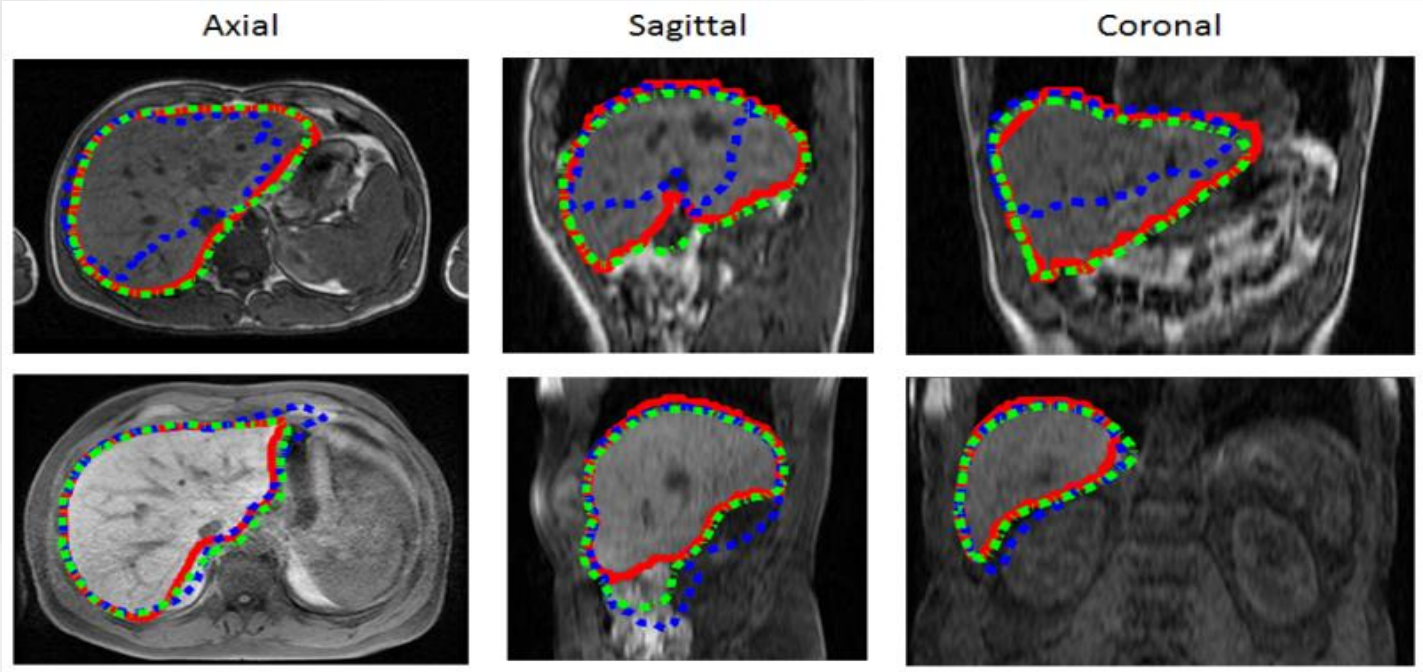
(b) Out-of-phase



Quantitative Tissue Assessment: Liver



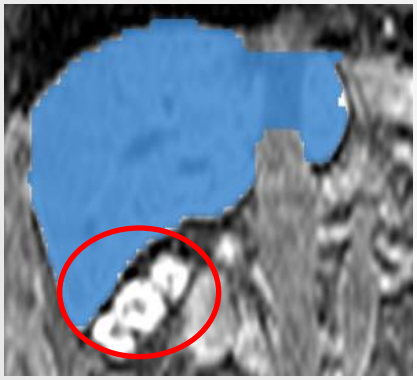
Quantitative Tissue Assessment: Liver
Deformable model



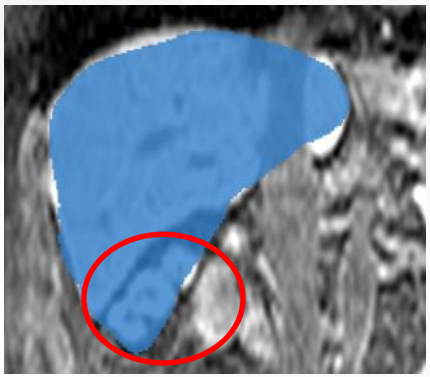
Axial view of two samples. Red solid lines are ground truth. Blue dotted lines are the surfaces of initializations. Green dotted lines are the surfaces of final segmentations.

