

Biomedical Informatics

@ emory.edu

BMI 500

There are only 10 types
of people in the world:
Those who understand binary
and those who don't.



EMORY
UNIVERSITY

**Department of
Biomedical Informatics**
Emory University School of Medicine

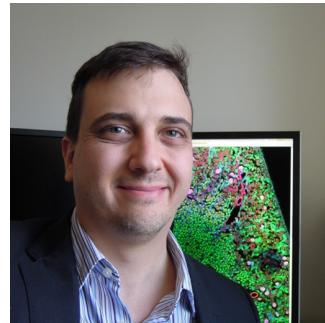
Who is BMI? 23 faculty, 11 for BMI 500



Gari Clifford
Interim Chair
Assoc. Prof.
(BMI/BME)



Jim Blum
Vice-Chair for
Clinical Research,
Chief, Critical Care,
Ass. Prof. Anesthesiology



Lee Cooper
Assistant Prof.
(BMI/BME)



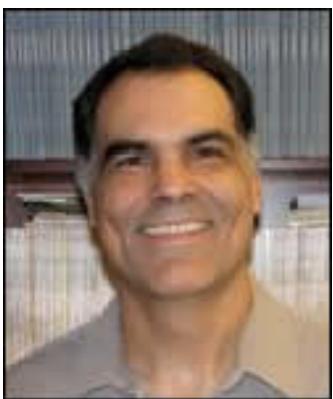
Jun Kong
Assistant Prof.



Shamim Nemati
Assistant Prof.



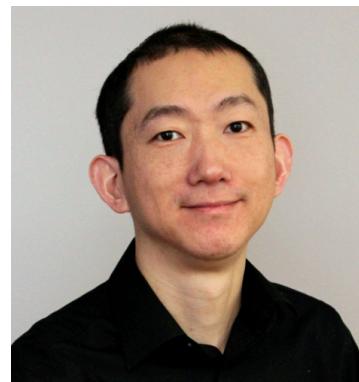
Babak Mahmoudi
Assistant Prof.



Carlos Moreno
Associate Prof.
Pathology & BMI



Andrew Post
Associate Prof.
BMI/ACTSI



Zhaohui (Steve) Qin
Associate Prof.
(Biostats & BMI)



Ashish Sharma
Assistant Prof. /
Co DGS



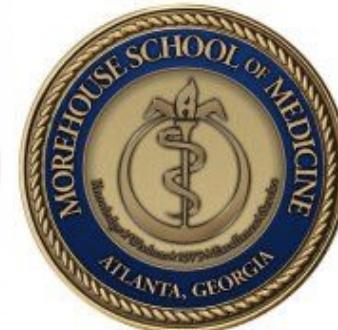
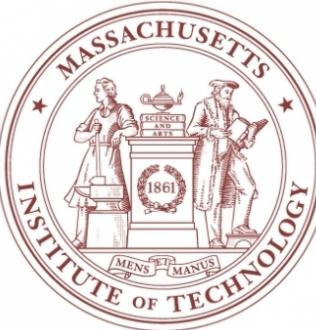
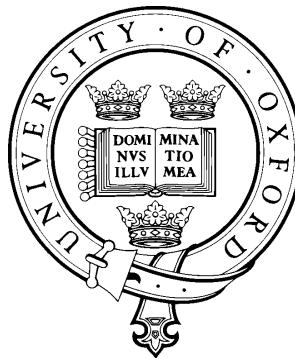
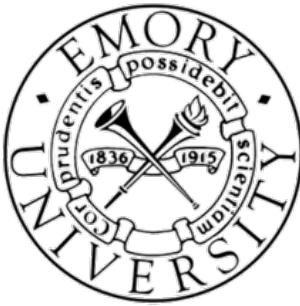
Li Xiong
Associate Prof.

Gari Clifford



<http://gdclifford.info>

BSc. Physics & Electronics (Exon)
MSc. Math & Theoretical Physics (Soton)
DPhil in AI / Sig Proc (Oxon)
Postdoc HST (MIT)
Principal Research Scientist (MIT)
Associate Prof BME (Oxford)
Directory, Center for Doctoral Training, BME (Oxford)
Associate Prof Emory (BMI) / GTech (BME)
Interim Chair BMI
Honorary Professor, University of Oxford
Adjunct Prof., Morehouse School of Medicine
Distinguished Guest Prof., Tsinghua University
Deputy Editor, Journal of Physiological Measurement, Institute of Physics.



Big Data (Streaming) Healthcare – Sig. Proc. & Machine Learning

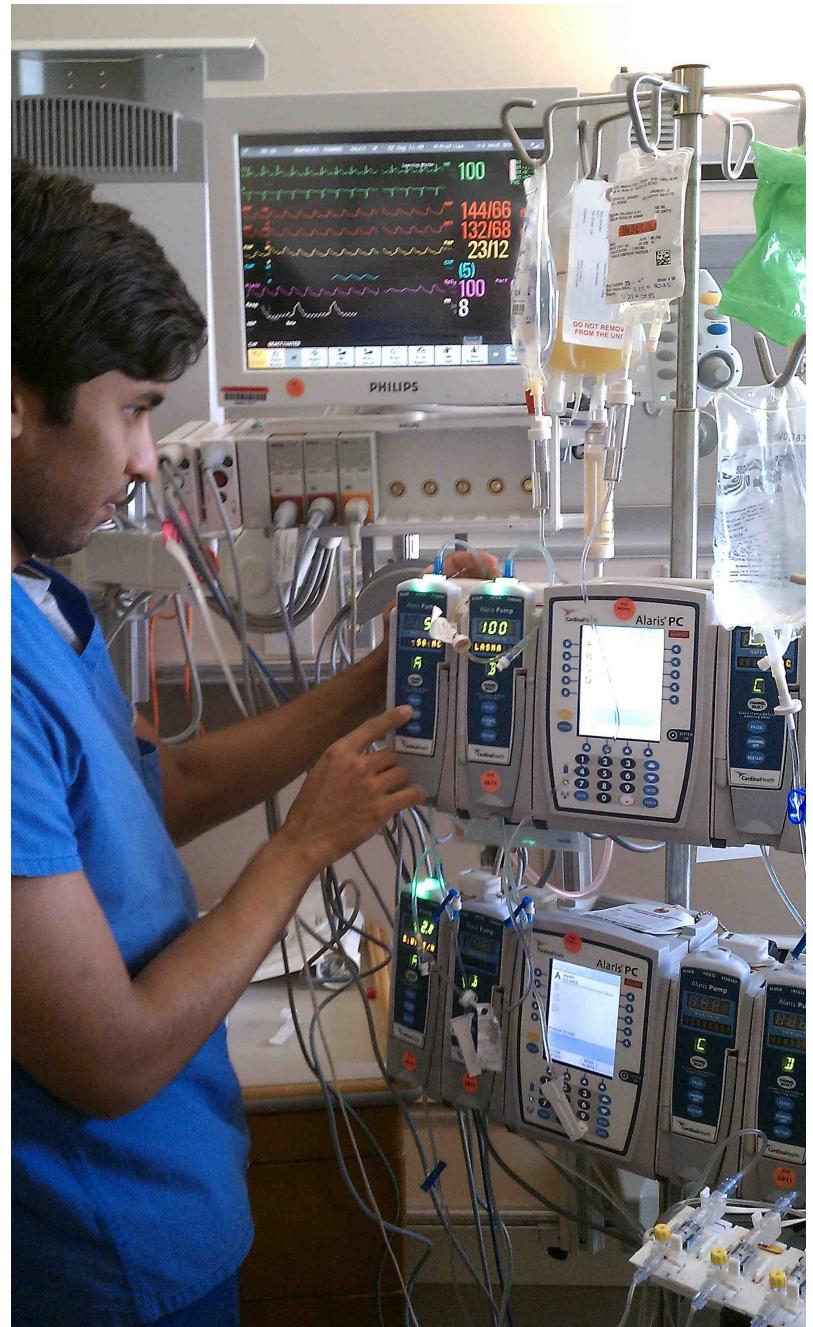
Currently at Emory's ICUs:

Volume – 24TB over the last two years

Velocity – 1 GB/person/day

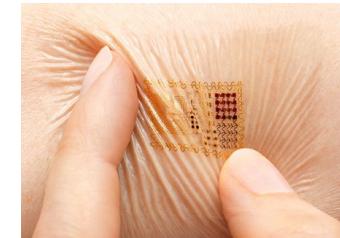
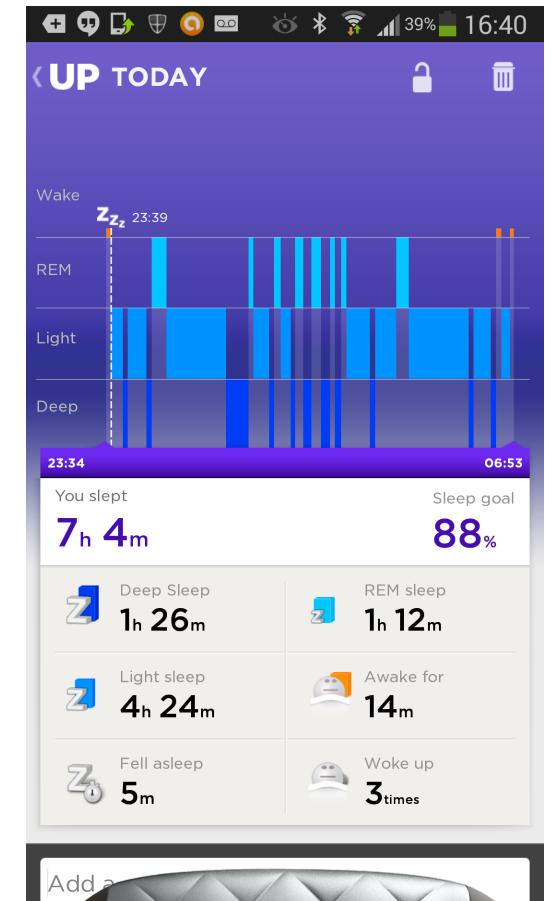
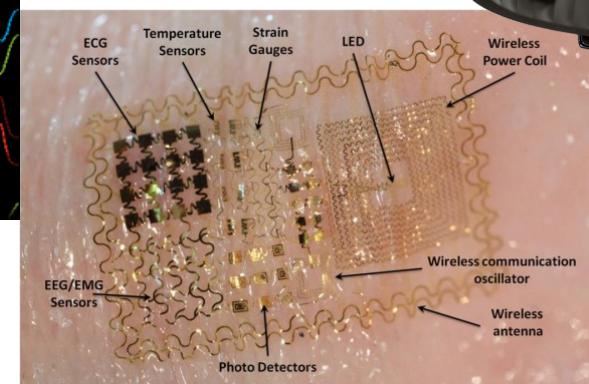
Variety – Up to 10,000 parameters
electrophysiology, images, free text,
microbiology, interventions, ICD9

Veracity – up to 90% of alarms are false!



