Project IX
Big Data Meets Medical Physics
Dosimetry

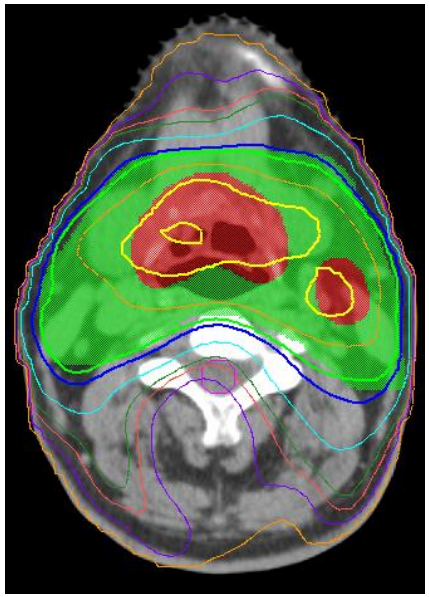
Fumbeya Marungo, Hilary Paisley, John Rhee

Mentors: Dr. Todd McNutt, Dr. Scott Robertson

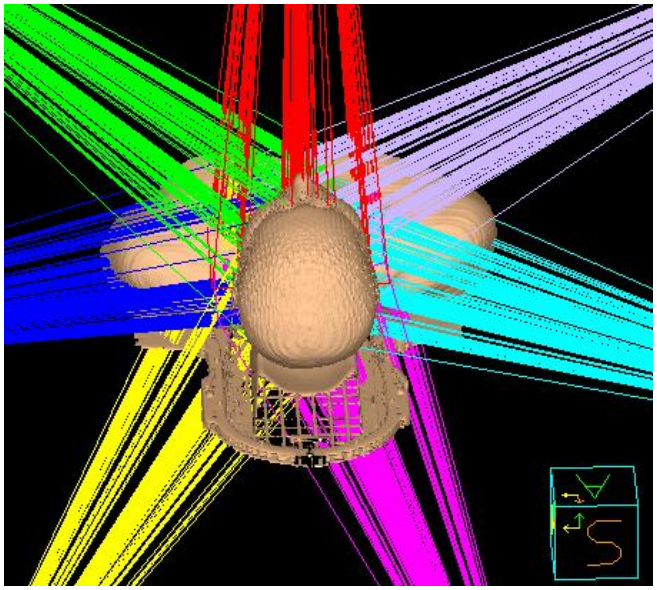
Second Project Checkpoint Presentation, April 29, 2014

Topic – Radiotherapy Planning

- In radiotherapy planning medical physicist wish to maximize the treatment dose while minimizing the risk of side effects.
- The conventional models for assessing treatment plans do not account for the multidimensional nature of the problem.



Images copied from http://oftankonyv.reak.bme.hu/tiki-index.php?page=The+Clinical+Application+of+IMRT



Present Complication Risk Models

- Lyman-Kutcher-Berman model reduces a dose volume histogram (DVH) to two values, V_{eff} and D_{max} .
- Uses values to calculate complication probability (NTCP).

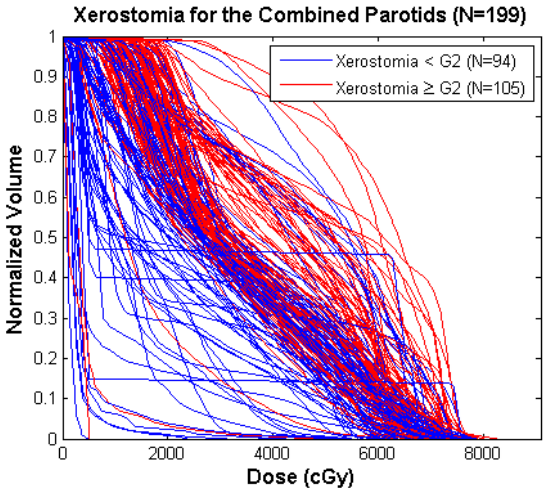
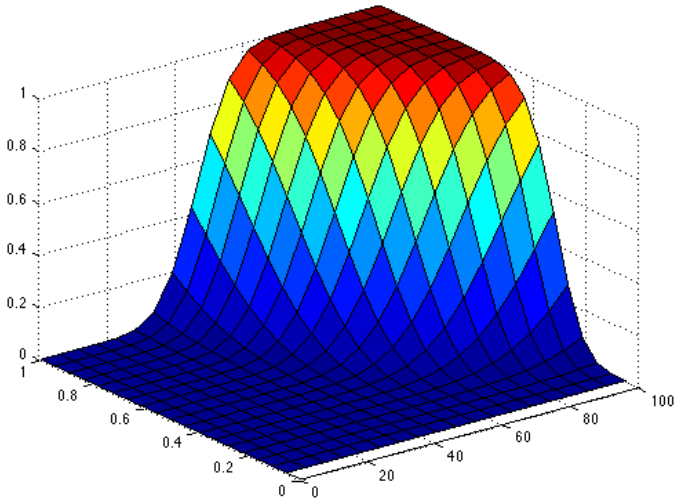


