Applied Math for Bioimaging

What is she up to @BIDS in the past 5 years?

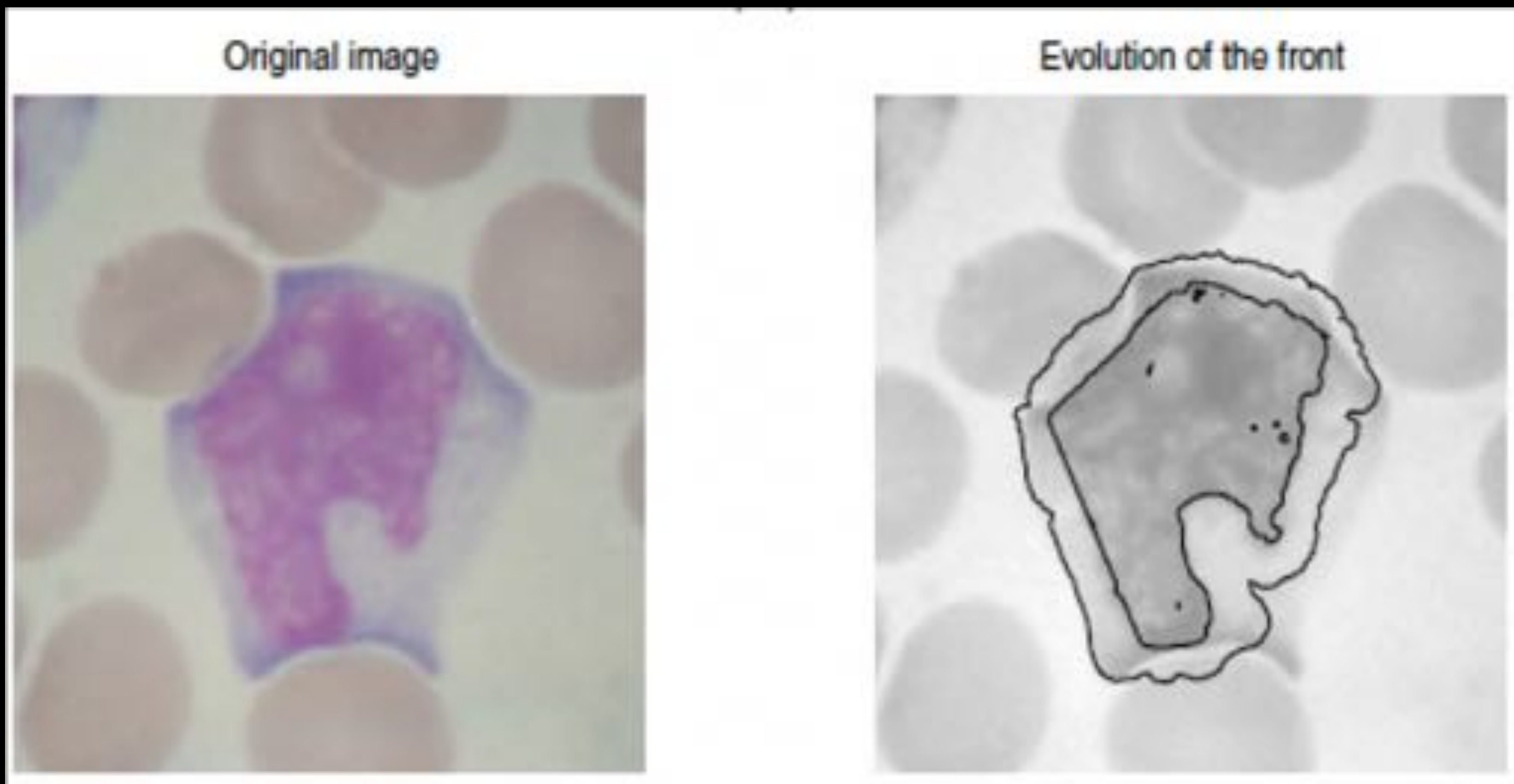
Dani Ushizima, Ph.D.