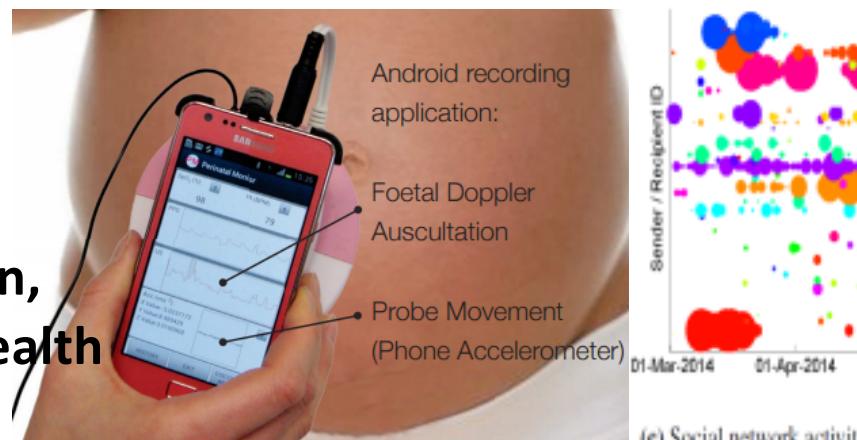
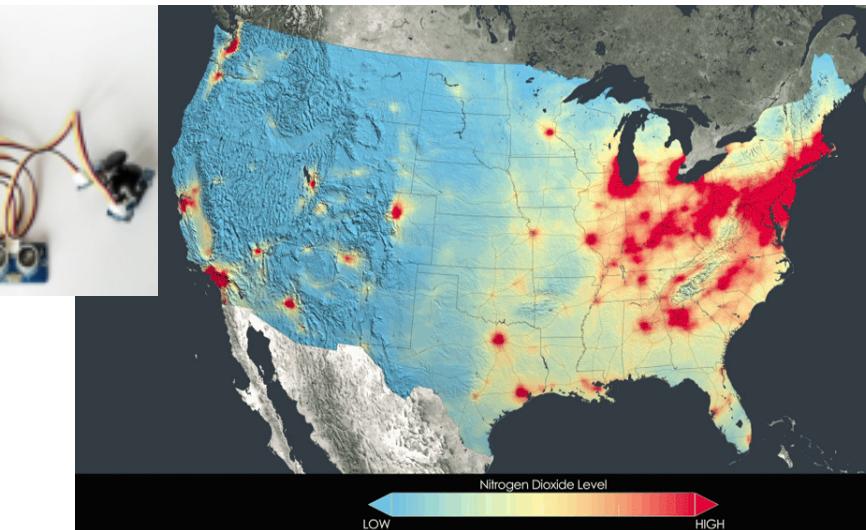
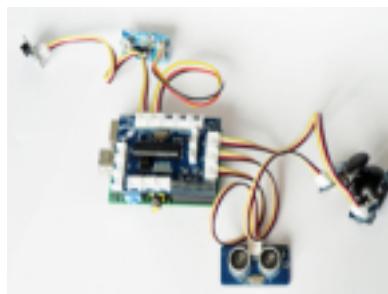
Sleep, Circadian Rhythms Neuroscience & Behaviors

- EEG & ECG analysis
- O₂, Respiration, Movement
- Phone Apps
- Wearables
- Actigraphy analysis
- Physiology-based
- Video
- Voice analytics

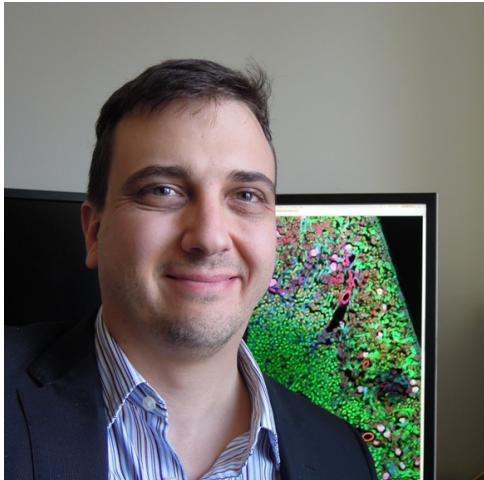


Projects: ML & Sig. Proc. for ...

- Sleep / mental health Computational Psych. (Depression, Schizophrenia, PTSD)
- Biopschosocial effects on CVD / PTSD
- Healthy Aging Study (Alzheimer's)
- Seizure detection (Rhett)
- CHF readmissions using wearables
- AF & Stroke Prediction
- Environmental factors @Exposome Center
- Context Aware Electronic Medical Records
- Video: Sedation and Agitation in the ICU
- Voice analytics (neurodegeneration)
- Eye tracking (vigilance, accuracy, disease progression)
- Perinatal: IUGR, fetal stressors, preconception, intrapartum monitoring, growth, maternal health



Lee Cooper



PhD in Electrical Engineering - Signal Processing & Control (Ohio State)

Assistant Professor of Biomedical Informatics (Emory)

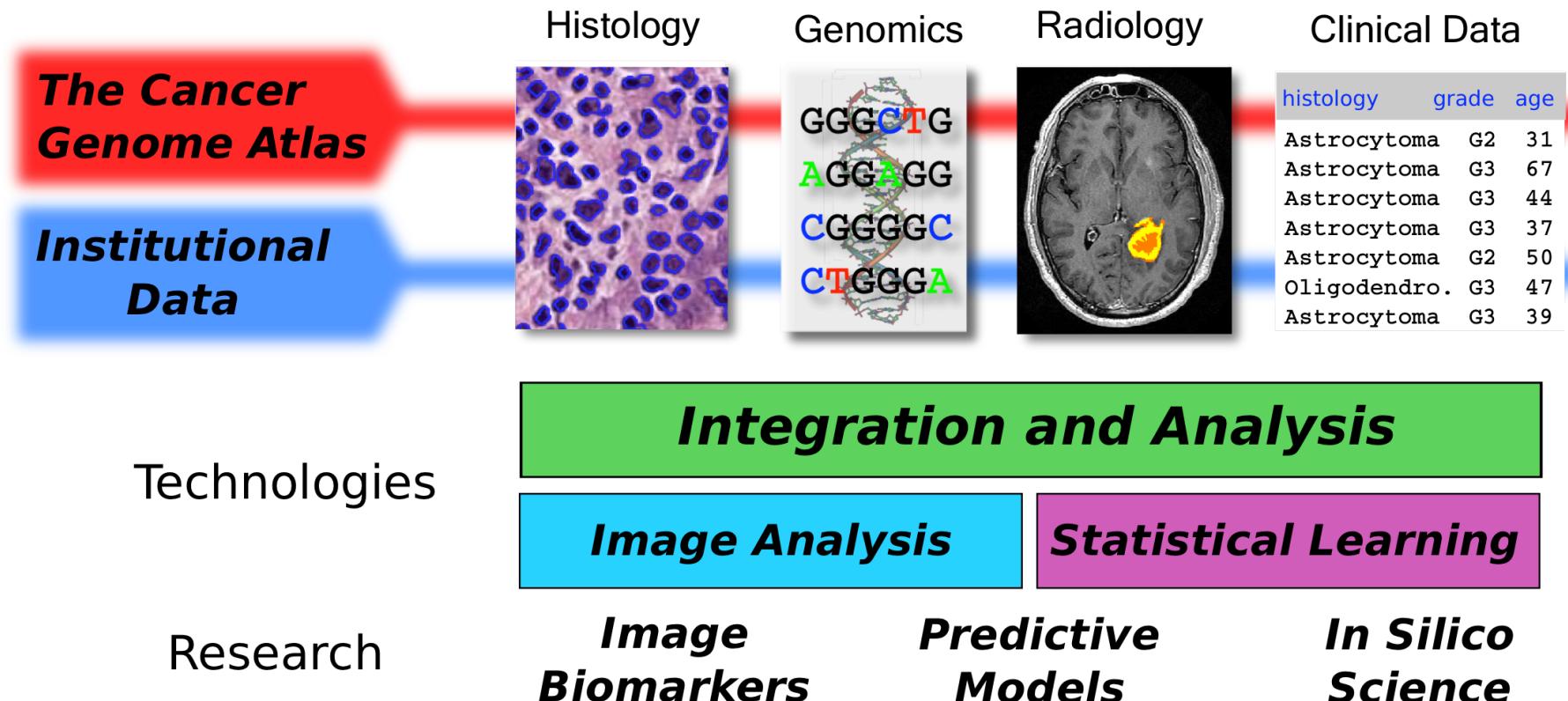
Assistant Professor of Biomedical Engineering (GT)

**Core Member, Cancer Genetics & Epigenetics Program,
Winship Cancer Institute**

Faculty, Cancer Biology Program (Emory)



Transforming Cancer Data to Knowledge & Practice



How can data + algorithms:

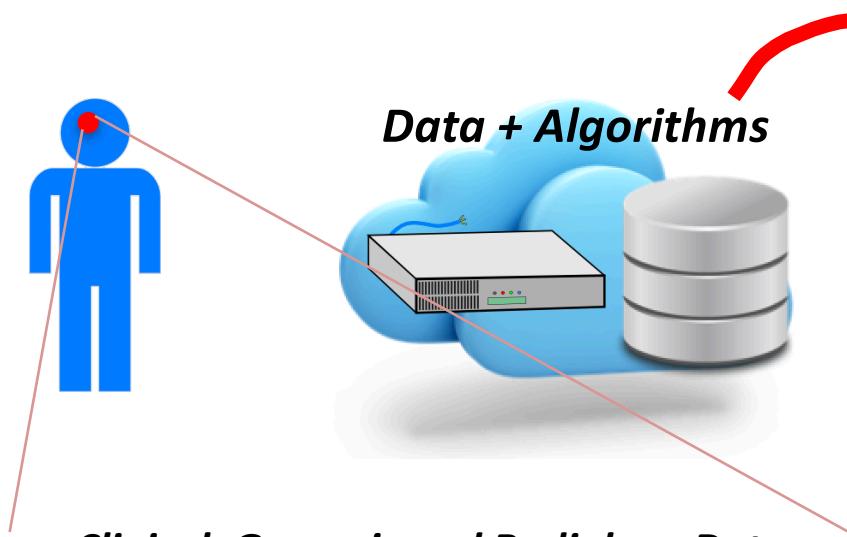
- Improve prediction of clinical outcomes and patient care? (translation)
- Provide understanding of fundamental biological processes? (science)

TCGA Research Network, *Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas*.

[N Engl J Med.](#) 2015 Jun 25;372(26):2481-98

L.A.D. Cooper

Predicting Outcomes from High-Dimensional Genomics

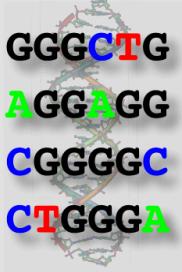


Clinical, Genomic and Radiology Data

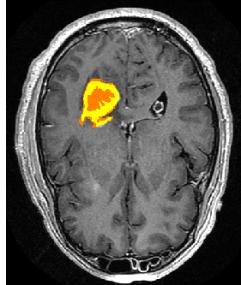
Clinical (100+)

histology	grade	age
Astrocytoma	G2	31
Astrocytoma	G3	67
Astrocytoma	G3	44
Astrocytoma	G3	37
Astrocytoma	G2	50
Oligodendro.	G3	47
Astrocytoma	G3	39

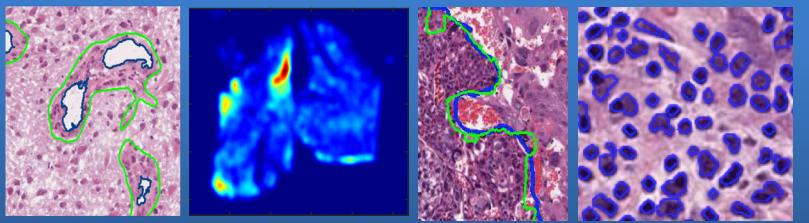
Genomics
(10,000+)



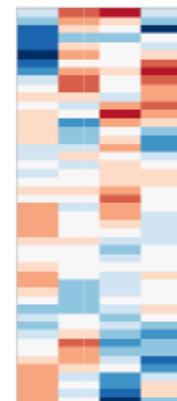
Radiology
(100M+)



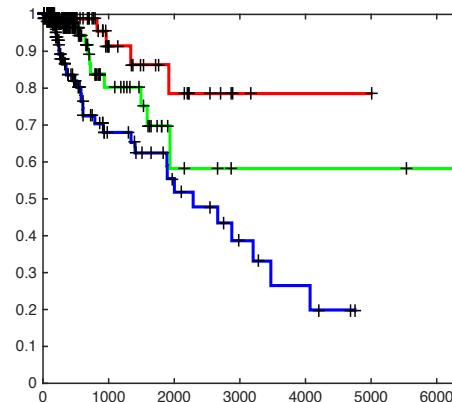
Pathology Data + Algorithms



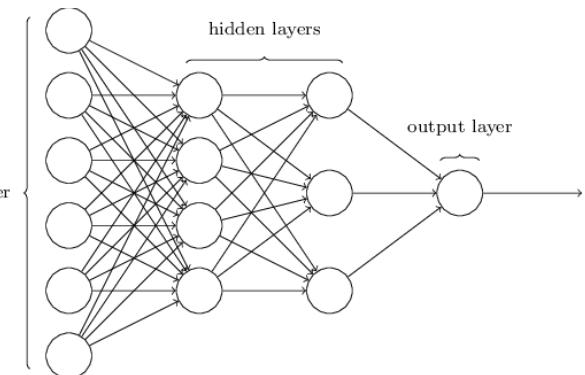
Features



Survival Predictions



Prognostic Models
Based on Deep Neural Networks & Bayesian Optimization



Clinical Decision Support

Surgery?

Radiation?

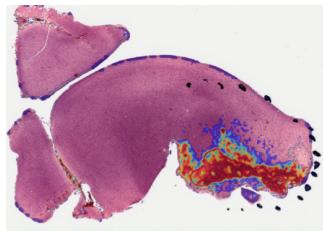
Chemotherapy?

Combination?