Image courtesy of Dr. Todd McNutt, Dr. Scott Robertson



JOHNS HOPKINS MEDICINE

ciis

ERC | CISST

Collaborative Laboratory for BIGSS Biomechanical- and Image-guided Surgical Systems

LABORATORY FOR Computational Sensing + Robotics THE JOHNS HOPKINS UNIVERSITY

Use Big Data to Reduce Complication Risk

- Current treatment planning can be informed by prior experience.
- Multi-factored view of risk.
- First step, use dose location data.

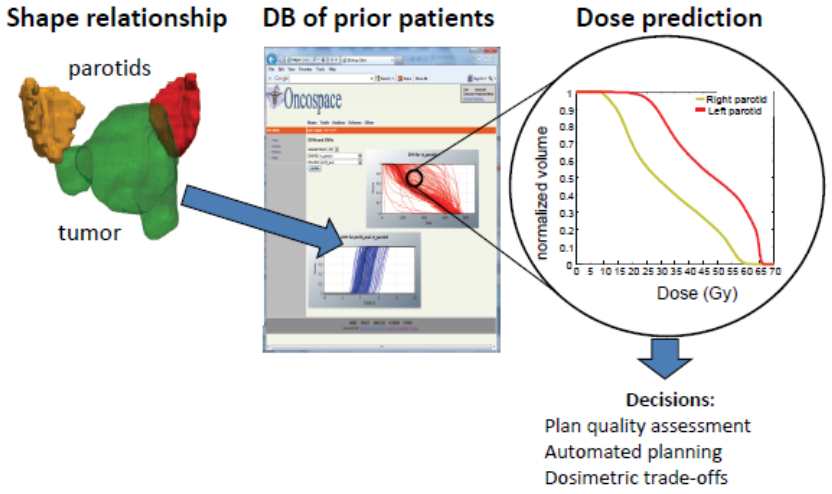


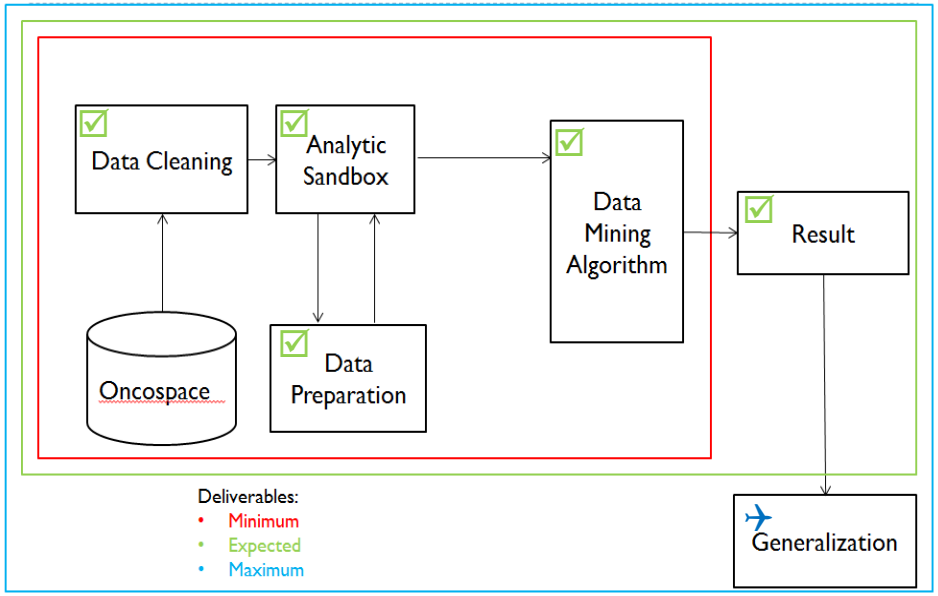
Image courtesy of Dr. Todd McNutt, Dr. Scott Robertson