Staff Scientist
Computational Research Division
Berkeley Lab

Data Scientist
Berkeley Institute for Data Science
UC Berkeley

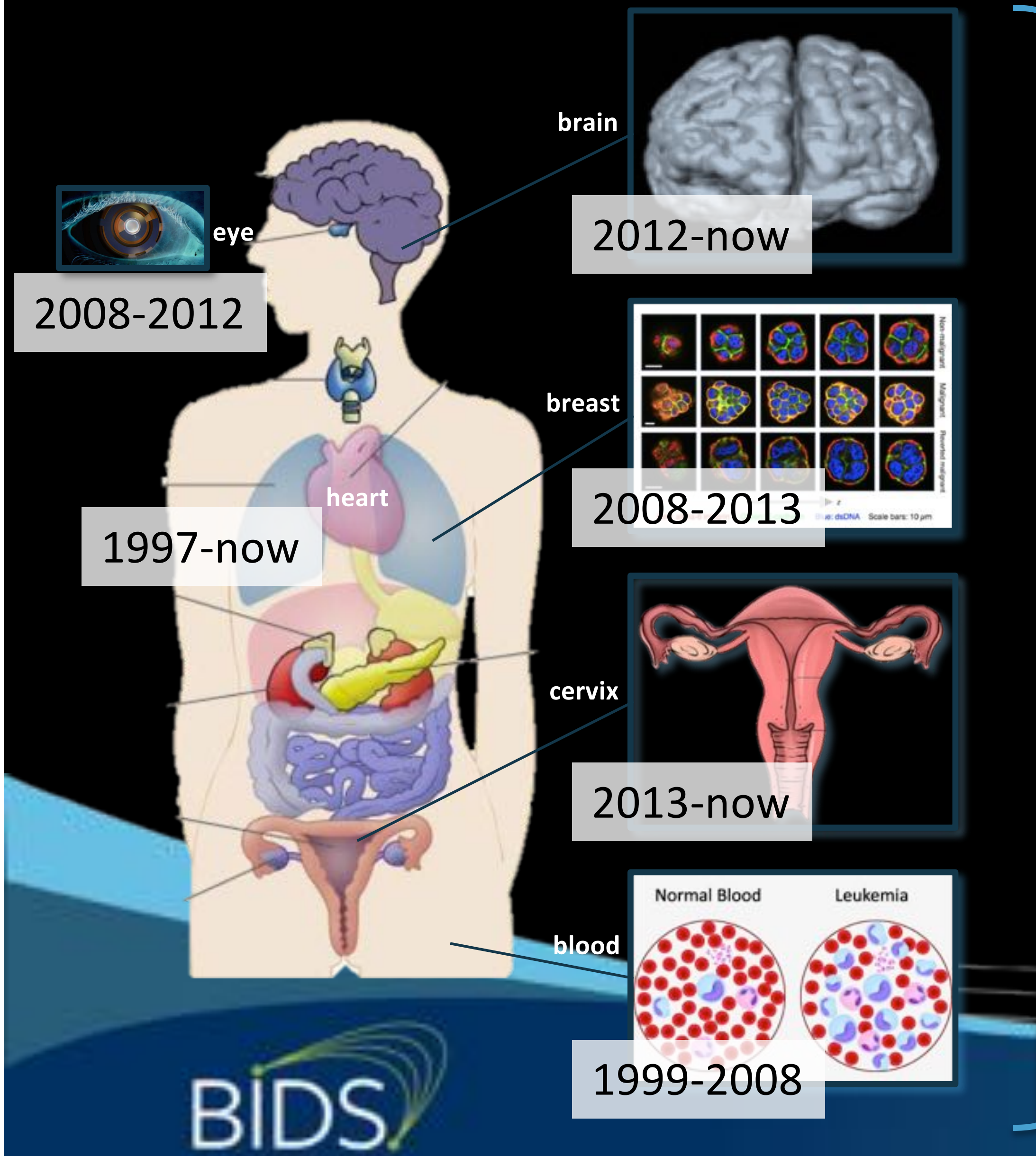
1999

Cybernetic Vision Group



Prolymphocytic Leukemia

Computational health science



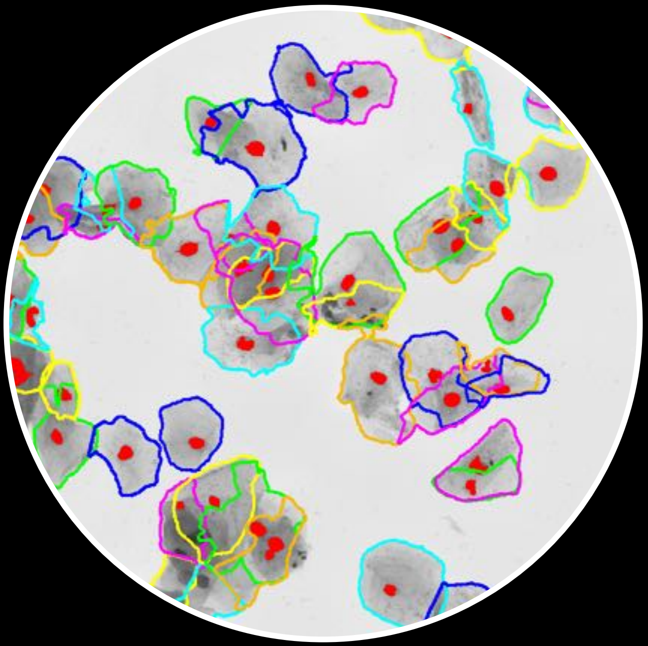
- Problem:
  - Quantification from images;
- Approach:
  - <u>Bio</u>: detect and track biomarkers associated to the progression of diseases;
  - <u>Math</u>: schemes for image representation, segmentation, characterization, classification, interpretation;
- Expected impact:
  - Software to measure experiments systematically;
  - Develop new treatments that target individuals more precisely.

2014- Science Without Borders, Moore Sloan Foundation, BrighFocus

The cancer project

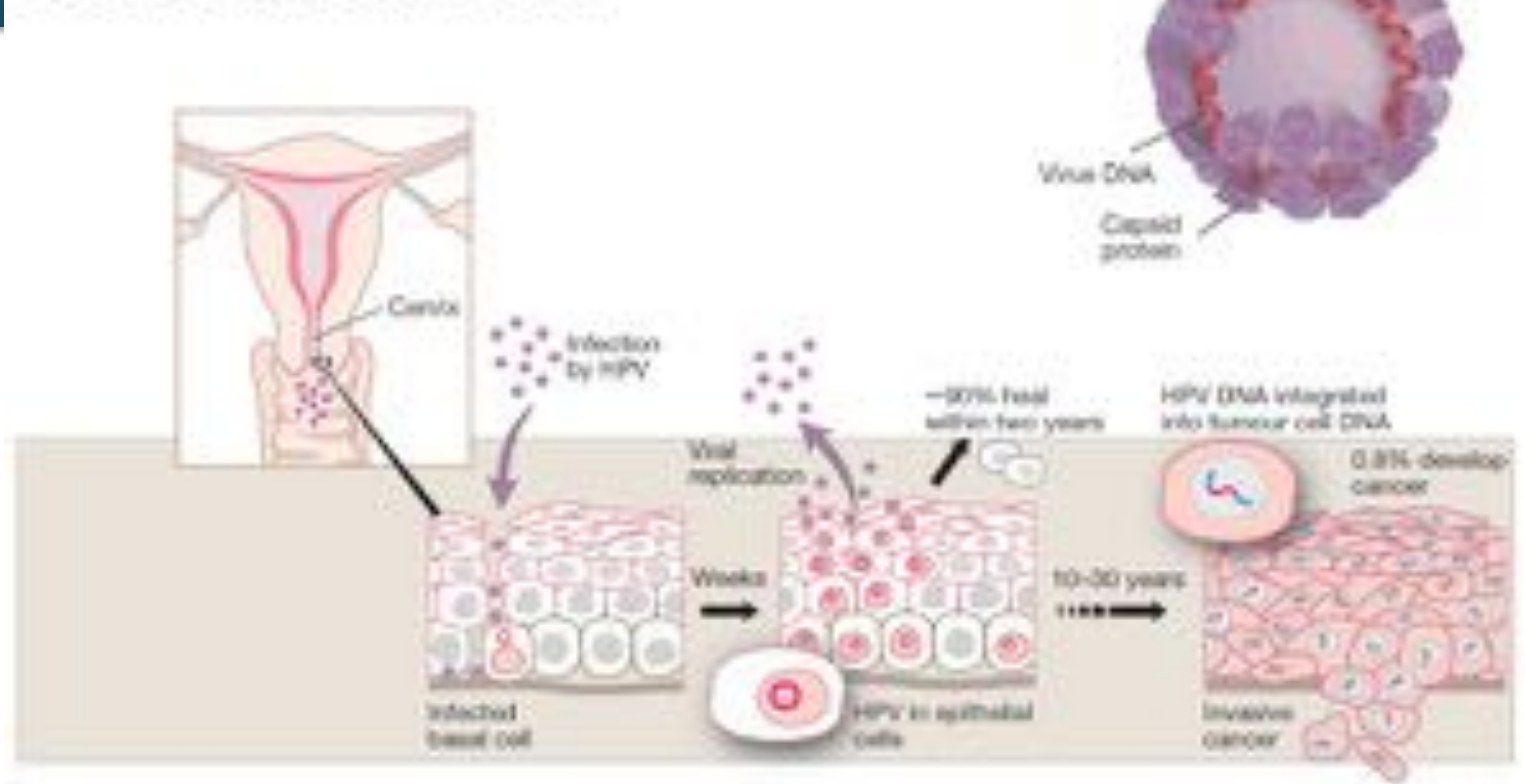
What is it?
How did it start?
Why should you care?