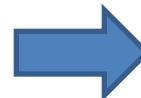
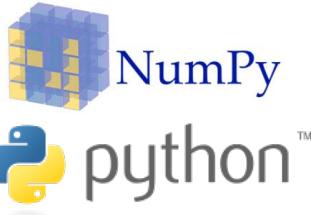
L.A.D. Cooper

HistomicsTK: Algorithms and Infrastructure for Pathology Image Analysis

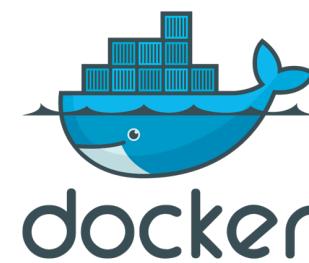
Image-Analysis Algorithms



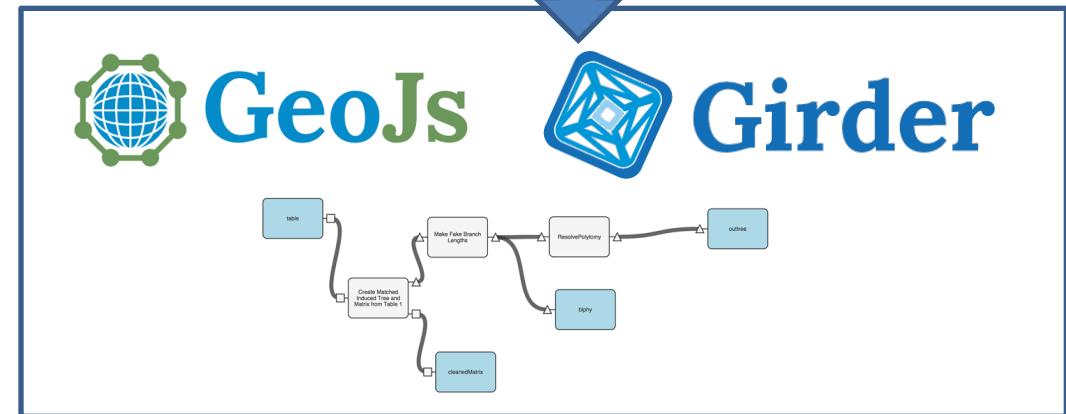
50K x 80K pix



Containerize and model metadata



Workflows, Execution, Database
for Terabytes of pathology
data and metadata



<https://github.com/DigitalSlideArchive/HistomicsTK>

20,0000 lines of code, 900 commits and growing

People

12 >



L.A.D. Cooper

Jun Kong



Information and Control Engineering
(Shanghai Jiao Tong University)

Electrical Engineering
(Shanghai Jiao Tong University)

Electrical & Computer Engineering
(Ohio State University)

Research Scientist and Sr. Research Scientist
(Emory)

Assistant Professor
(Emory)

Member of Winship Cancer Institute
(Emory)



EMORY
—
**WINSHIP
CANCER
INSTITUTE**

A Cancer Center Designated by
the National Cancer Institute



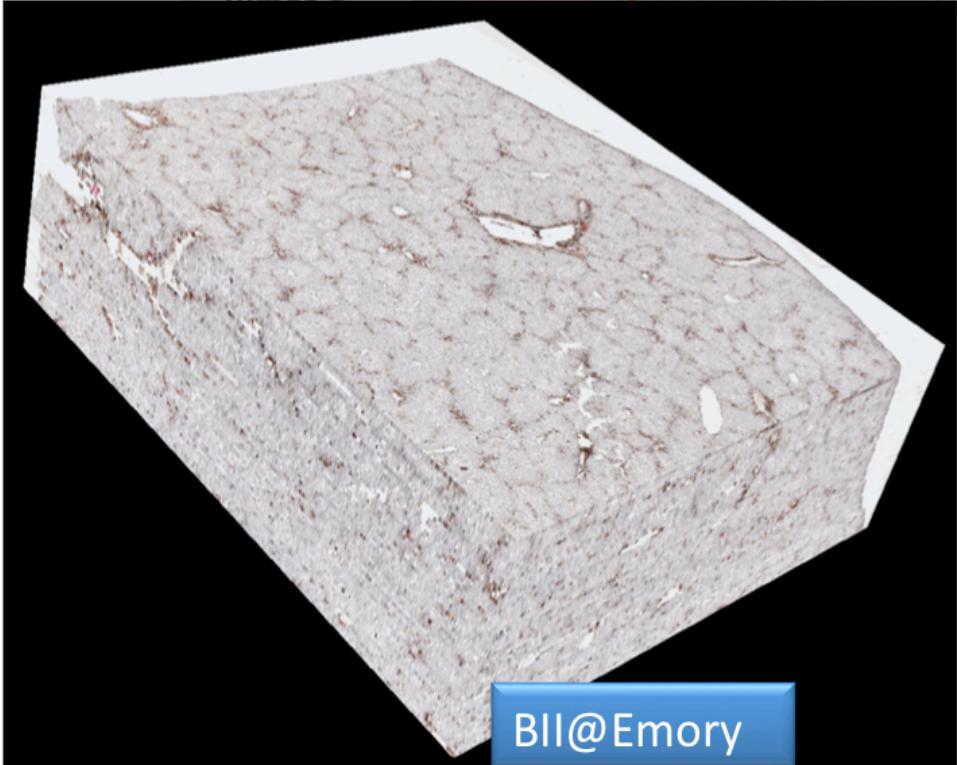
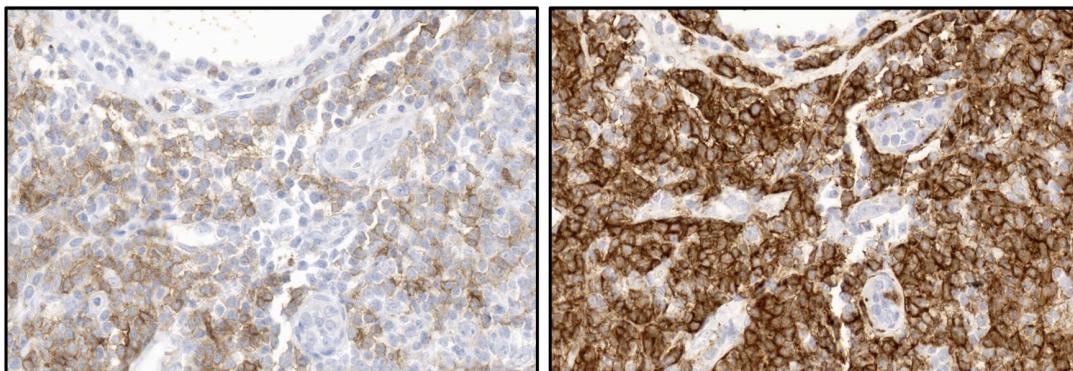
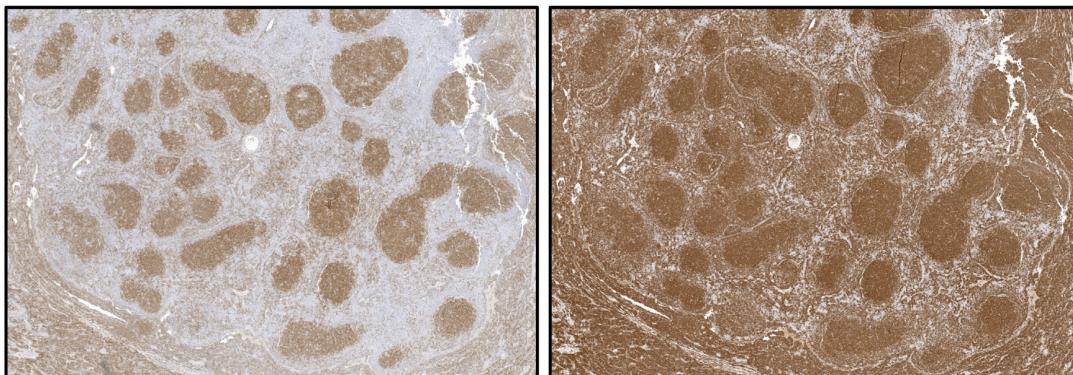
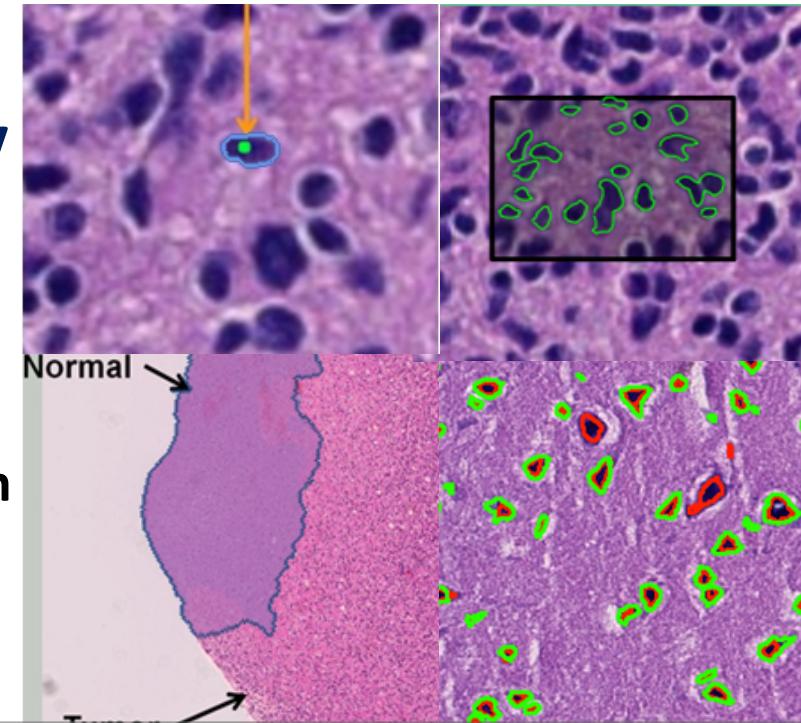
Big Histopathology Imaging Data – Currently @Emory Pathology