Team Member Responsibilities

- Fred Brooks presents a "Surgical Team" approach developed by Harlan Mills and F. Baker at IBM.
- Chief programmer leads the effort.
- Other roles naturally divide into administrative and technical functions.
- Our project requirements:
  - Thirteen project related artifacts.
  - Creating a software pipeline
  - PM and SE responsible for "1st drafts"
  - TL responsible for refining drafts and final data mining.

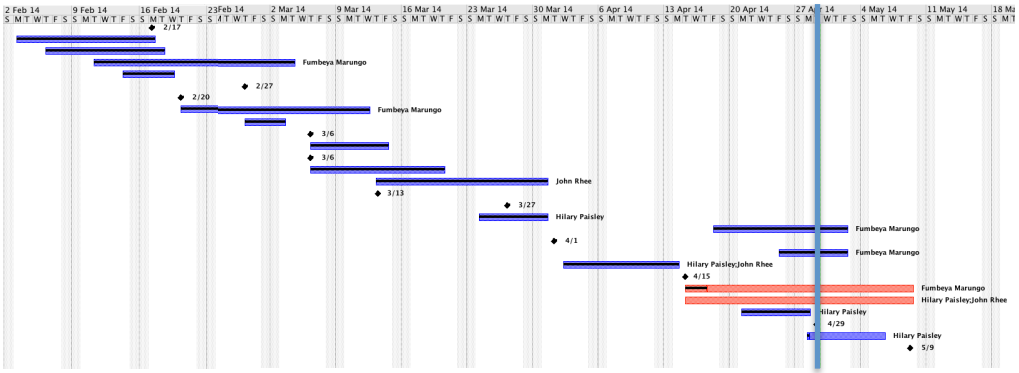
Deliverables

- We have created an analysis pipeline.
- We have obtained usable results for parotid glands via data mining.
- We are working on applying the approach to the larynx as part of our maximum deliverables.



Project Plan Management

- Demo of ProjectLIBRE.
  - Project tasks completed.
  - Task notes.
  - GANTT chart.



		Name	Duration	Start	Finish	Resource Initials
2		Maintain Wiki	74 days	1/28/14 8:0...	5/9/14 5:0...	
3	✓	Project Planning Presentation	1 day	2/11/14 8:0...	2/11/14 5:...	
4	✓	Project Planning Report	1 day	2/17/14 8:0...	2/17/14 5:...	
5	✓	Project Planning	11 days	2/3/14 8:00...	2/17/14 5:...	
6	✓	Setup Development Environme	9 days	2/6/14 8:00...	2/18/14 5:...	
7	✓	Literature Review	16 days	2/11/14 8:0...	3/4/14 5:0...	FM
8	✓	IRB	4 days	2/14/14 8:0...	2/19/14 5:...	
9	✓	Database Access	1 day	2/27/14 8:0...	2/27/14 5:...	
10	✓	Meeting with Mentors	1 day	2/20/14 8:0...	2/20/14 5:...	
11	✓	Begin Preparing Paper Seminar	15 days	2/20/14 8:0...	3/12/14 5:...	FM
12	✓	Define Initial Data Preprocessir	3 days	2/27/14 8:0...	3/3/14 5:0...	
13	✓	Target Database Access	1 day	3/6/14 8:00...	3/6/14 5:0...	
14	✓	Data Clensing and Preprocessi	7 days	3/6/14 8:00...	3/14/14 5:...	
15	✓	Meeting with Mentors	1 day	3/6/14 8:00...	3/6/14 5:0...	
16	✓	Develop Target Database	11 days	3/6/14 8:00...	3/20/14 5:...	
17	✓	Data Reduction and Transform	13 days	3/13/14 8:0...	3/31/14 5:...	JR
18	✓	Seminar Presentation - Fumbe	0.438 d...	3/13/14 1:3...	3/13/14 5:...	FM
19	✓	Meeting with Mentors	1 day	3/27/14 8:0...	3/27/14 5:...	
20	✓	Plan for Check Point	6 days?	3/24/14 8:0...	3/31/14 5:...	HP
21	✓	Data Mining	11 days	4/18/14 8:0...	5/2/14 5:0...	FM
22	✓	Check Point Presentation	1 day	4/1/14 8:00...	4/1/14 5:0...	
23	✓	Assess Models	6 days	4/25/14 8:0...	5/2/14 5:0...	FM
24	✓	Seminar Prep	9 days?	4/2/14 8:00...	4/14/14 5:...	HP;JR
25	✓	Seminar Presentation - Hilary/	0.125 d...	4/15/14 8:0...	4/15/14 9:...	HP;JR
26		Writing Report	19 days	4/15/14 8:0...	5/9/14 5:0...	FM
27		Work on Poster	19 days	4/15/14 7:0...	5/9/14 5:0...	HP;JR
28	✓	Prep Final Checkpoint	6 days?	4/21/14 8:0...	4/28/14 5:...	HP
29	✓	Final Checkpoint Presentation	1 day	4/29/14 8:0...	4/29/14 5:...	
30		Code Review and Documentati	7 days?	4/28/14 8:0...	5/6/14 5:0...	HP
31		Poster Day	1 day	5/9/14 8:00...	5/9/14 5:0...	FM;HP;JR