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE



Cervical cells

Cervical Cancer



Worldwide: 2nd leading cause of women death [20-40] years; 570K new cases/year.

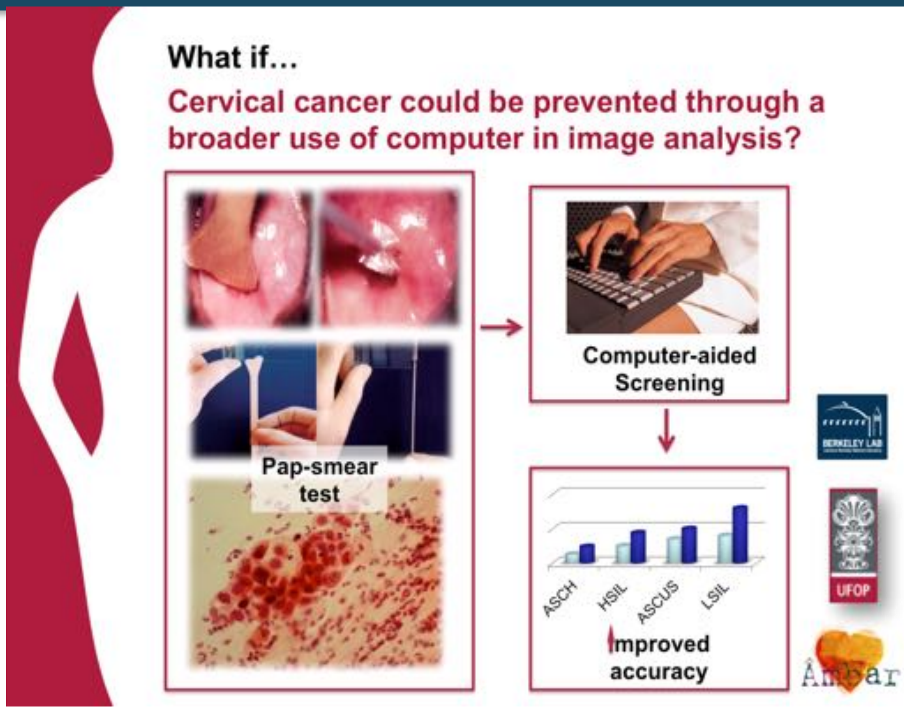
"… type of cancer that **kills more women** in East Africa and parts of Brazil...",

"... the chances to be cured are **100%** when discovered in early stages."

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

Figure credit: http://www.healthyandfitworld.com/how-to-prevent-cervical-cancer/

4. Cell recognition using computer vision



- ***Problem:*** screening of cervical cells based on Pap smears for improved cytology analysis for underserved communities;
- ***Approach:*** we invented a method for cell identification (SPVD);

- ***Results and Impact:***
  - Algorithms detects subcellular structures in 12 seconds;
  - Highest precision in comparison with other available algorithms;

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

IEEE ISBI14
Overlapping Cervical Cytology Image Segmentation Challenge

- **Xmas 2013;**
- **Input:**
  - 512x512 pixels, 2+ cells/img with different degrees of overlap, contrast, and texture;
- **961 images:**
  - GT: train 45, test: 90 realist
  - Unknowns: 16 real + 810 realistic
- **Our algorithm SPVD:**
  - Filters + superpixels + binary classifiers + Voronoi diagram

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

Super Pixel Voronoi Diagram for Giemsa-stained Cells

$$h(x) = k^{-1}(x) \int_{-\infty}^{+\infty} f(\xi) \overbrace{c(\xi, x)}^{\text{geometric}} \; \overbrace{s(f(\xi), f(x))}^{\text{photometric}} \, d\xi$$

E. Wes Bethel, "High Performance, Three-Dimensional Bilateral Filtering", 2009
Tomasi et al, Bilateral Filtering for Gray and Color Images, p. 839, ICCV 1998

Super Pixel Voronoi Diagram for Giemsa-stained Cells

Bilateral → **SRM** → **Phansalkar** → **Fast marching** → **Voronoi**

$$P(R_i, R_j) = \begin{cases} 1 \; if \; |R_i, R_j| \le \sqrt{b^2 R_i + b^2 R_j} \\ 0 \, , otherwise \end{cases}$$

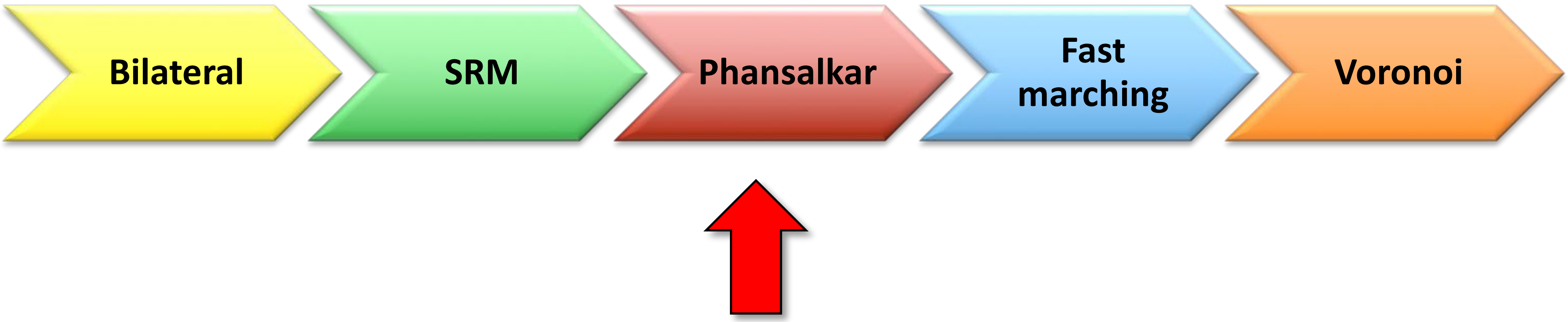
merging predicate

$$b(R) = g\sqrt{\frac{1}{2Q|R|}\left(\frac{\ln S_{|R|}}{\delta}\right)}$$

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

Nock et al, Statistical Region Merging, IEEE Trans. PAMI, 26, p. 1452-1458, 2004.