Volume – TB over the last few years

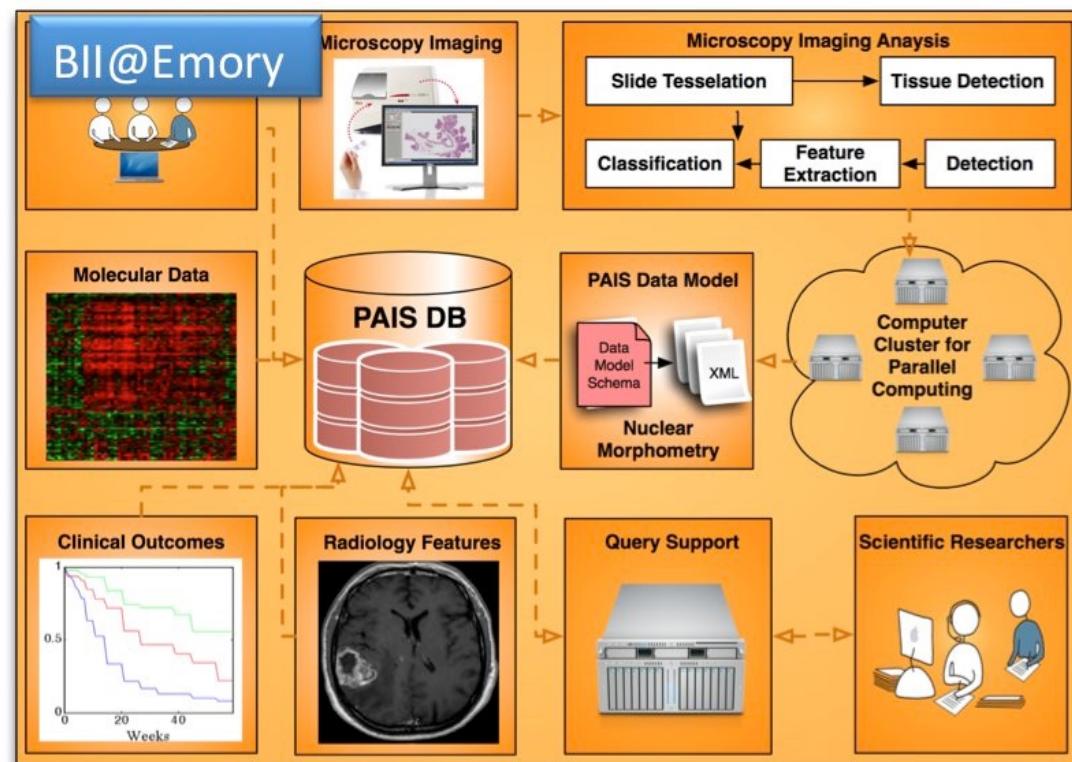
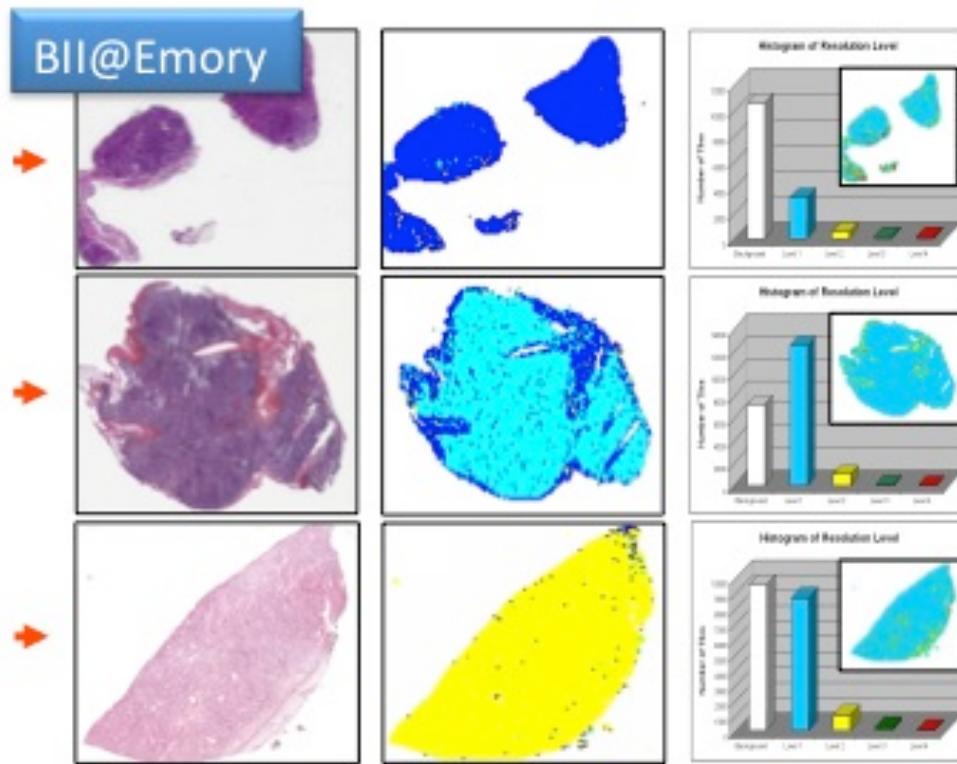
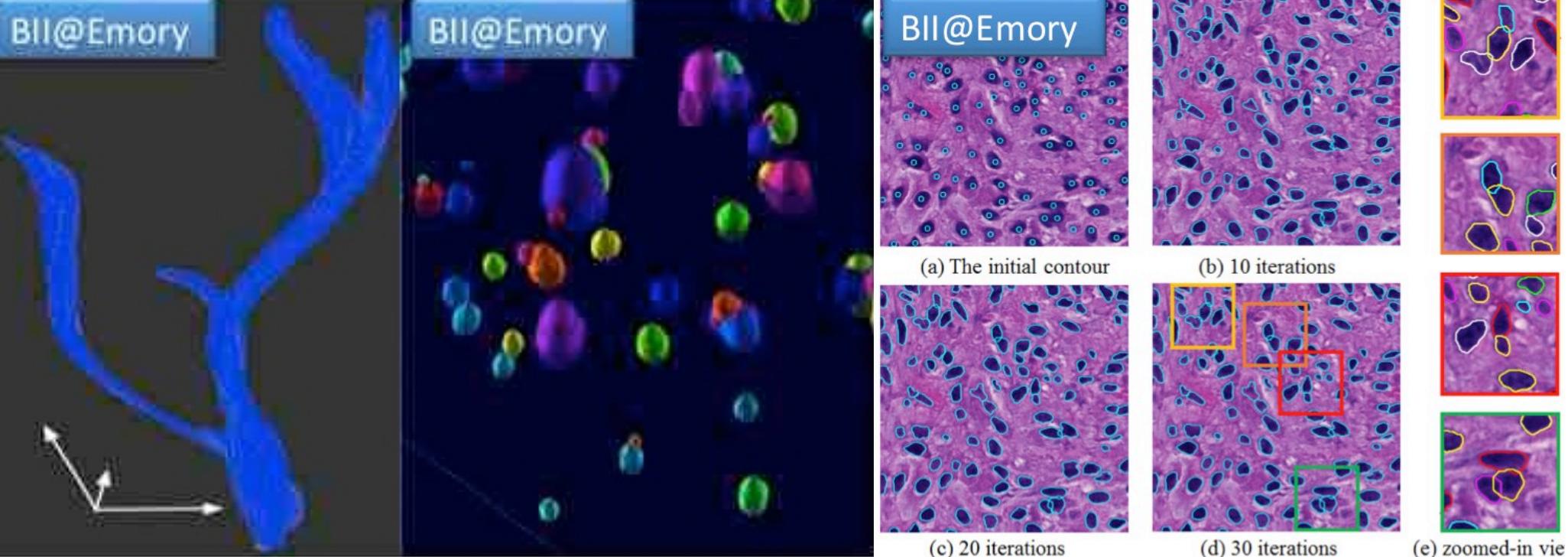
Velocity – 100 GB/day

Variety – Bright field (H&E, IHC), Fluorescence,
Confocal Imaging, Multiphoton, Electron
Microscope

Pathology structures – cells, vessels, etc.

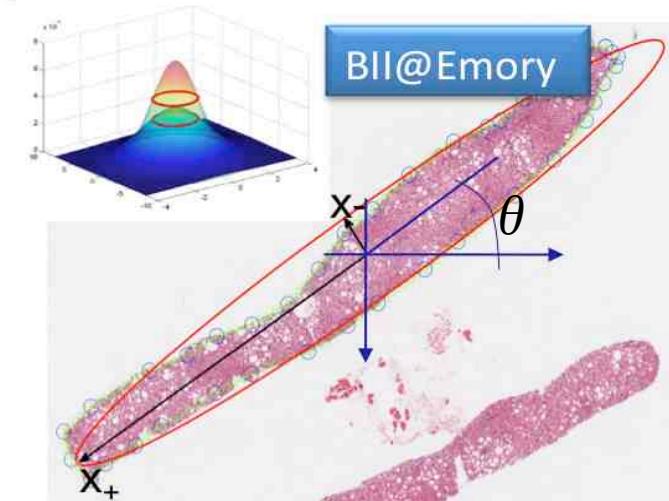
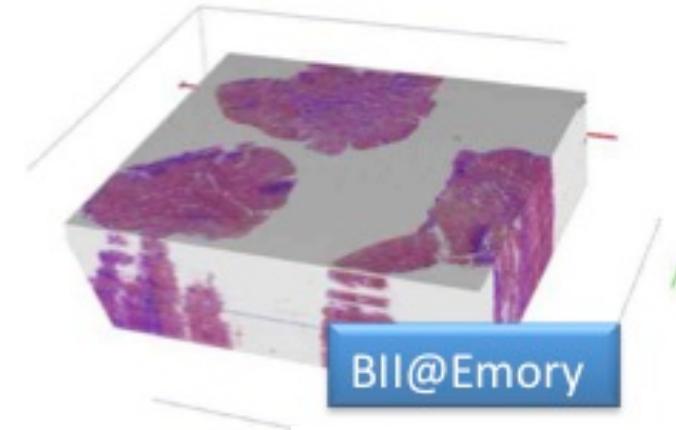


BII@Emory



Projects

- Live cell tracking with 3-D fluorescent microscopy images
- 2-D and 3-D pathology image analysis
- Histology and biomarker imaging integration
- Video analysis for pollinator tracking analysis
- Image analysis and machine learning for Computer-aided diagnosis
- 3-D pathology spatial analytics
- Deep learning classification with radiology image volumes
- Deep learning classification with 2-D and 3-D pathology images
- Geometric object representation



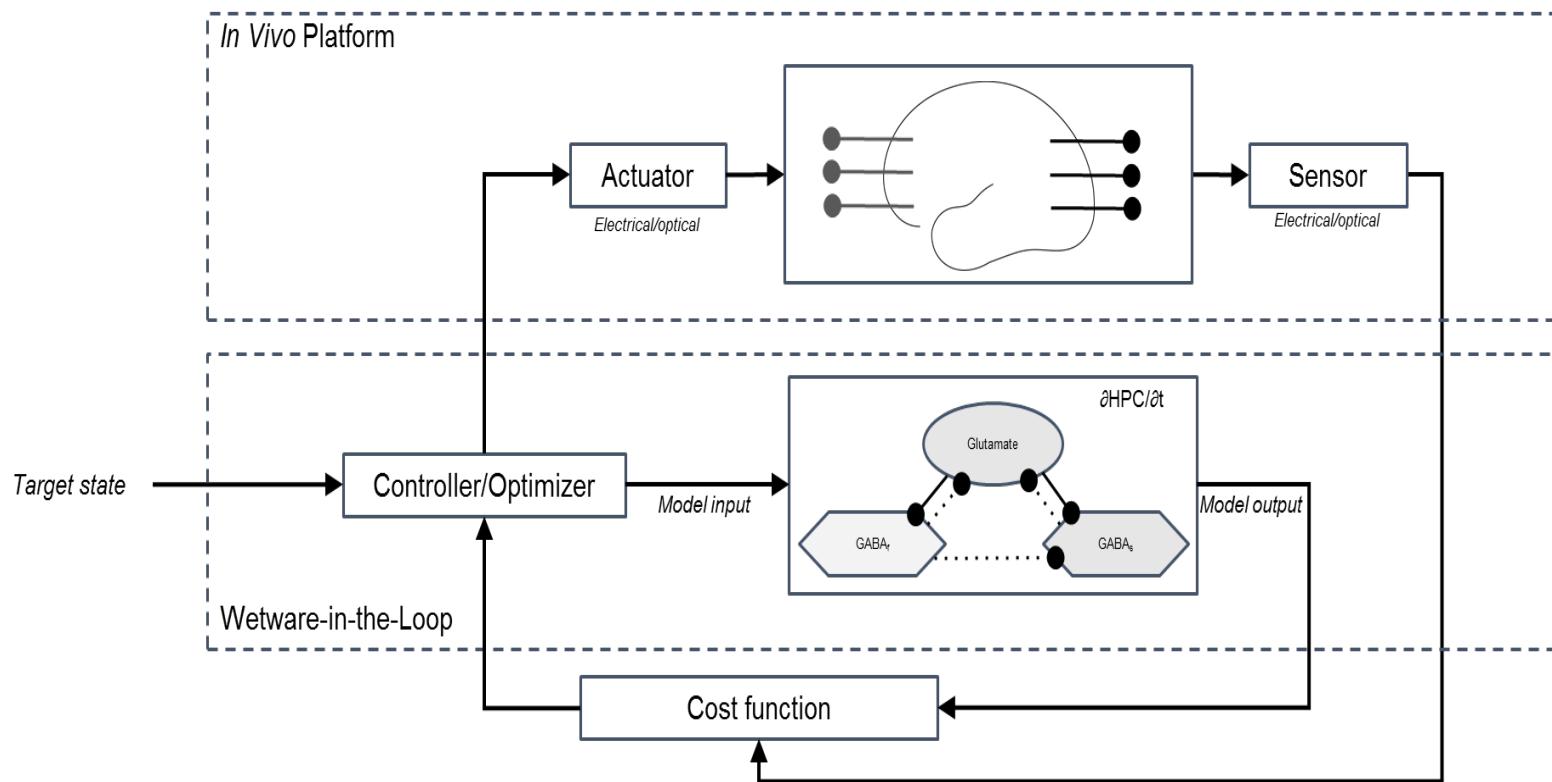
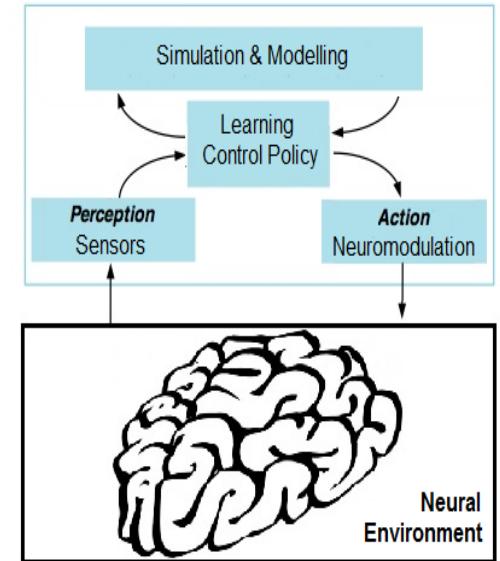
Babak Mahmoudi, PhD



- **BSc in Electrical Engineering (University of Tehran)**
- **MSc in Machine learning and Biomedical Signal Processing (Iran University of Science & Technology)**
- **MSc & PhD in Neural Engineering, Machine Learning and Artificial Intelligence (University of Florida)**
- **Postdoctoral fellowship in Translational Neural Engineering (Emory University & Georgia Tech)**
- **Clinical Research Scientist in Deep Brain Stimulation modelling (Emory University)**
- **Assistant Professor of Biomedical Informatics**