Timeline

February 2014

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
						1
2	3 Intro Discussion M @ 10pm	4	5	6 Report Completed M @ 3pm	7	8
9	10 Presentation Prep M @ 10pm	11 Project Presentation	12	13 Discuss Project MM @ 4pm	14	15
16	17 Setup dev environment M @ 10pm	18	19 IRB	20	21	22
23	24	25	26	27 Step 1, Read papers M @ 3pm	28	

March 2014

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
						1
2	3 Data Processing M @ 10pm	4 Literature Review	5	6 Database Access MM @ 4pm	7	8
9	10 Review of MM M @ 10pm	11	12	11 Seminar Presentation - Fumbeya	14	15
16	17 Prep Checkpoint M @ 10pm	18	19	20	21	22
23	24 Data Reduction M @ 10pm	25	26	27 Write Initial Paper MM @ 3pm	28	29
30	31 Look at Data M @ 10pm					

Timeline

April 2014

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
		1 Checkpoint Presentation	2	3	4	5
6	7 Bugs in Code M @ 10pm	8	9	10 Fix Coding Bugs MM @ 4pm	11	12
13	14 Seminar Prep M @ 10pm	15 Seminar Presentation– Hilary/John	16	17	18	19
20	21 Data Mining M @ 10pm	22	23	24 How to get Results MM @ 4pm	25	26
27	28 Checkpoint Prep M @ 10pm	29 Mini Checkpoint	30			

May 2014

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
				1 Discuss Results MM @ 4pm	2 Assess Models	3
4	5 Poster Prep M @ 10pm	6 Code Review	7 Finish Poster M	8 Solidify Presentation M	9 Poster Session	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