Quantitative Tissue Assessment: Liver

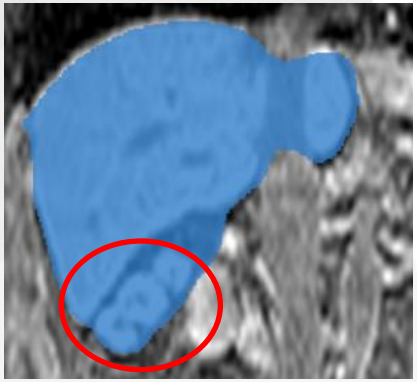
Sparse shape constraint



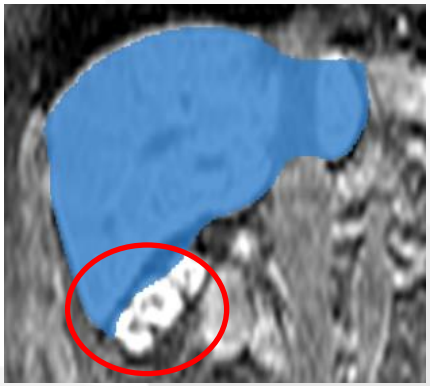
Ground truth



Initialization



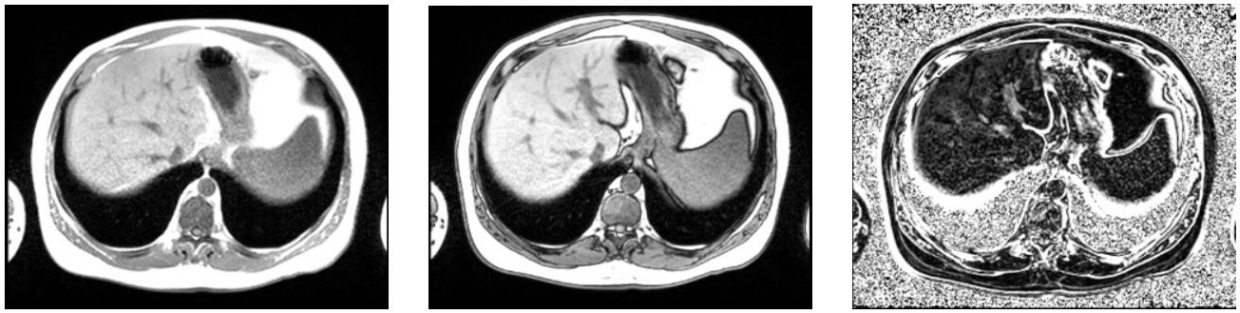
Without shape prior



With shape prior

Liver Fat Analysis:
In/Out Phase Imaging

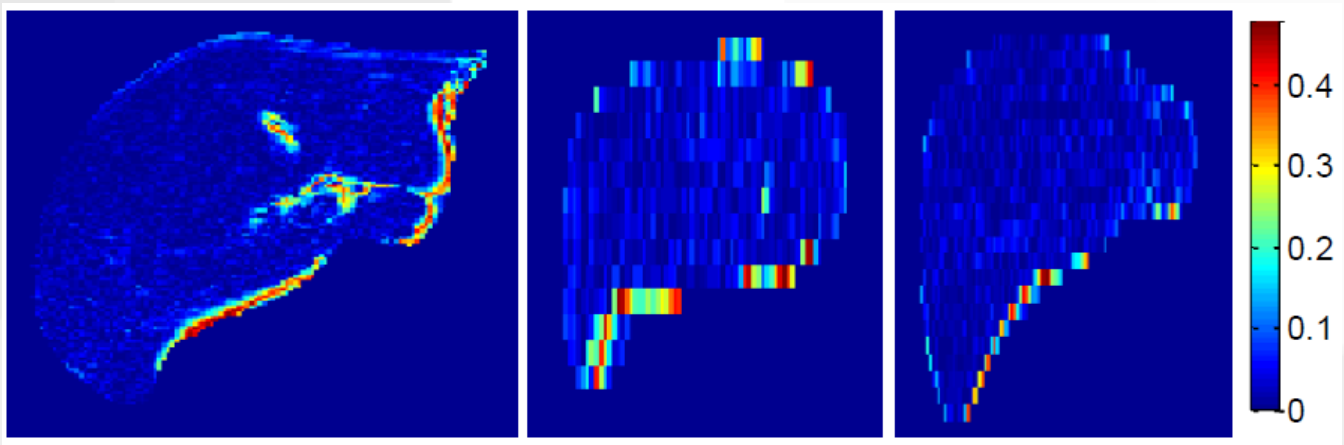
- Traditional fat-fraction distribution



(a) In-phase (b) Out-of-phase (c) Fat fraction map

Liver Fat Analysis cont.

- Liver specific fat-fraction map



Lessons Learned

- Analytics methods allow:
- Integration and assessment of new data based on Parameters Extracted
- Key improvement: Analytics allow us to obtain valuable statistics on shape, appearance and 3D data representations that can't be achieved based on 2D images only and human observation
- The extracted parameters allow us to compute the necessary properties from the data and Abstract them.
- Learning and statistical methods allow us to prune the usefulness of the data and add only those which are suitable to the database

Quality Assessment of MRI Sequences

- The quality of the MRIs must be sufficient to allow for an accurate determination of morphological features of the area for evaluation.
- The following image characteristics will be evaluated by BioClinica as part of the quality assessment.
- Occasionally, problems such as the below are unavoidable. These should be noted in the comments section of the Data Transmittal form, particularly if they are likely to recur in any repeat examination.

Quality Assessment of MRI Sequences

Good Anatomical Coverage	Proper FOV placement and adequate number of slices for each sequence.
Aliasing	Aliasing can be tolerated as long as structures of interest are not obscured
Subject Motion	Usually can be avoided by use of pads and sandbags, and by instructions to the subject
Pulsatile Motion	Artifacts from the popliteal artery are often unavoidable but can be minimized by proper placement of sat bands