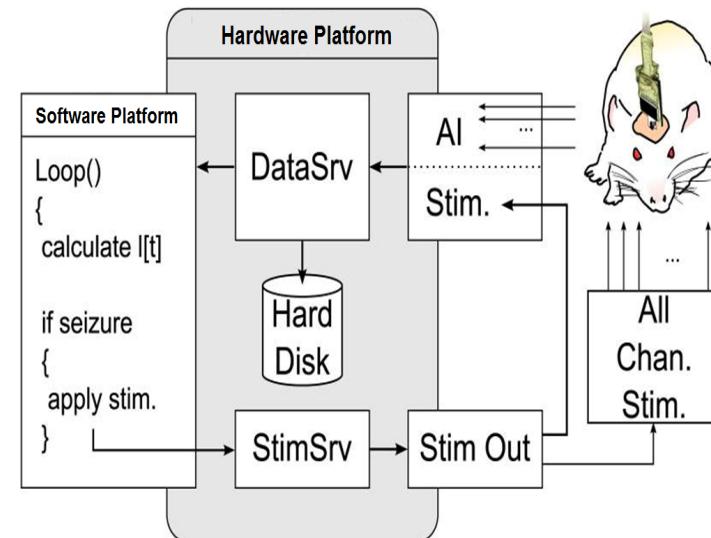
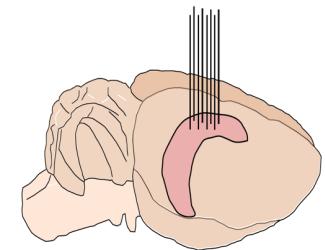
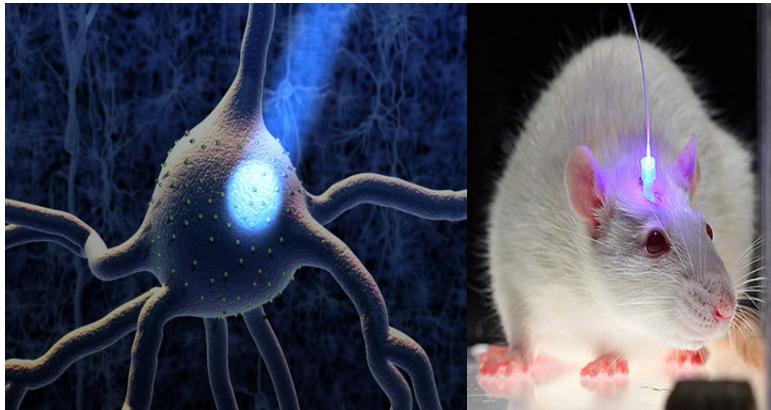
Neuroinformatics & Designing Intelligent Systems

- Neural Data Science
- Intelligent neural interface systems
- Simulation and modelling the neural networks



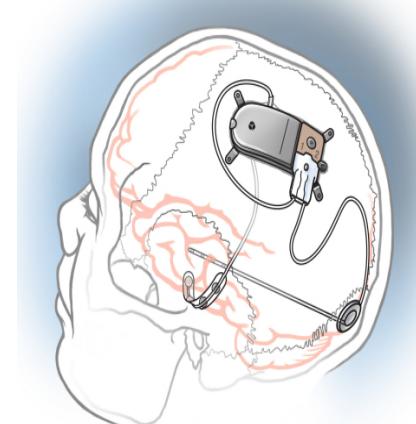
Experimental implementation and translational research

- Closed-loop distributed microstimulation
- Optogenetic neuromodulation



Clinical applications

- Neuromodulation therapies for epilepsy and memory enhancement
- Alzheimer's disease



Projects

- Developing machine learning and artificial intelligence algorithms
 - Decoding neural data for the identification of biomarkers
 - Optimization of neuromodulation therapies
- Mathematical modelling the neural dynamics
 - Biologically realistic models of neural circuits
 - Black-box modelling the input-output relationship between stimulation and neural response
- Implementation of the algorithms in experimental/clinical setting
 - Cloud-based computation
 - Mobile computing and developing apps
 - Real-time computation for closed-loop interaction with the brain

Shamim Nemati



**PhD in Electrical Engineering & Computer Science
(MIT)**
Postdoc in Machine Learning (Harvard)
Assistant Professor (Emory BMI department)
Adjunct Faculty (Emory CS department)
Program Faculty (Emory CSI program)

shamim.nemati@emory.edu
BMI Department Room#579
<http://nematilab.info>



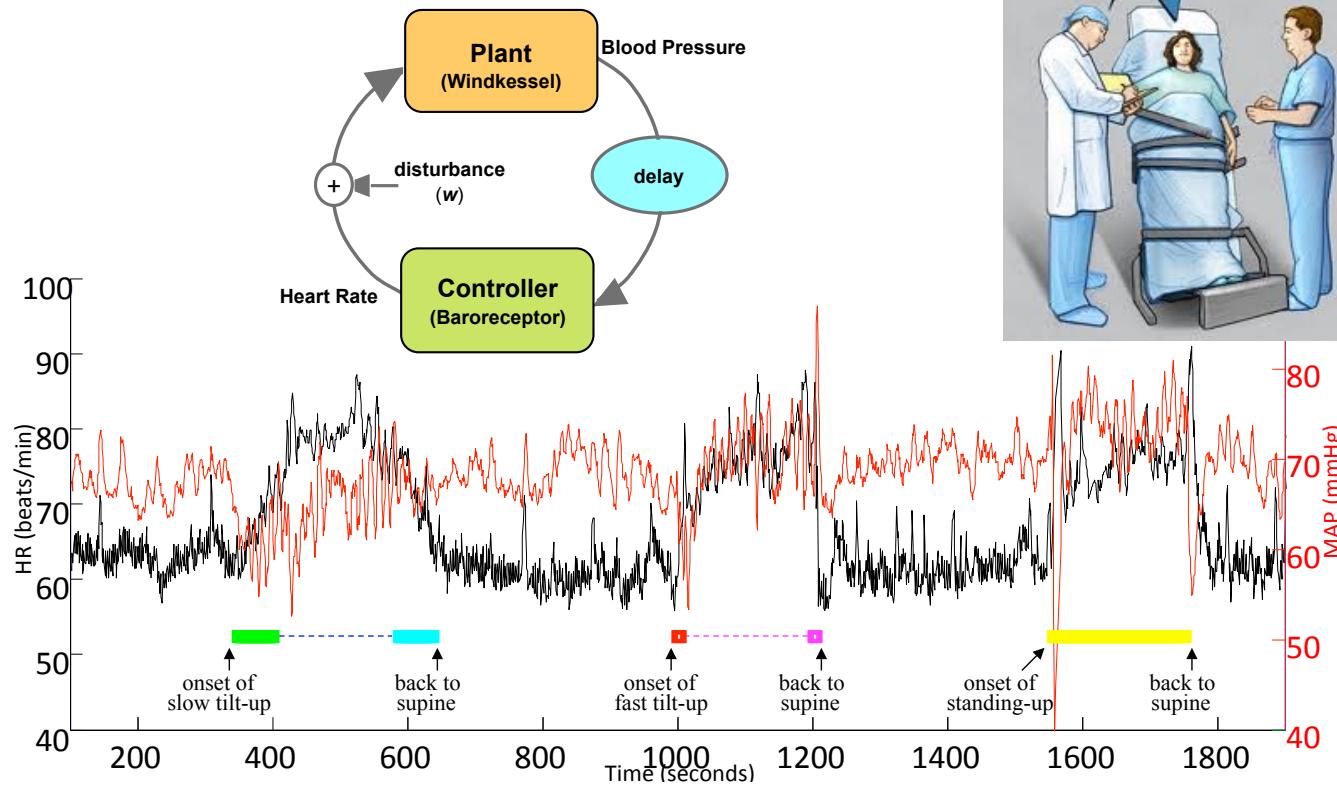
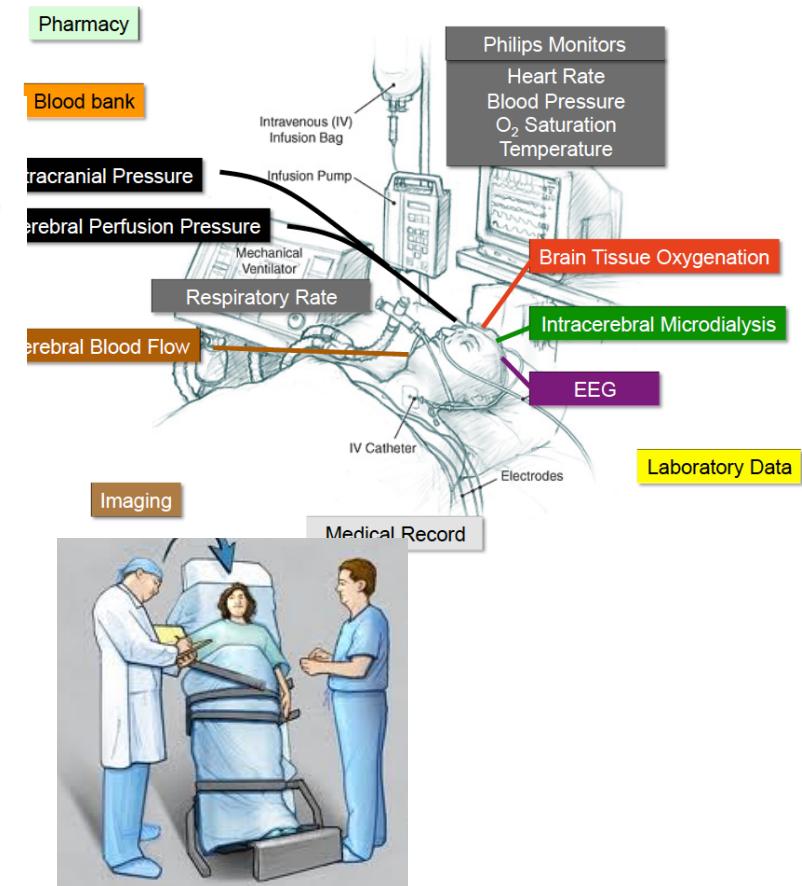
EMORY
UNIVERSITY

**Department of
Biomedical Informatics**
Emory University School of Medicine

Machine Learning for Physiological Time Series

Intensive Care Unit (ICU) Analytics

- Prediction of Impeding Adverse Events (e.g., Sepsis)
- Time series Prediction/Classification
- Model-based Machine Learning (ML)
 - Incorporating physiological models within ML algorithms



Nemati et al., J Appl Physiol, 2011, Nemati et al., IEEE EMBS, 2012

Deep Learning for Time Series Prediction & Classification

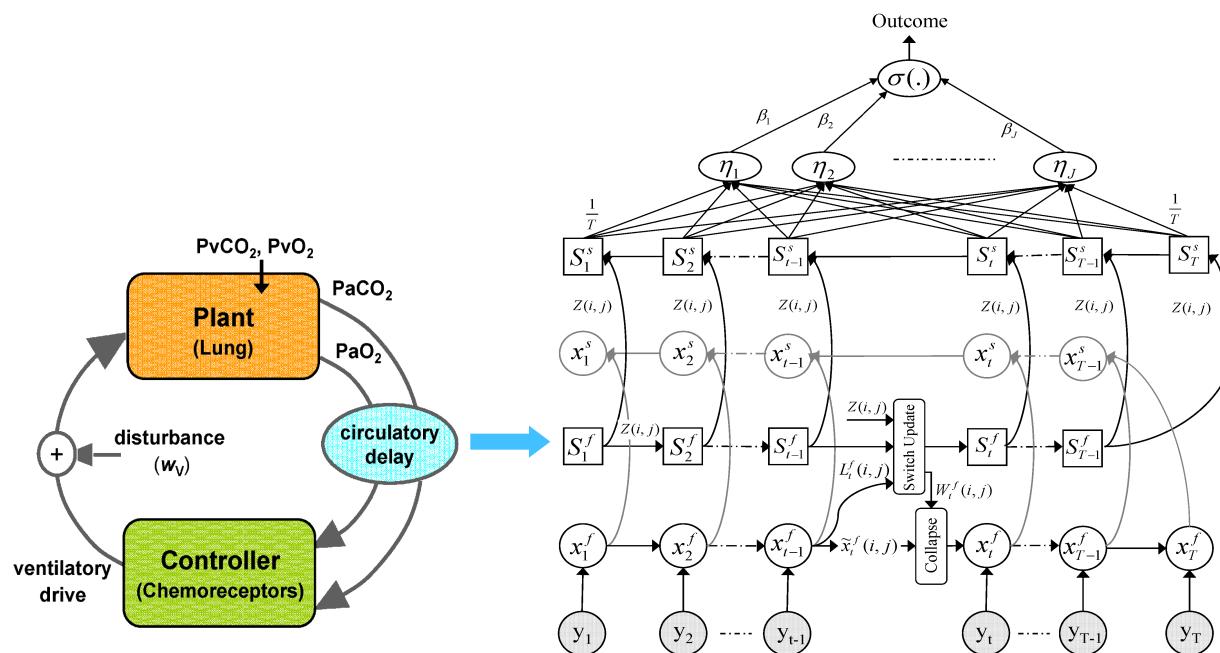
Intensive Care Unit (ICU) Analytics

- Prediction of Impeding Adverse Events (e.g., Sepsis)
- Time series Prediction/Classification
- Model-based Machine Learning (ML)
 - Incorporating physiological models within ML algorithms