Super Pixel Voronoi Diagram for Giemsa-stained Cells



$$t = \mu * (1 + p * \exp(-q * \mu) + k * ((\sigma / s) - 1))$$

Phansalskar et al "Adaptive local thresholding for detection of nuclei in diversity stained cytology images", ICCV 2011

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

Super Pixel Voronoi Diagram for Giemsa-stained Cells

$$\boxed{|\nabla T|\,F(x,y)=1}$$

$$[\max(D^{-x}_{i,j}T, -D^{+x}_{i,j}T, 0)^2 + \max(D^{-y}_{i,j}T, -D^{+y}_{i,j}T, 0)^2] = {}^{1}\!/_{F^2_{i,j}}$$

$$D^{-x}_{i,j} = \frac{T_{i,j} - T_{i-1,j}}{\Delta x}$$

J. A. Sethian. Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science, 1999.

Super Pixel Voronoi Diagram for Giemsa-stained Cells

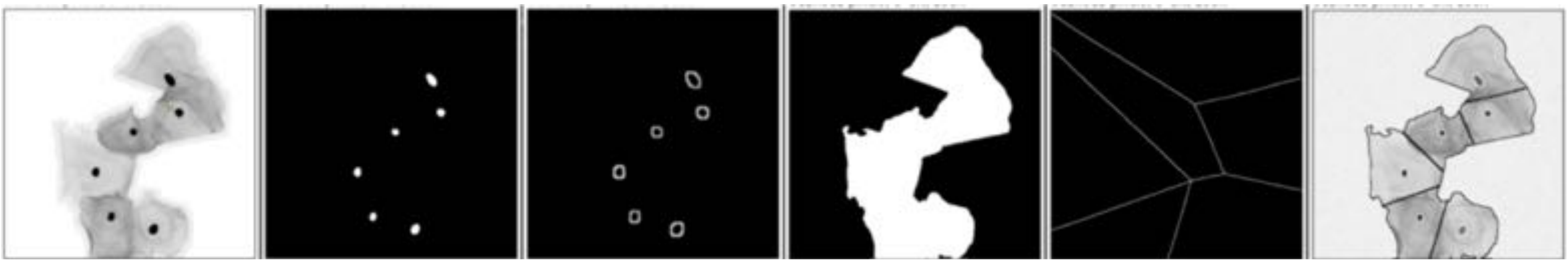
Bilateral → SRM → Phansalkar → Fast marching → **Voronoi**

$$R_k = \{x \in X \mid d(x, P_k) \le d(x, P_j) \text{ for all } j \neq k\}$$

Schindelin 2008

Super Pixel Voronoi Diagram for Giemsa-stained Cells

Bilateral → **SRM** → **Phansalkar** → **Fast marching** → **Voronoi**



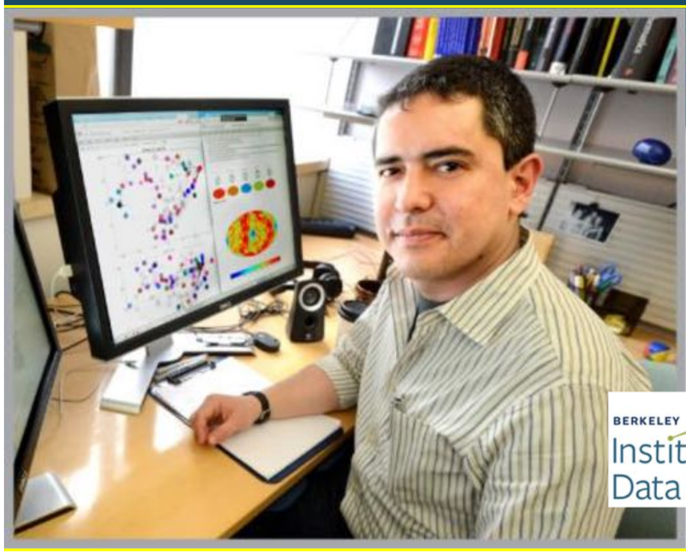
a b c d e f

1st place in code competition

IEEE Int. Symposium on Biomedical Imaging 2014

And now what?