Chemical Shift	Minimized by proper choice of band-width and frequency encode direction. The optimal phase encode direction and receiver bandwidth has been pre-programmed into each sequence, and should not be changed.
Fat Saturation Failure	Partial fat sat failure is tolerable at the image periphery as long as the contrast of structures of interest (for example, cartilage or bone marrow edema) is not compromised
Susceptibility Effects	May arise from implants and other metallic objects; acceptable if they do not obscure structures of interest

Clinical Data Management

- Clinical Data Managers
  - Experienced data managers trained in each therapeutic area and modality
  - Provide end-to-end oversight to ensure data integrity
  - Work closely with study teams to map out the big “data picture” and ensure the proper data quality controls are in place
- Deliverables
  - Consultative interaction to present best options for the study
  - Data and Image Transfers (Test & Production)
  - Experienced in CDISC standards
- Data Validation
  - Data validation embedded throughout processes and systems
  - Extensive range of standard checks with custom study-specific checks developed as needed

Publications

- Z. Yan, S. Zhang, C. Tan, B. Belaroussi, Y. Zhou, H. J. Yu, C. Miller, D. Metaxas: Automatic Liver Segmentation and Hepatic Fat Fraction Assessment in MRI. ICPR, 2014.
- C. Tan, Z. Yan, S. Zhang, B. Belaroussi, H. J. Yu, C. Miller, D. Metaxas: A Robust and Automatic Framework for Assessment of the Muscle/Fat Fraction in Thigh. ICPR, 2014.
- Z. Yan, S. Zhang, C. Tan, B. Belaroussi, H. J. Yu, C. Miller, Dimitris Metaxas: Atlas-Based Liver Segmentation and Hepatic Fat-Fraction Assessment for Clinical Trials. Computerized Medical Imaging and Graphics. 2014
- H. J. Yu, C. Tan, Z. Yan, B. Belaroussi, Y. Zhang, D. Metaxas, J. Carrino, C. Miller: Tissue Quantification: Thigh MRI in Osteoarthritis. IWOAI, 2014.

If you have any questions, please feel free to contact: dnm@cs.rutgers.edu