Professors Clifford

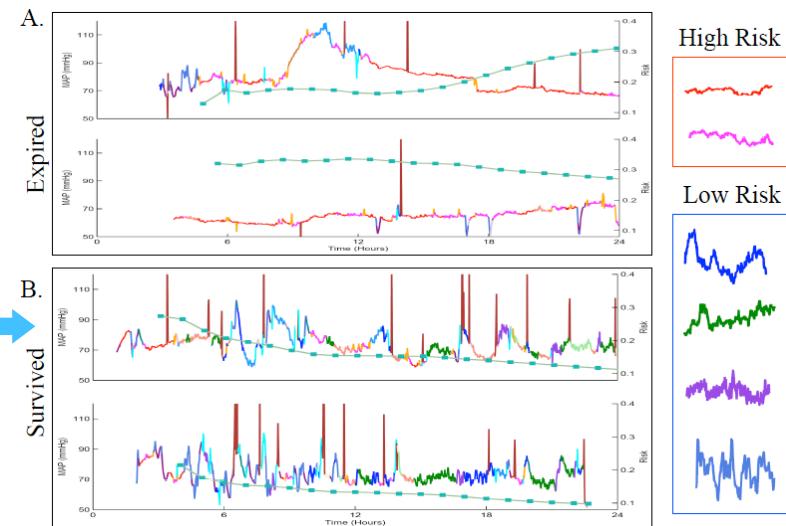
Professors Buchman



Physiological Models

Deep Learning for Time Series

Streaming Analytics



Finding Predictive Patterns in multivariate time series

Precision Medicine: Sequential Decision Making

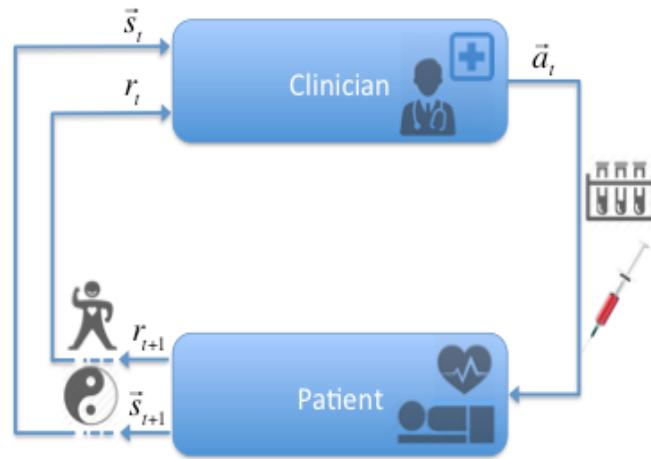


Figure 1. Sequential Decision Making for Medication Dosing in the ICU

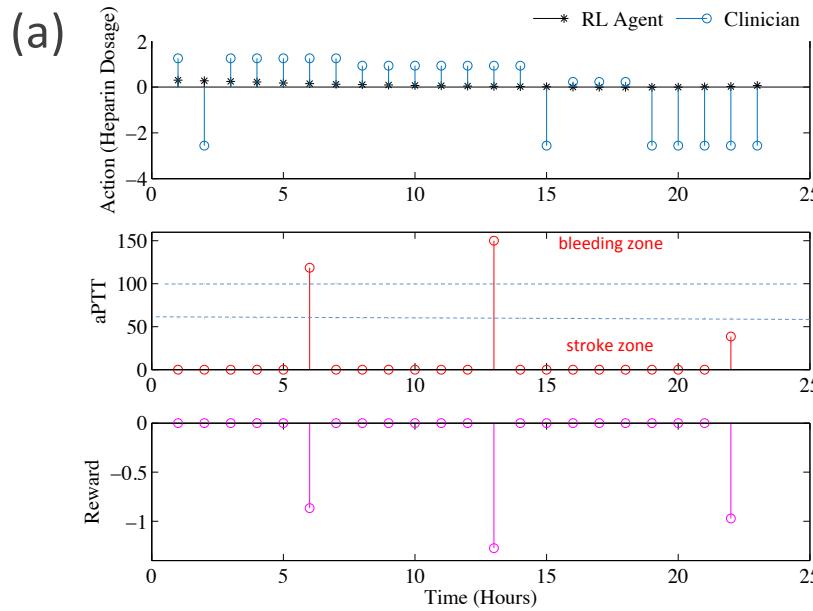


Figure 3. An examples of Heparin dosing (panel a), and population-level performance of the RL agent (panel b).

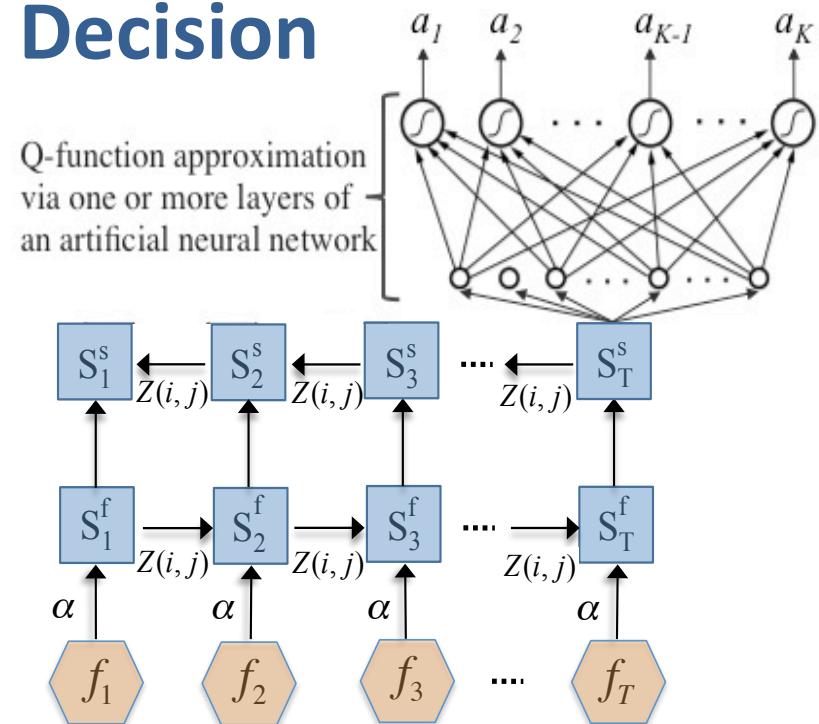
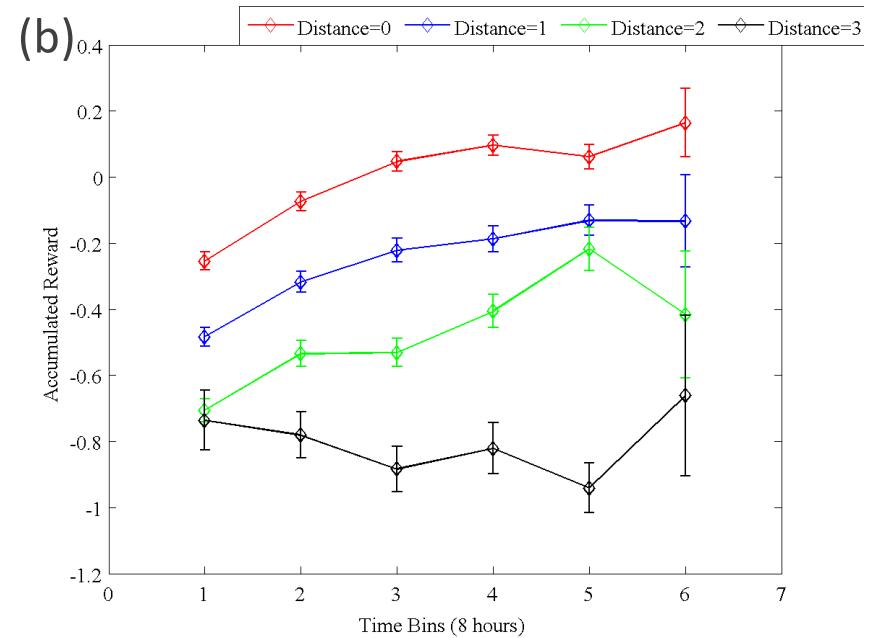


Figure 2. Q-network for sequential decision making.

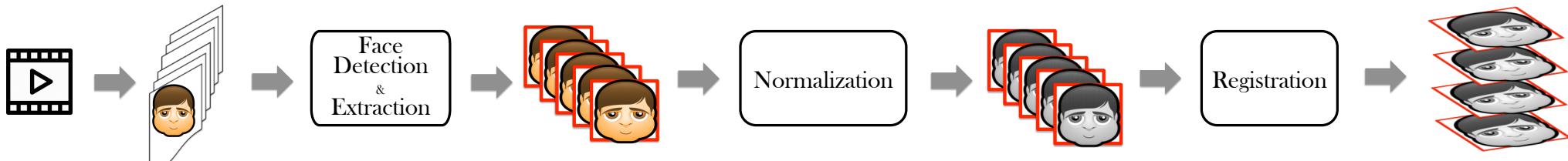


Computational Psychiatry

- Finding objective metrics of patient improvement
 - Video Analysis (dynamical systems, multiscale entropy)



Professors Helen Mayberg



Sahar Harati et al., EMBC 2016

- Physiological markers of stress & physical health

- Wearable technology (smart-watches and patches)