JOHNS HOPKINS MEDICINE

ciis

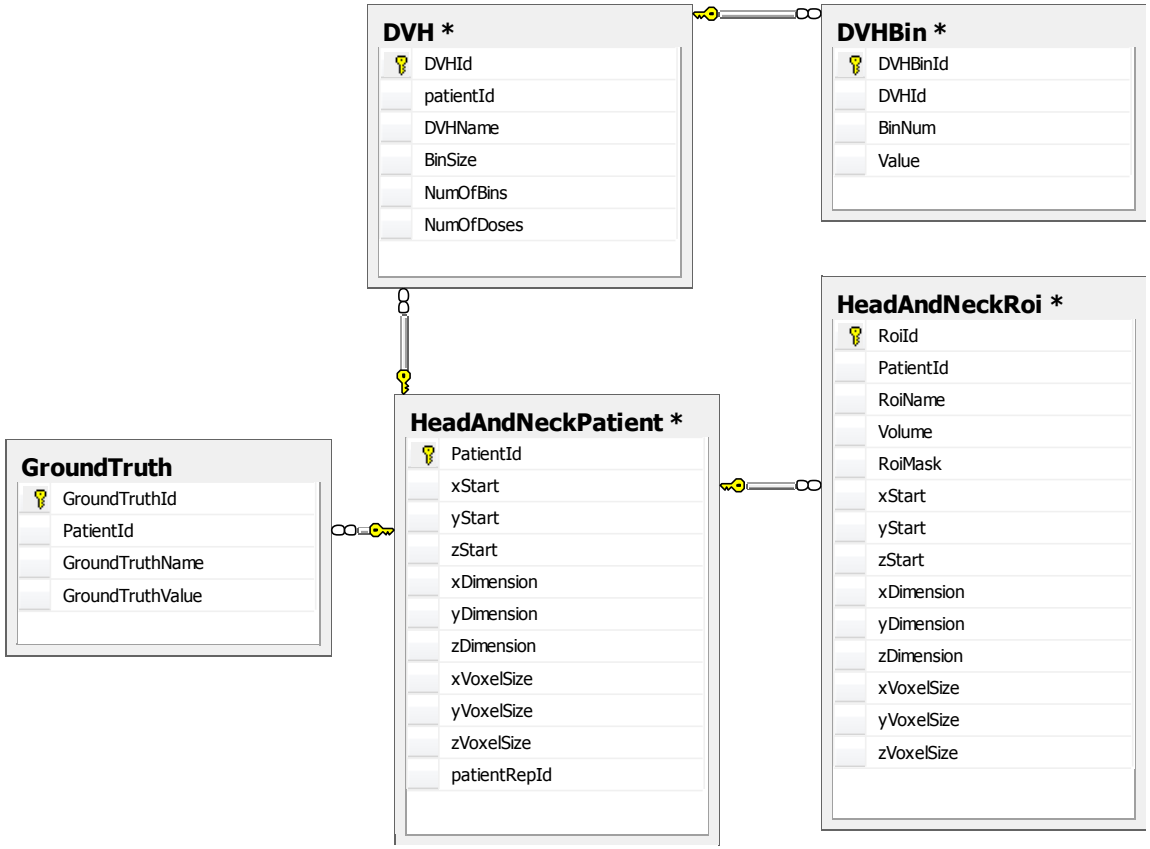
ERC | CISST

Collaborative Laboratory for BIGSS Biomechanical- and Image-guided Surgical Systems

LABORATORY FOR Computational Sensing + Robotics THE JOHNS HOPKINS UNIVERSITY

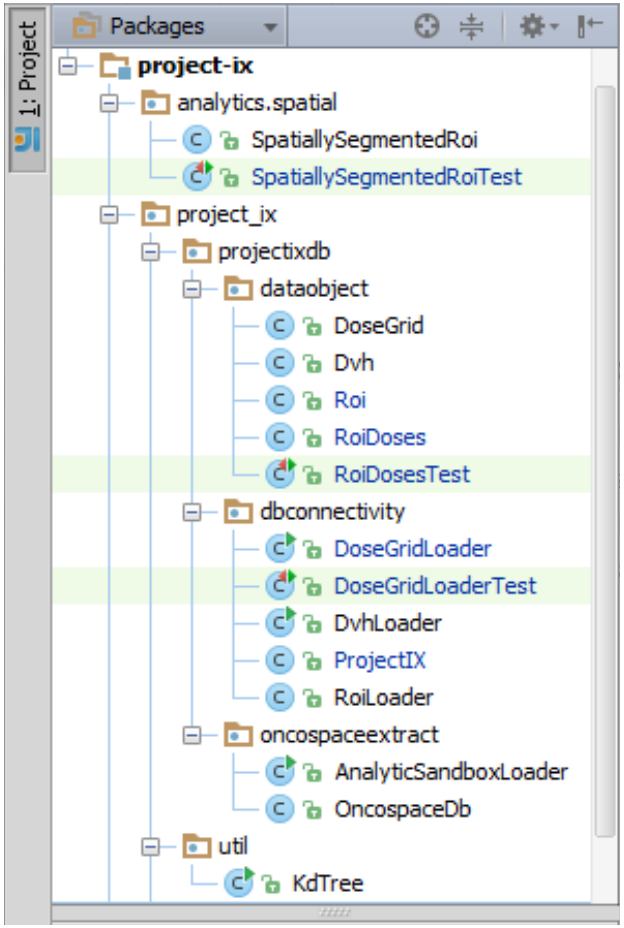
The Analytic Sandbox

- Separate from Oncospace.
- Dose grids benefit from ~75% data compression.
- Consistent BigEndian encoding.