CerviCal



2014-2019

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

Cytopathology

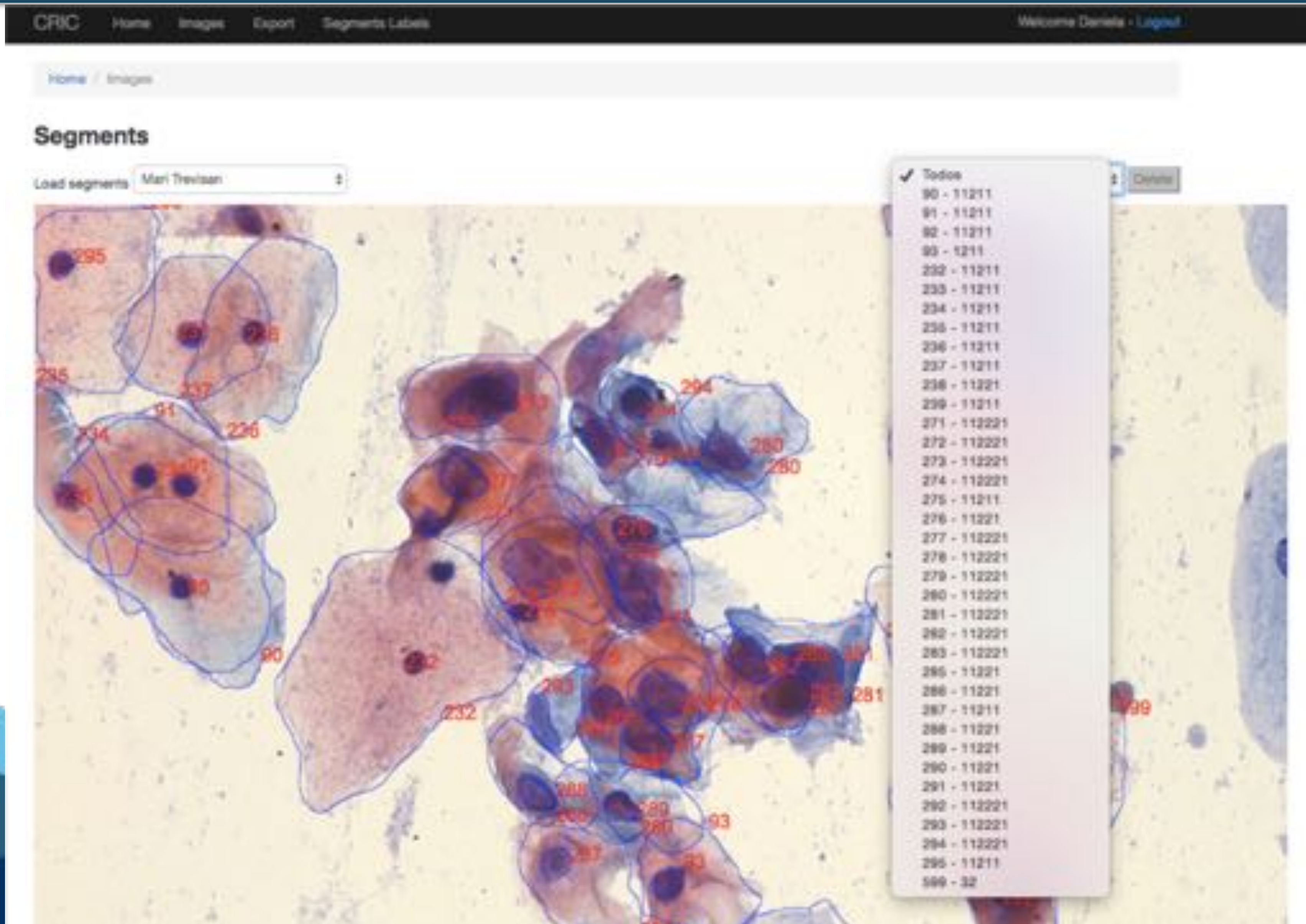
Engineering

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

Computer Science

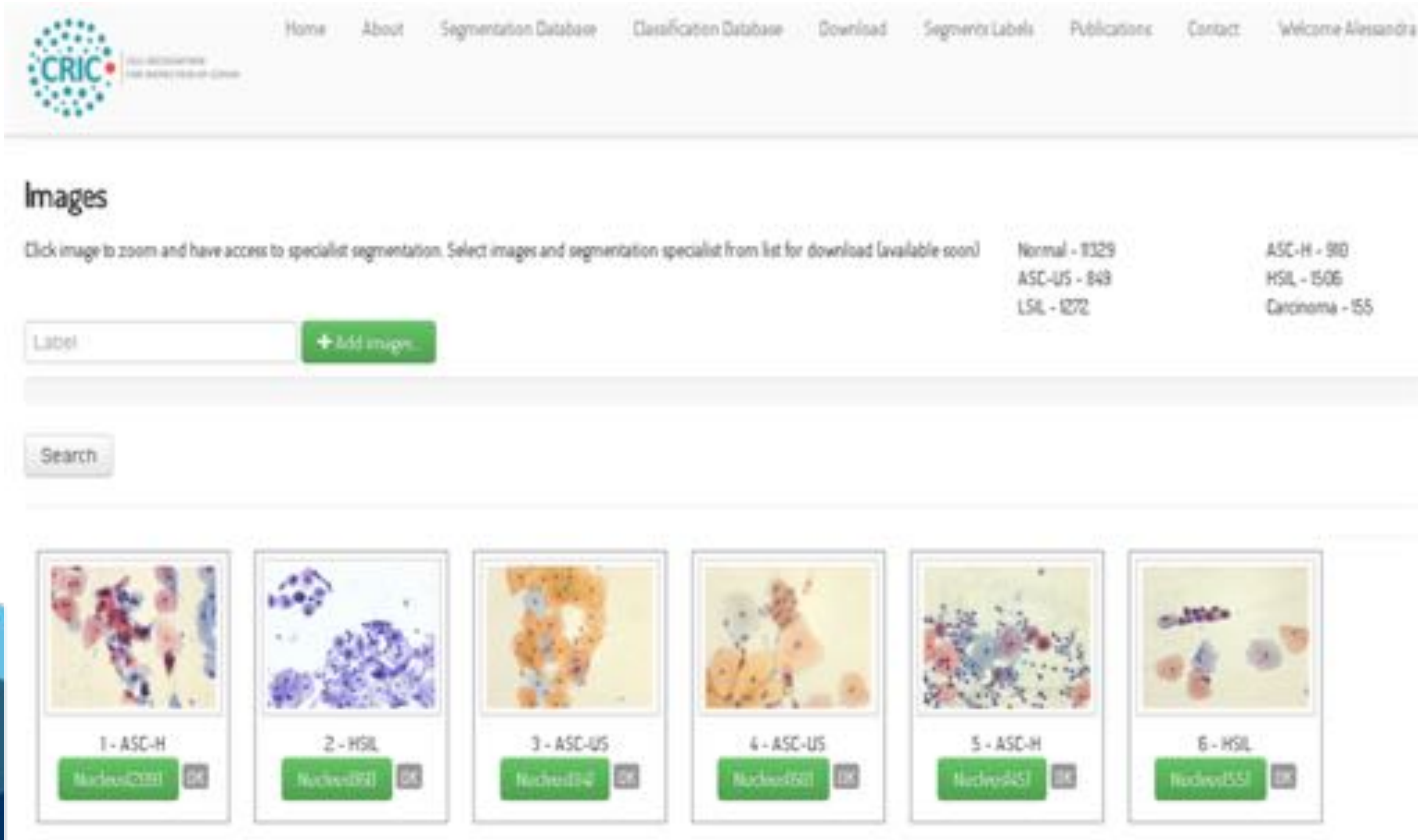
CRIC

CRIC Cervical cells database



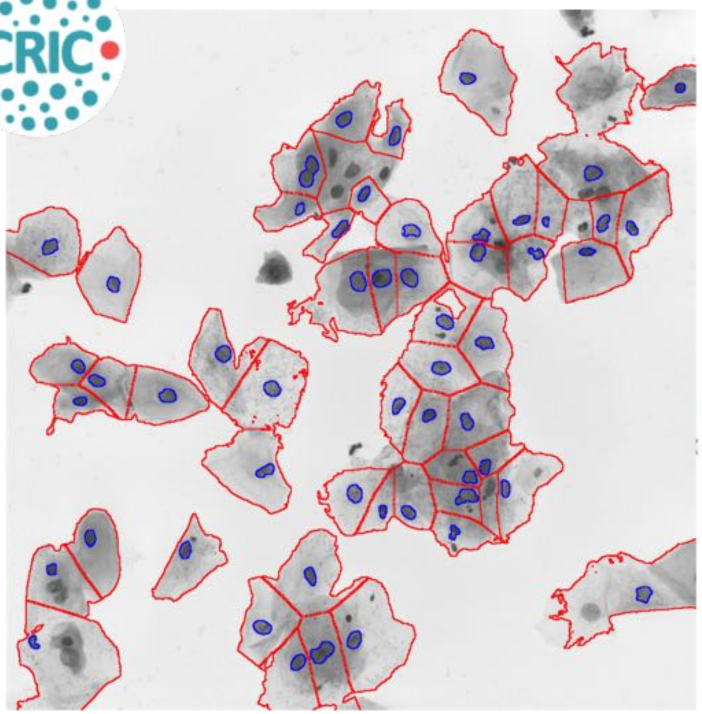
CRIC: Center for Recognition and Inspection of Cells

Dani Ushizima, Ph.D (LBNL/BIDS), Fatima Medeiros (UFC), Claudia Carneiro (Univ. Manchester/UFOP), Andrea Bianchi (UFOP)