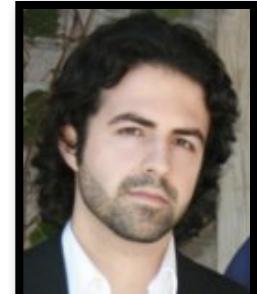
Nemati et al, EMBC 2016

- Audio and Text Analysis



Professors Clifford

Sentiment Analysis on Clinical Text



Dr. Ghassemi

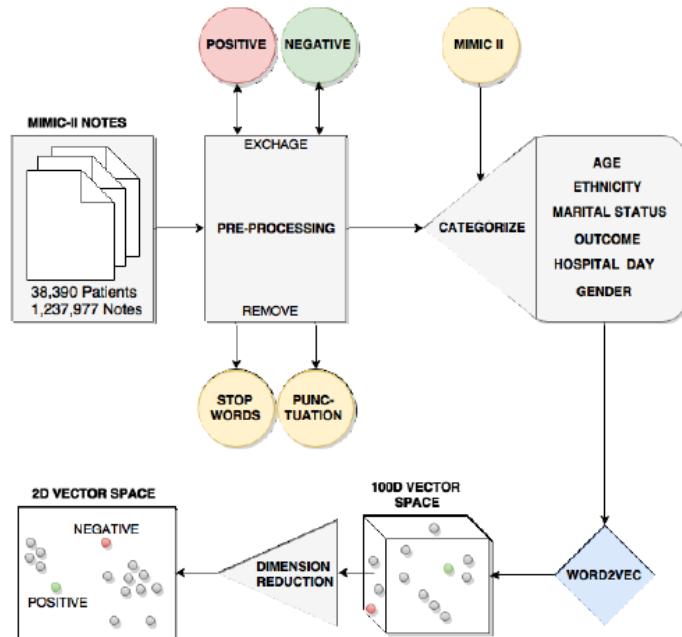


Figure 1. Schematic diagram of the proposed sentiment analysis technique

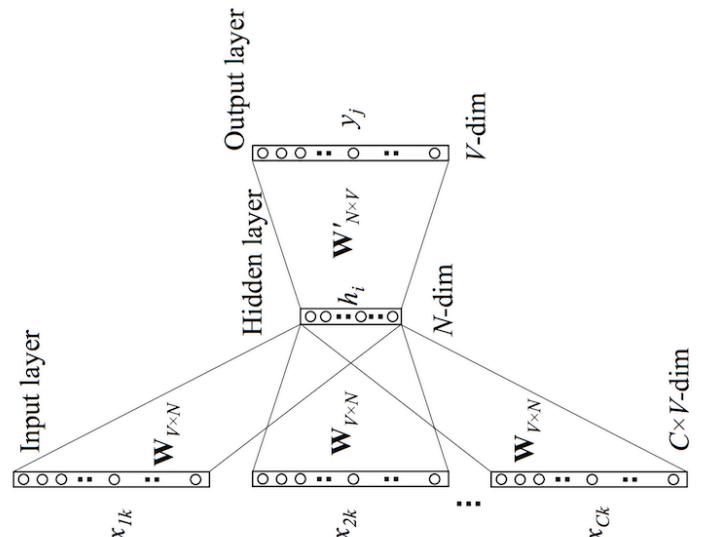


Figure 2. Schematic diagram of the proposed sentiment analysis technique

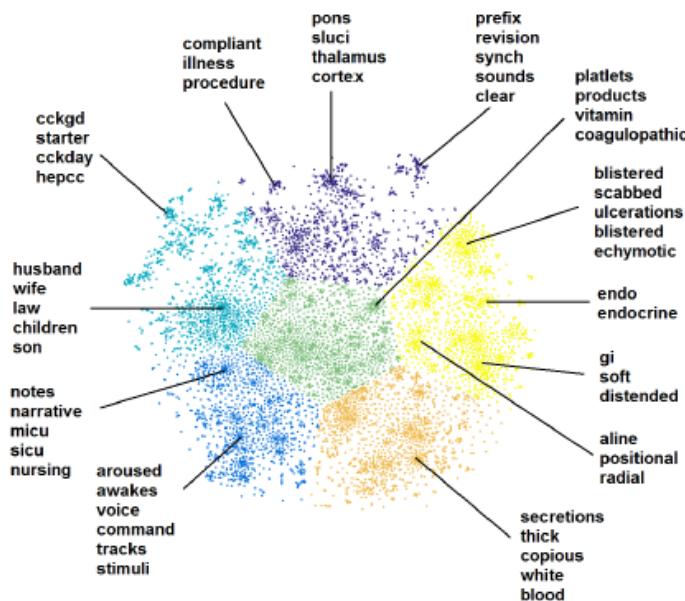


Figure 3. Clustering of patient data in the word2vec vector-space.

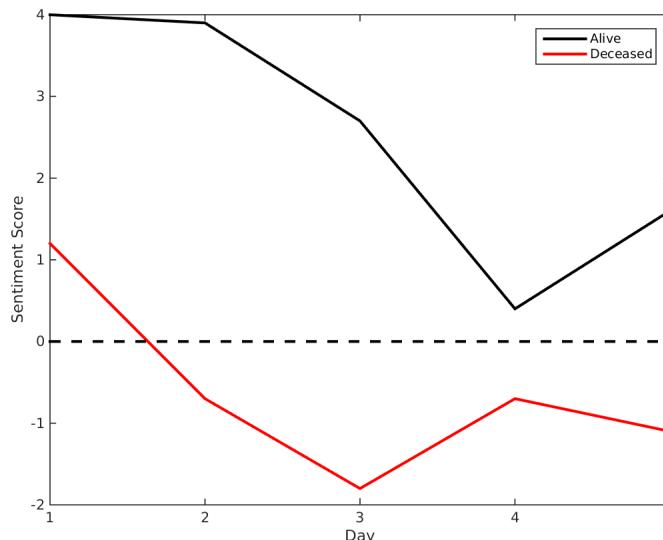


Figure 4. The sentiment score of the notes, partitioned by outcome class, and hospital stay day.
Ghassemi, Nemati, Cinc (2015)



Dr. Falgun Chokshi

Andrew Post



MD (Univ. of Pennsylvania)

PhD in Biomedical Informatics (Pitt)

Associate Professor (Emory)

**Director, Biomedical Informatics Program, Atlanta
Clinical and Translational Science Institute**

Clinical Informatics Architect (Emory Healthcare)



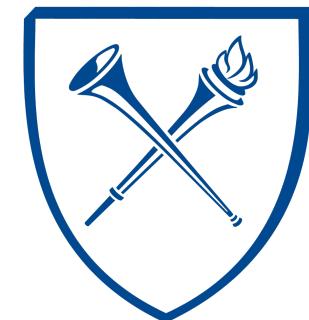


- ◊ Imaging Informatics
- ◊ CDW Data Reports
- ◊ Database Design & Support
- ◊ Data Management
- ◊ Cloud Design & Consulting
- ◊ Electronic Case Reports
- ◊ Training & Consultations
- ◊ Laboratory IS Support
- ◊ Clinical Research Databases
- ◊ Programming Support



Biomedical Informatics Program

EMORY



Emory i2b2 – Cohort Discovery

“I Want to Recruit Patients with Rheumatoid Arthritis...”



i2b2 Query & Analysis Tool Project: i2b2 Demo User: Shrine_i2b2 Find Patients | Analysis Tools | Help | Change Password | Logout

Query Tool

Query Name: Rheum-Femal- 35-@09:45:42

Temporal Constraint: Treat all groups independently

Group 1: Dates Occurs > 0x Exclude
Treat Independently
Rheumatoid arthritis and other inflammatory polyarthropathies
one or more of these

Group 2: Dates Occurs > 0x Exclude
Treat Independently
Female
one or more of these

Group 3: Dates Occurs > 0x Exclude
Treat Independently
35-44 years old
45-54 years old
one or more of these

AND AND AND

Run Query Clear Print Query 3 Groups New Group

Show Query Status Graph Results

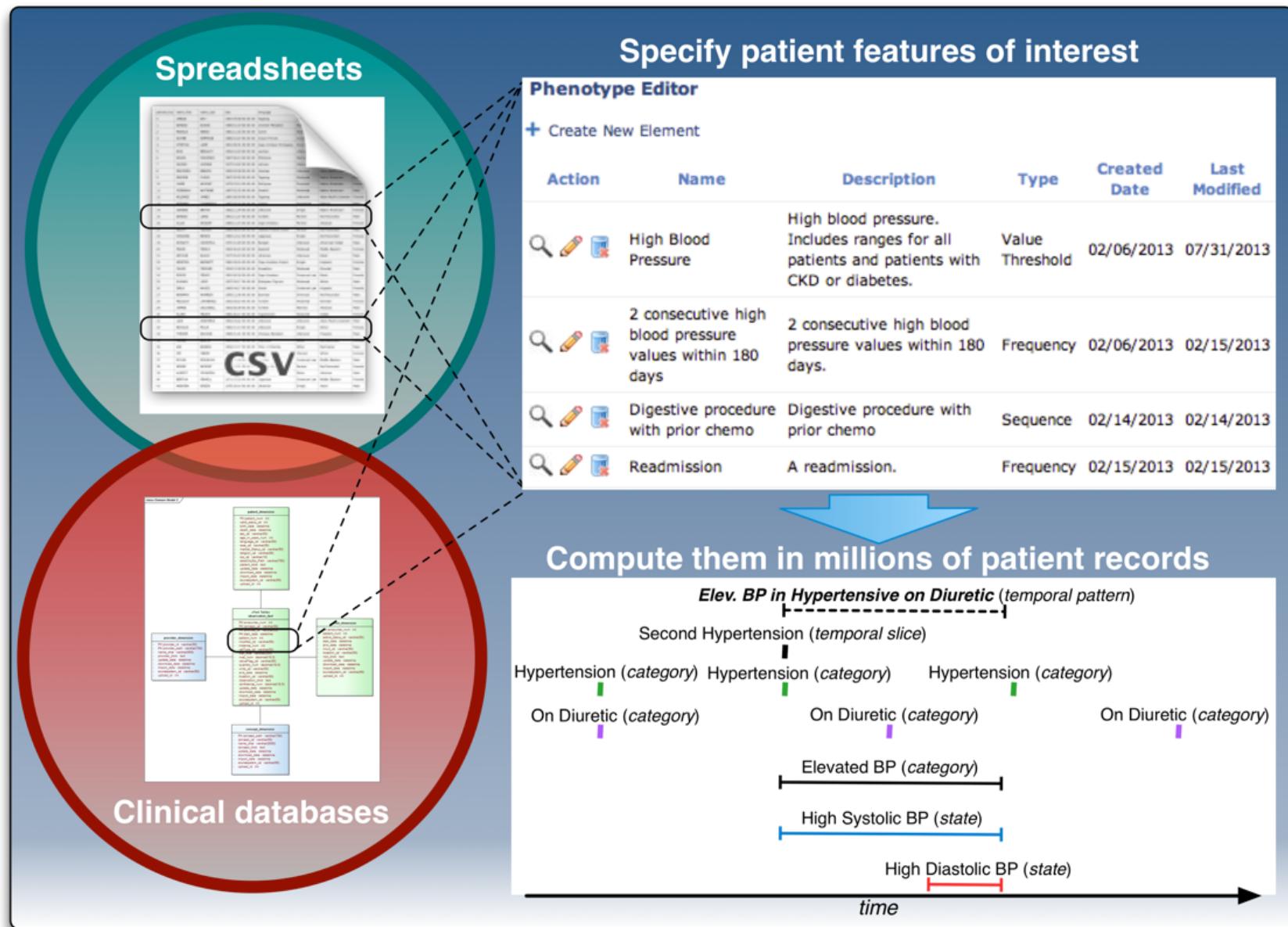
Number of patients
1687
For Query "Rheum-Femal- 35-@09:45:42"

Navigate Terms Find

ACT Demographics
0.5 Version
Age
0-9 years old
10-17 years old
18-34 years old
35-44 years old
45-54 years old
55-64 years old
65-74 years old
75-84 years old
85-89 years old
>= 65 years old
>= 85 years old
>= 90 years old
Not recorded
Hispanic
Race
Sex
Vital Status
ACT Diagnoses
ACT Laboratory Tests
ACT Medications
ACT Procedures
ACT Visit Details

Eureka! EHR Phenotyping

“I Want to Create a Predictive Model of Hospital Readmissions...”

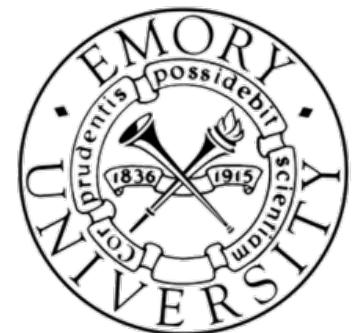


Ashish Sharma



Electrical Engineering (PEC, India)
PhD Scientific Visualization and Data Mining
(University of Southern California)
Postdoc Biomedical Informatics (Ohio State)
Research Scientist (Ohio State)
Assistant Professor (Emory)

Ashish Sharma (ashish.sharma@emory.edu)
Psychology Building — #573
<http://sharmalab.info>



Research Data: everywhere & every which way

Large, Diverse, Heterogeneous, Noisy

- Explore this data
- Integrate (with other datasets)
- Analyze (increasingly on the cloud)
- Visualize

The screenshot shows the TCGA homepage with several overlapping windows:

- Top Left Window:** "Four Subtypes of Stomach Cancer Identified". It features a circular diagram of stomach cancer subtypes and a summary: "Recent finds with the TCGA Research Network have found that stomach cancers, also called gastric cancers or gastric adenocarcinomas, fall into four distinct molecular subtypes."
- Top Middle Window:** "Enrich Data Portal". It provides a platform for measurement of enriched genes, download, and analysis. It notes: "The Cancer Genome Atlas (TCGA) Data Portal provides a platform for measurement of enriched genes, download, and analysis. Data were generated by TCGA."
- Top Right Window:** "TCGA Data Portal Overview". It describes the portal's mission: "The Cancer Genome Atlas (TCGA) Data Portal provides a platform for measurement of enriched genes, download, and analysis. Data were generated by TCGA. This portal is a central resource for cancer genomics information. Please visit the 'About' page for more information."
- Bottom Left Window:** "TCGA In Action". It highlights recent news: "July 2014: Steps Towards Precision Medicine Using FFPE Specimens for Comprehensive Genomic Characterization", "Leadership Update", and "TCGA In Action".
- Bottom Middle Window:** "Available Cancer Types". A table lists cancer types with counts of genes analyzed, samples analyzed, and last updated date.
- Bottom Right Window:** "Software release". It announces: "The TCGA team successfully completed their software update released for today. Details about the update can be found on the TCGA website. If you have concerns about the release, please email us at tca@nih.gov".

Footer:

The Cancer Genome Atlas Understanding genomics to improve cancer care cancergenome.nih.gov/

Data Integration and Cloud-Based Analysis

- **Data Café** — Dynamic data warehousing that allows one to connect disparate datasets. Highly scalable. Adding some intelligence.
- **Cloud-Pipe** — Cloud based analysis and workflows. Designed for quantitative imaging. Ease of use.
- **Bindaas** — Middleware system developed in our lab for easy development of secure and robust web-services.