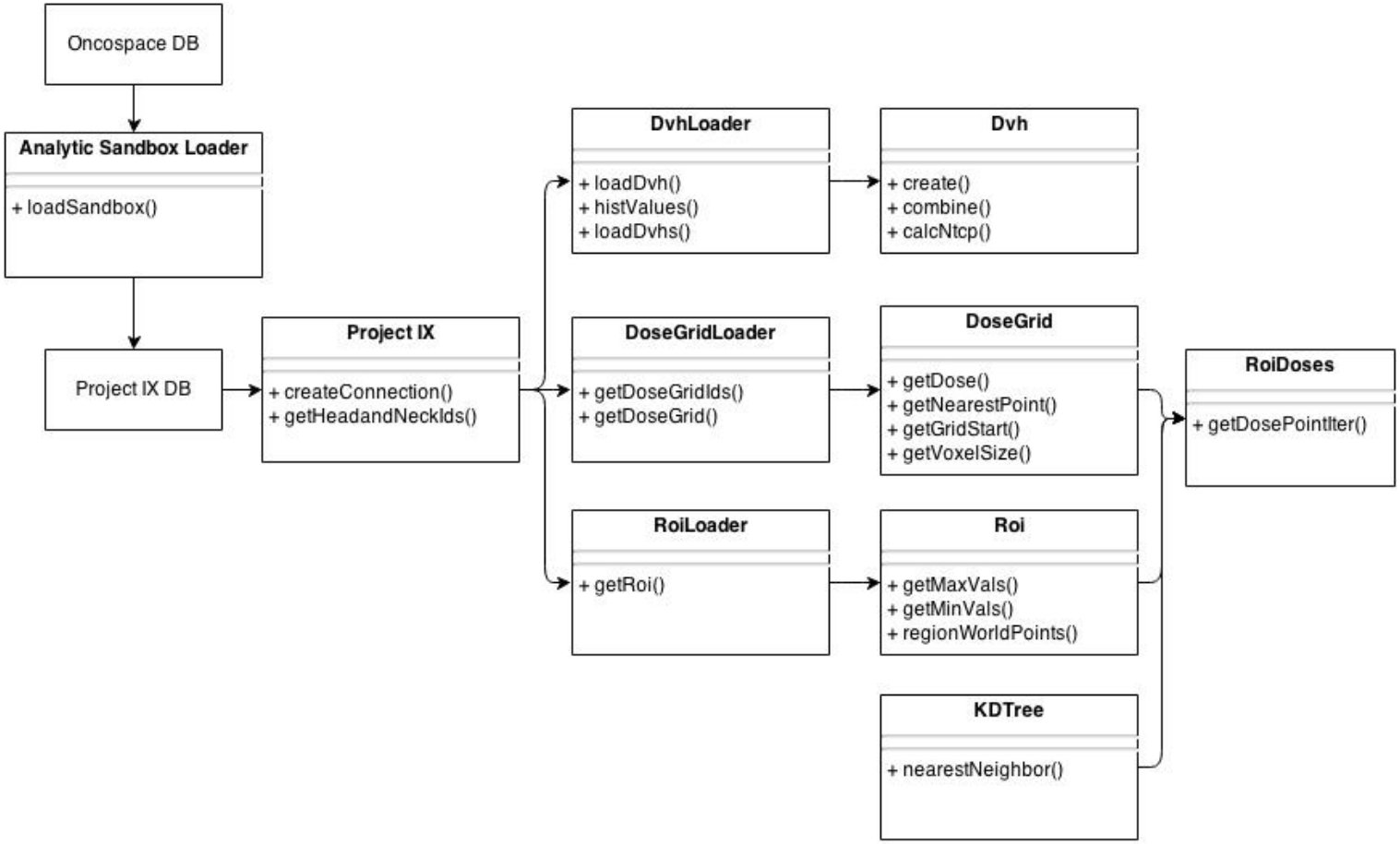


Software Design

- Data loader classes load data from analytic sandbox.
- Roi class is a volumetric mask.
- DoseGrid is a mapping of voxels to dose. It uses a k-d Tree to find the nearest voxel to a given point.
- RoiDoses uses a Roi and a DoseGrid to return a dose value for each Roi voxel in world frame coordinates.