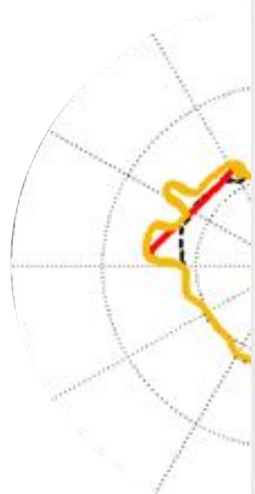


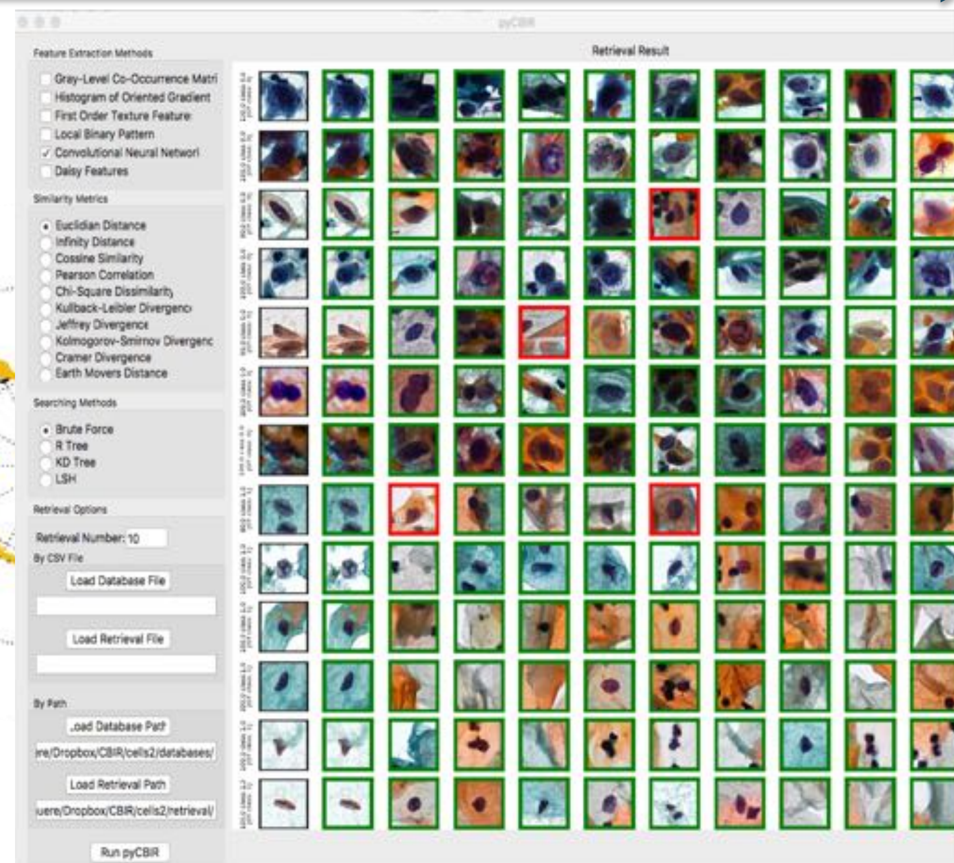
From cell simulations to routine Pap smear data

2012 → 2019



IEEE Int. Conf on Biomedical Imaging
Ushizima's team in 2014

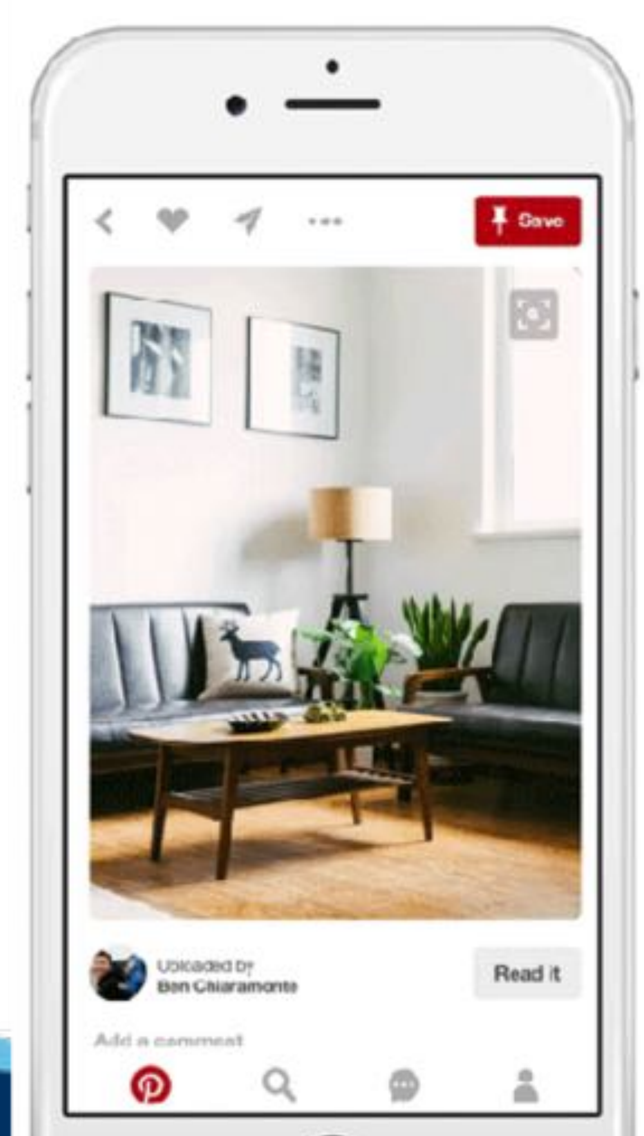




pyCBIR for scientific image search
CRIC team in 2019

https://www.youtube.com/watch?v=G4JOcZo8GyM

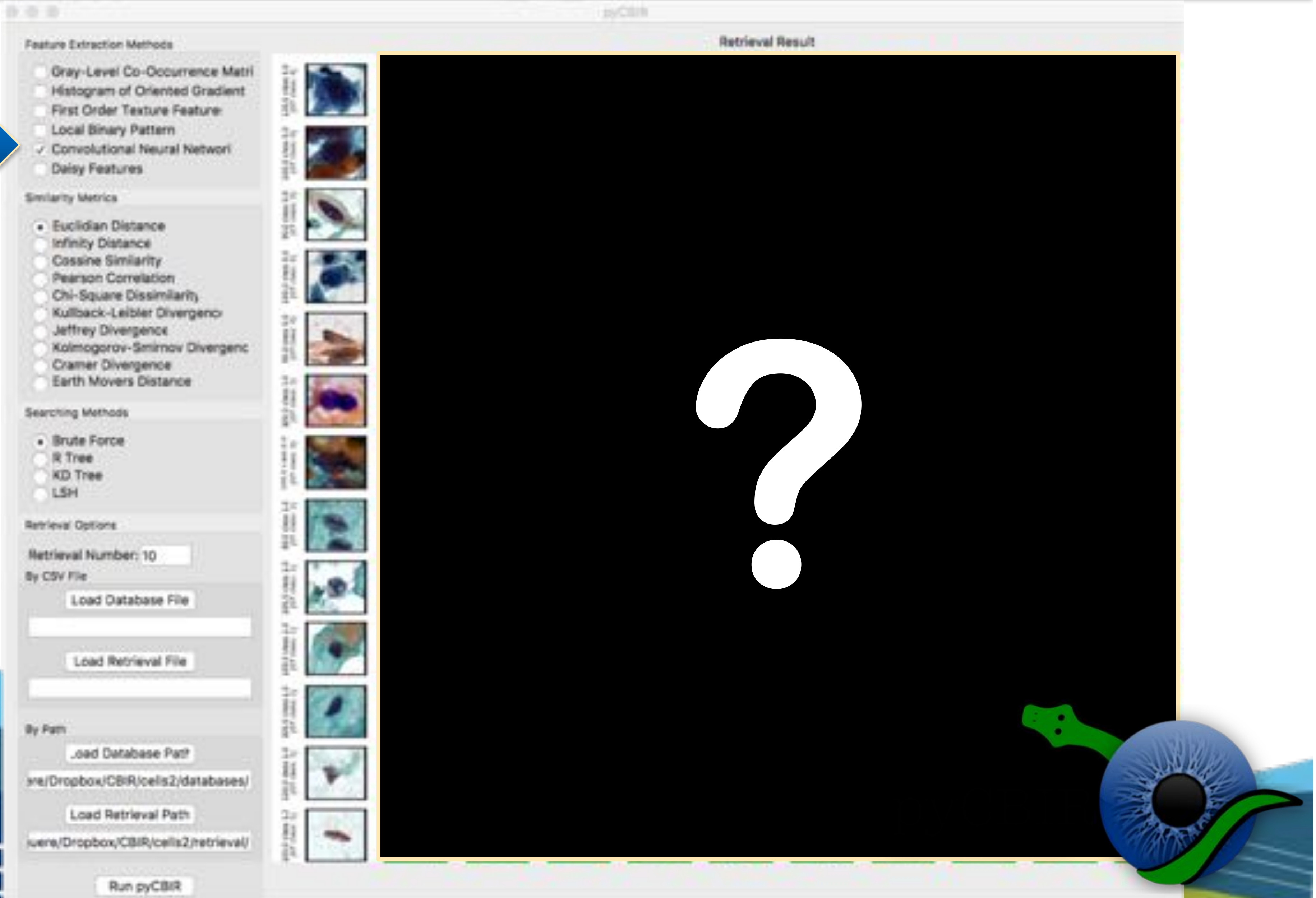
Pinterest Pinboard for science?



BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

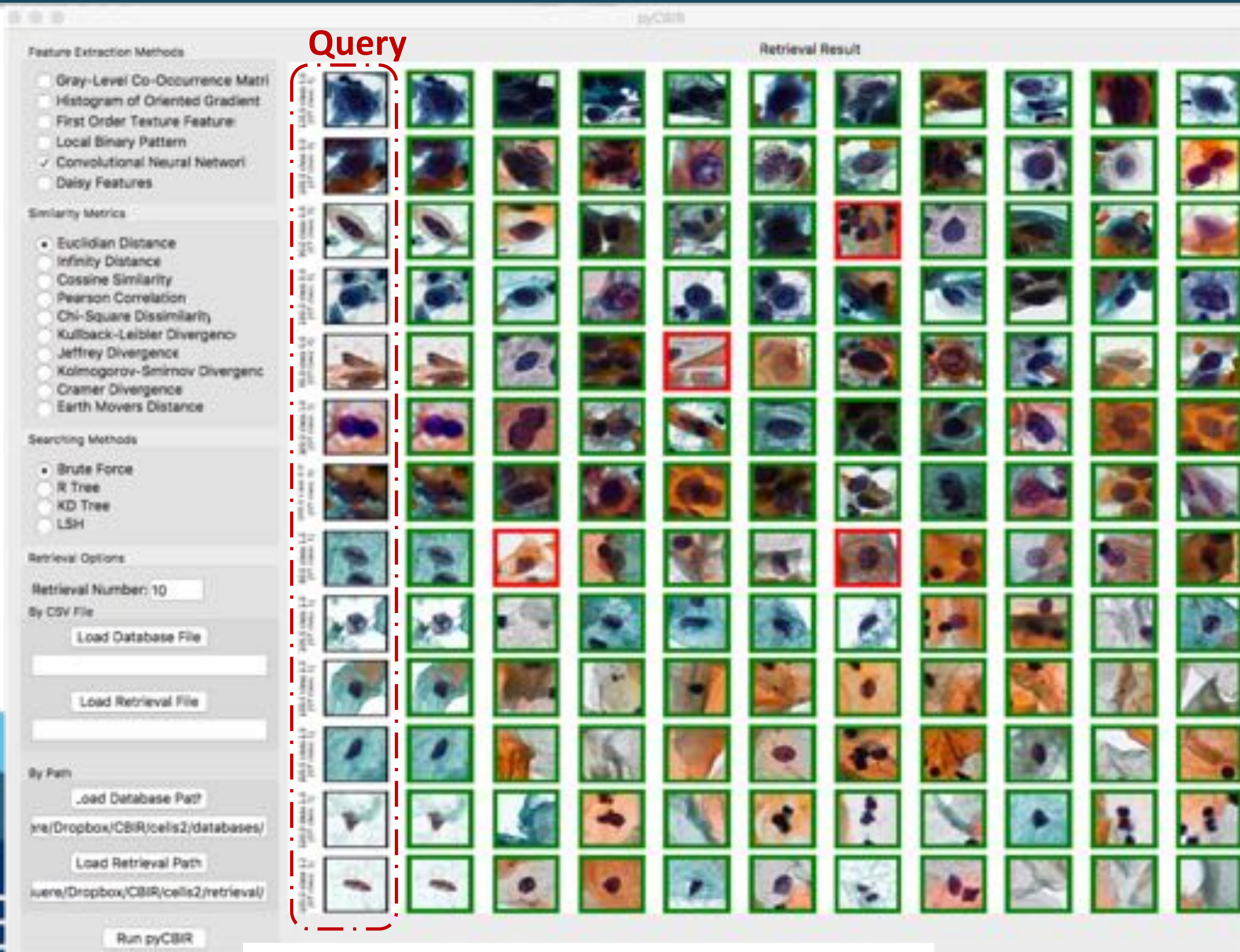
Searching and ranking cells by similarity