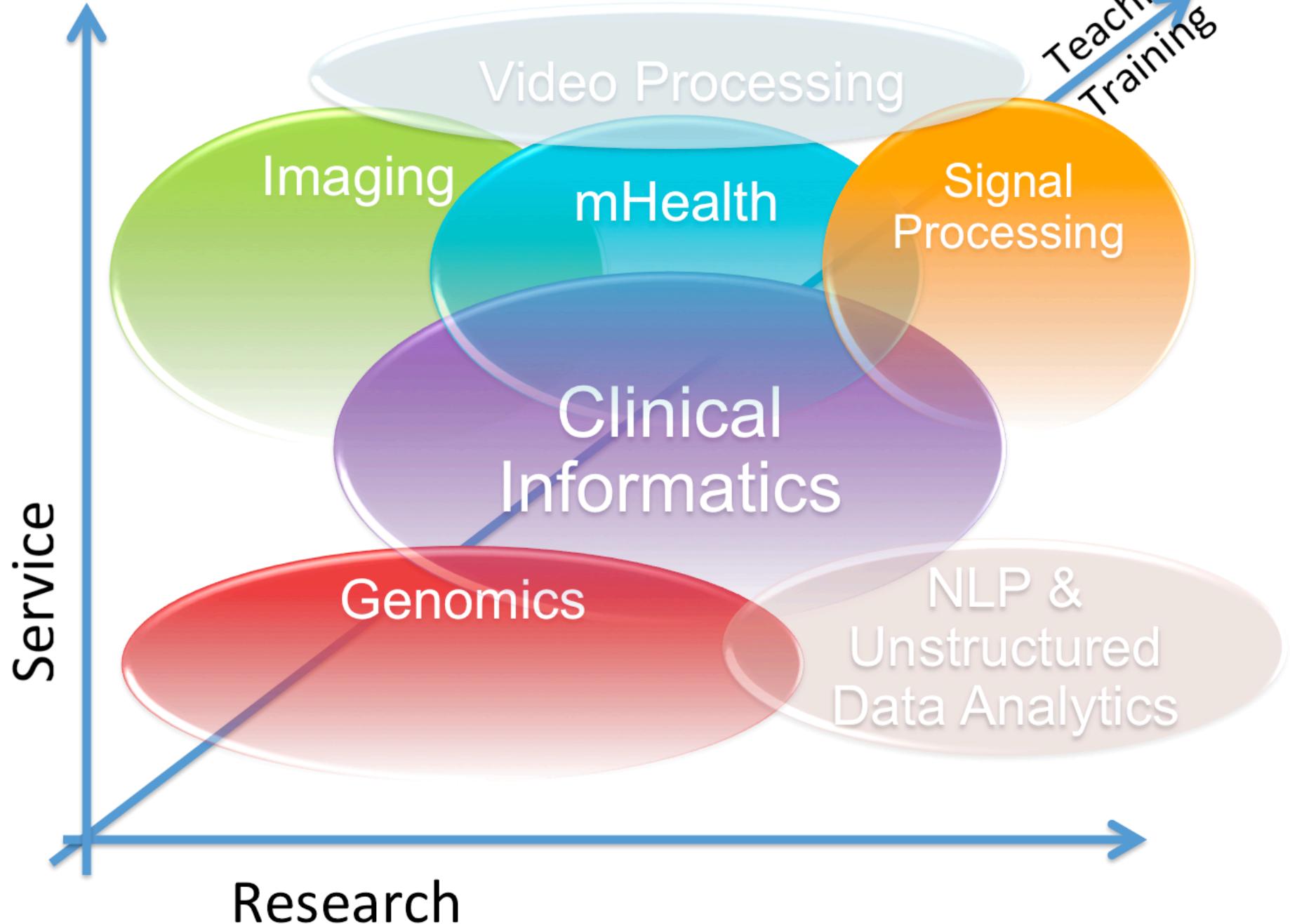
End-User: Quantitative Imaging
for Cancer Research

Data Exploration and Visualization

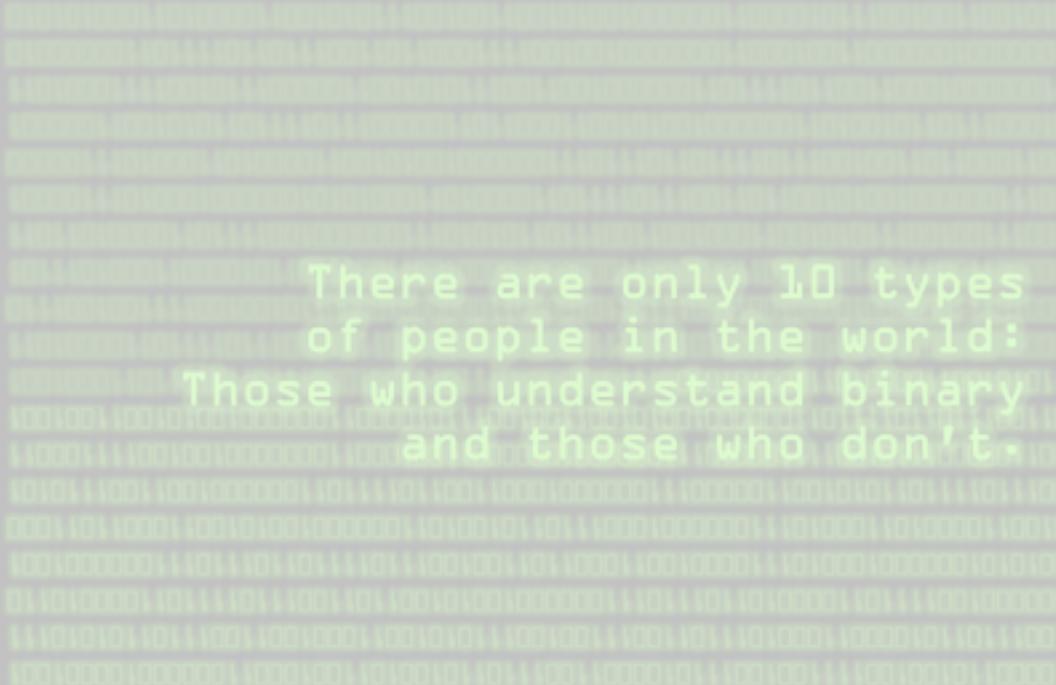
- **DataScope** — Visual exploration of integrated datasets. Interactive; Fuse multiple sources – scientific mashups; Geo-spatial data support; Scalable;
- **caMicroscope** — Platform for visualizing, exploring and analyzing digital pathology data.

End-User: Co-Clinical Trials; Cancer Research; Public Health Informatics.

What is BMI @ Emory?



BMI @Emory – Basic Things about working with us



There are only 10 types
of people in the world:
Those who understand binary
and those who don't.



EMORY
UNIVERSITY

**Department of
Biomedical Informatics**
Emory University School of Medicine

Things to know before you start

Expectations:

<http://gdclifford.info/ethics.html>

(Please read and email me to acknowledge or protest)

Key things:

- Plagiarism / Collaboration / Ethics
- Attendance (not optional – evidence-based excuses in advance please)
- Licensing/ownership of code ... public vs private
- Presentations, authorship, patents and COI's

Backing up code, data and docs / SCC

Everything we present here is on github – learn that now:

<https://guides.github.com/activities/hello-world/>

That's where your course materials are located!

Backup your (course)work in real time:

- Onedrive (1TB with office365 account)
- Google Drive (good for collaborative docs)
- Emory.box.com (100GB HIPAA compliant)
- Overleaf.com – online collaborative Latex system
- Mendeley – collecting and sharing references
- Github – Code (but keep it private and share it with your supervisor)
- AWS - \$100 for cloud computing ☺

Lab Rotations

(finding a group and a topic you like)

- 1 semester or summer in first 21 months.
- Intensive research topic leading to a publication (we hope)
- You will then sit with our research group and be expected to work 6-8 hours a day (40 hours a week) on a research topic, attend lab meetings, and give a final presentation/hand in a 4-12 page conference-style article on the work.
- Advise doing rotations with us during the first summer (preferably) or a subsequent fall or spring semester in their second year, when they are not as busy - i.e. don't bother doing a rotation if they are doing a high course load that semester.

Expectations / research management

- It's a 2 way exchange between your mentor and you, but you need to manage your own PhD. **Be persistent** (but not annoying)
- Moving to independence and taking control of your research (eventually)
- Go from weekly to month meetings as time progresses
- You are part of a group – pay it forward / help out
- End goals:
 - Excellent rewarding job at the end of the PhD,
 - Several journal publications (to ensure you pass),
 - Impactful research (so the money spent on you was worth it to society and the research can continue to be funded)
 - Have fun and learn that hard work is rewarding – it's about the journey

Further reading:

- Richard Hamming, **You and Your Research'**, Bell Communications Research Colloquium Seminar 7 March 1986:
www.cs.virginia.edu/~robins/YouAndYourResearch.html
- Nick Feamste, **Managing your Advisor**, August 14, 2013
<https://greatresearch.org/2013/08/14/managing-your-advisor/>
- <http://blogs.berkeley.edu/2013/11/15/how-to-build-a-bad-research-center/> or
<https://goo.gl/f4aXJv>
- <https://people.eecs.berkeley.edu/~pattrsn/talks/BadCareer.pdf>

Course Overview/Timetable

	Date			Taught by
Wk. 1	8/29	Introduction to Biomedical Informatics:	Course overview; History; Applications; Interdisciplinary nature; informatics in action — case studies	Clifford & Sharma
Wk. 2	9/12	Data Representation, Standards and Management	Role of Interoperability; Data exchange and management; SNOMED-CT; UMLS; Domain/Problem specific standards & management	Post
Wk. 3	9/19	Artificial Intelligence in Medicine	From perception to sequential decision making; deep learning and reinforcement learning applications in clinical decision support systems.	Nemati
Wk. 4	9/26	Security, Privacy and Confidentiality	PHI, HIPAA; Fundamentals — Cryptographic Hashes; Encryption (symmetric vs public key); secure information exchange; relevant standards and policies; Authentication, authorization and RBAC; OpenID, OAuth 2.0; OpenID Connect (http://goo.gl/Nua0rE)	Xiong
Wk. 5	10/3	Electronic Health Records CPOE and CDS	key components of the modern EHR; standards that make EHR possible; Pros & Cons of modern EHR systems; patient engagement	Post
Wk. 6	10/17	Imaging Informatics	Fundamentals of Imaging (Rad, Path, RT); Data Standards (DICOM incl. WG26-Pathology), Data Management (PACS, RIS, VNAs, LIMS), Common acquisition modalities (CT, MR, PET) Lab: use 3D slicer and TCIA/Dcm4chee; DicomWeb-REST	Sharma
Wk. 7	10/24	Quantitative Biomedical Imaging Processing	Fundamentals of image processing, signal processing, pattern recognition; Video analysis, 2D & 3D Pathology Image analysis. Lab: processing images with fundamental image processing filters	Kong

Course Overview/Timetable

Wk. 8	10/31	Quantitative Imaging and Radiology	Radiomics; Case Studies – RIDER, LIDC, QIN (Aerts. et. al.)	Cooper
Wk. 9	11/7	Neuroinformatics & Signal processing	Brain Health, Neuropsychiatric diseases, Sleep, Behavioral phenotyping (Audio and video), Neurophysiological time series, Neural interface technologies, Neural decoding and neuromodulation, modelling and computational neuroscience,	Mahmoudi
Wk. 10	11/14	Critical Care Informatics	Cardiovascular physiology,integration of time series data at different sampling frequencies,	Clifford and Nemati
Wk. 11	11/21	Cloud Computing and Rapid Search on Terascale Clinical Databases	Big Data in the age of precision medicine; Introduction to Hadoop; AWS; Cloud Based Data Warehousing (netezza vs redshift vs teradata)	Sharma & Blum
Wk. 12	11/28	'omics	Analysis of molecular data including genomics. Case studies in cancer for optimal classification of patients and identification of driver alterations. TCGA and other Public Resources.	Cooper and Qin
Wk. 13	12/5	mHealth Informatics and evaluating/monitoring systems	Opportunities in mHealth, Global Health, Reverse Innovation, Ethics, Barriers and Standards, (Research Kit, Research Stack, HL7, SMART on FHIR 510(k) predication, after market vigilance) App Inventor Lab - Apps for Dummies.	Clifford

Grading

- Each week you will be graded /10
- Final mark is average of all grades
- No properly excused absence (job interview or family illness) = 0 for the week
- Leaving early or arriving late -20%
- Handing in late = -10% for each day or part of day that passes
- Partial credit for partially completed assignments
- No credit for corrupt PDFs
- We won't be chasing you
- No excuse for not achieving 90-100%