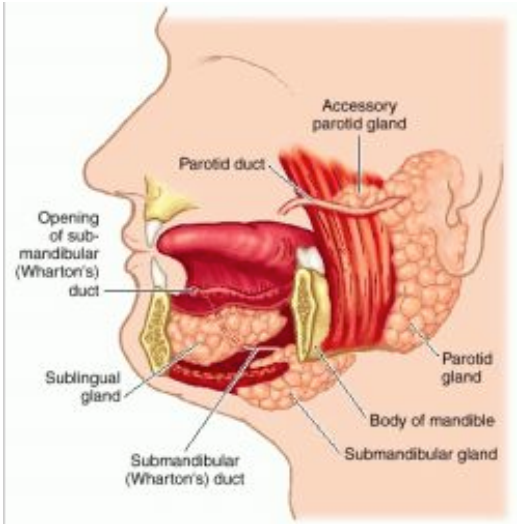


Software Design



Parotid Structure

- Parotid gland is irradiated during treatment of head and neck cancer patients.
- Oncospace provides 3D treatment data.



From medical-dictionary.thefreedictionary.com

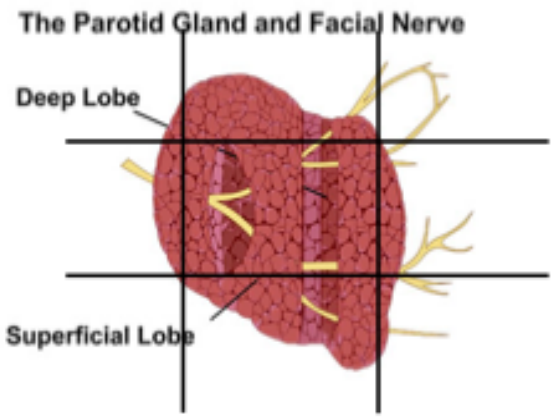
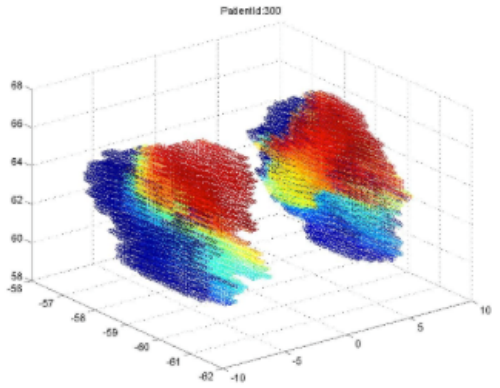


Image courtesy of Dr. Todd McNutt, Dr. Scott Robertson



JOHNS HOPKINS MEDICINE

ciis

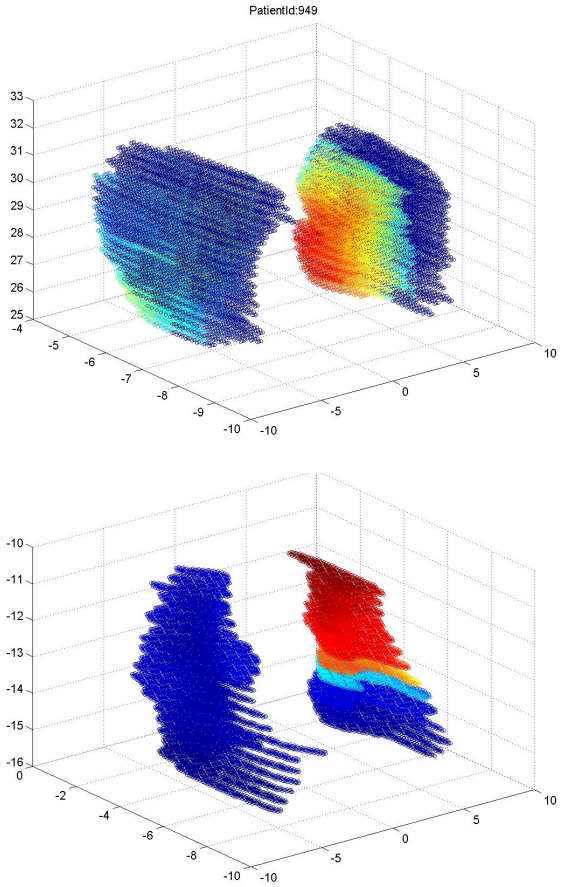
ERC | CISST

Collaborative Laboratory for BIGSS Biomechanical- and Image-guided Surgical Systems

LABORATORY FOR Computational Sensing + Robotics THE JOHNS HOPKINS UNIVERSITY

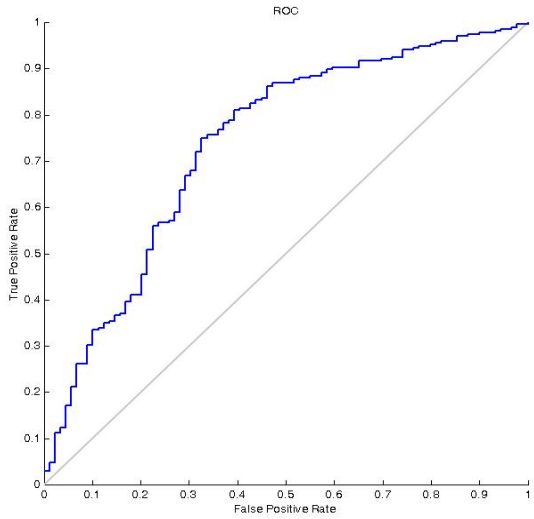
Features

- Included LKB values for V_{eff}, D_{max}, and NTCP.
- Defined five dose bands.
  - 5-25 Gy, 25-40 Gy, 40-55 Gy, 55-70 Gy, and >70 Gy
- Created gross left, right, and both DVHs for each band.
- Percentage of each dose band in fifths of each dimension.
- Selected 18 features based on information entropy reduction.



Results

- Evaluated ROC for LKB, Random Forest and Linear Regression.
- Performed stratified cross-validation with k = 10.
- Both Linear Regression and LKB outperformed Random Forest (n=364):
  - Linear Regression: 0.731
  - LKB: 0.723
  - Random Forest (1000 trees): 0.683
- Opportunities with larger data?



Summary

- Used recognized knowledge discovery procedures.
- Completed data mining of parotid data.
- Will be presenting results to mentors for evaluation.
- Beginning work on larynx data.

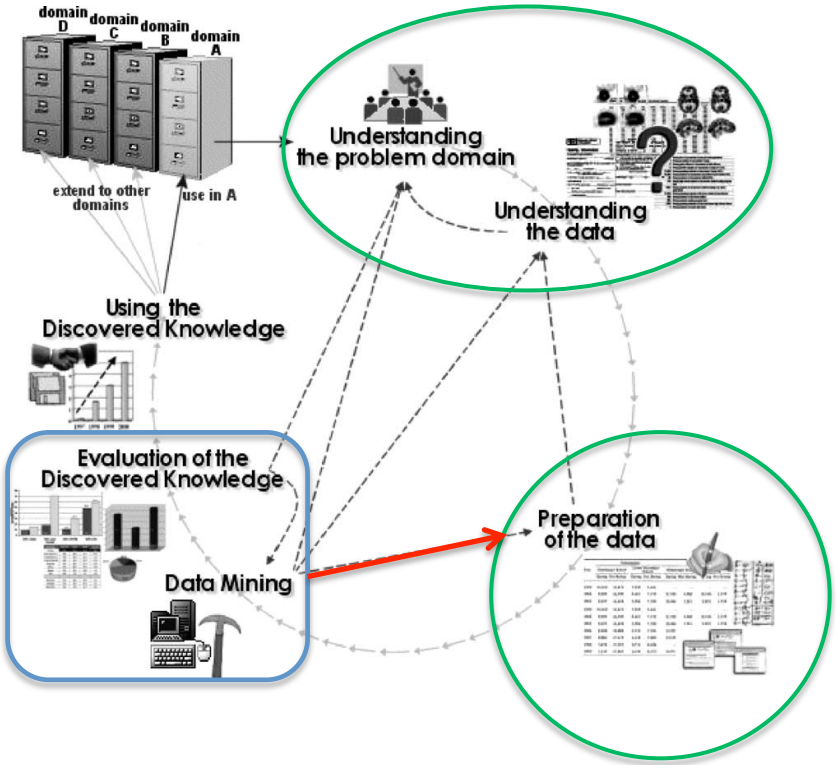


Image from (Cios et al. 2002)

Thank You

- Dr. Todd McNutt, Mentor
- Dr. Scott Robertson, Mentor
- Dr. Russell Taylor, Instructor
- CIS II Classmates…