Convnet



BERKELEY INSTITUTE FOR DATA SCIENCE

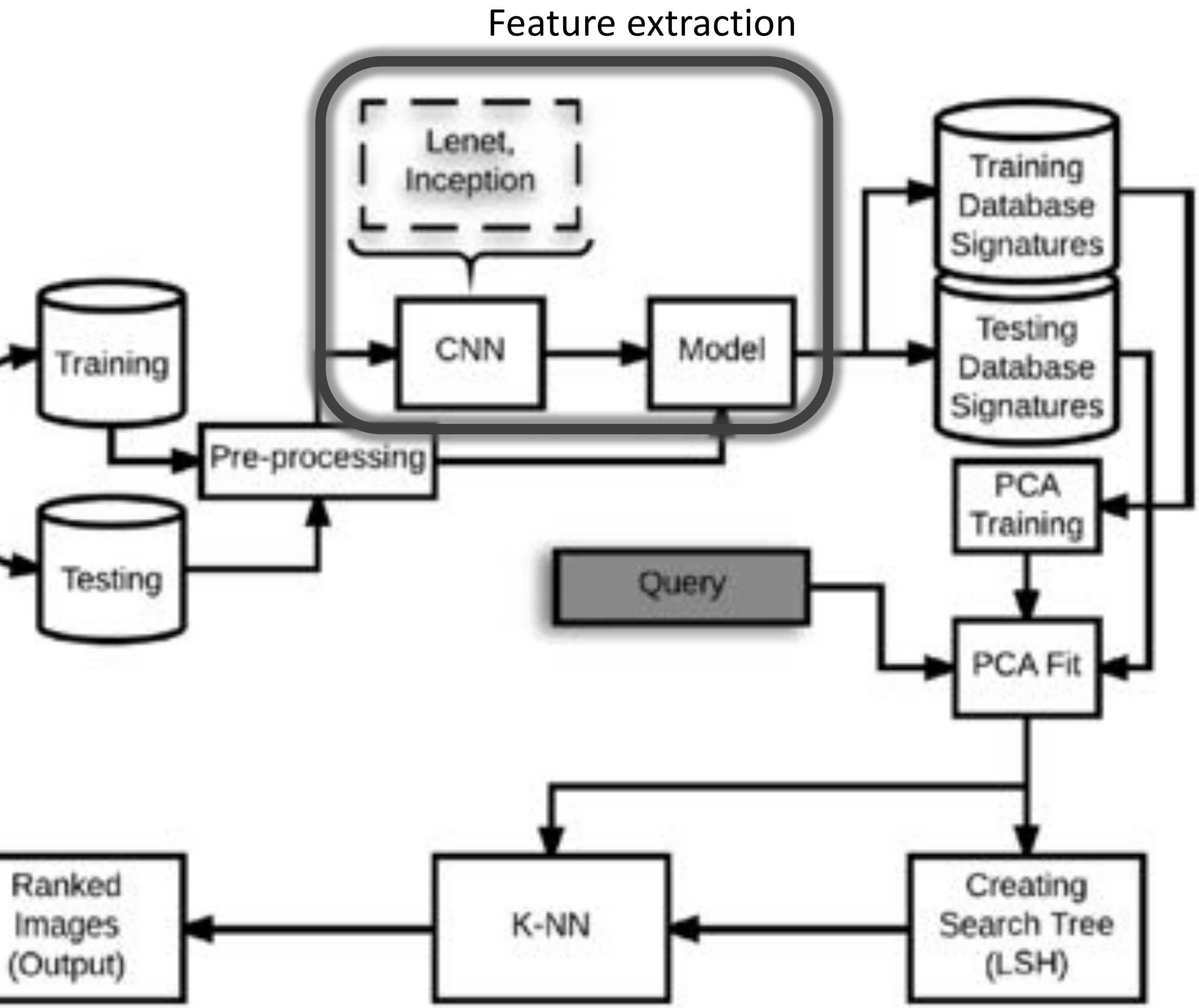
pyCBIR: AI for content-based image retrieval



https://bit.ly/aimagesearch

BERKELEY LAB
LAWRENCE BERKELEY NATIONAL LABORATORY

pyCBIR diagram



Berkeley
Institute for
Data Science

Feature Extraction Methods

- **Signature = index = feature vector = descriptors;**

1.Gray Level Co-Occurrence Matrix;

2.Histogram of Oriented Gradient;

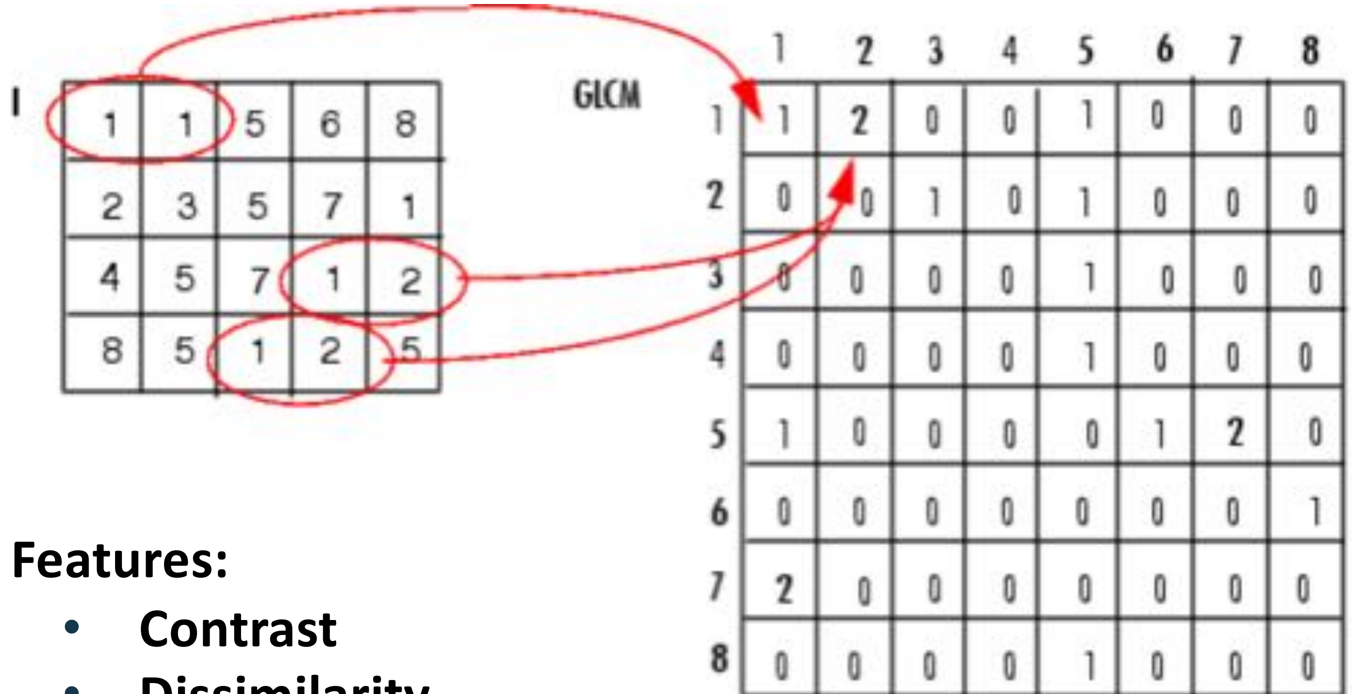
3.First Order Texture Features;

4.Local Binary Pattern;

5.Convolutional Neural Network.

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

Gray Level Cooccurrence Matrix (GLCM)



Features:

- **Contrast**
- **Dissimilarity**
- **Homogeneity**
- **Energy**
- **Correlation**
- **ASM**

First Order Texture Features

Mean

$$\mu = \sum_{i=0}^{G-1} i p(i)$$

Variance

$$\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 p(i)$$

Skewness

$$\mu_3 = \sigma^{-3} \sum_{i=0}^{G-1} (i-\mu)^3 p(i)$$

- **Kurtosis**

$$\mu_4 = \sigma^{-4} \sum_{i=0}^{G-1} (i-\mu)^4 p(i) - 3$$

- **Entropy**

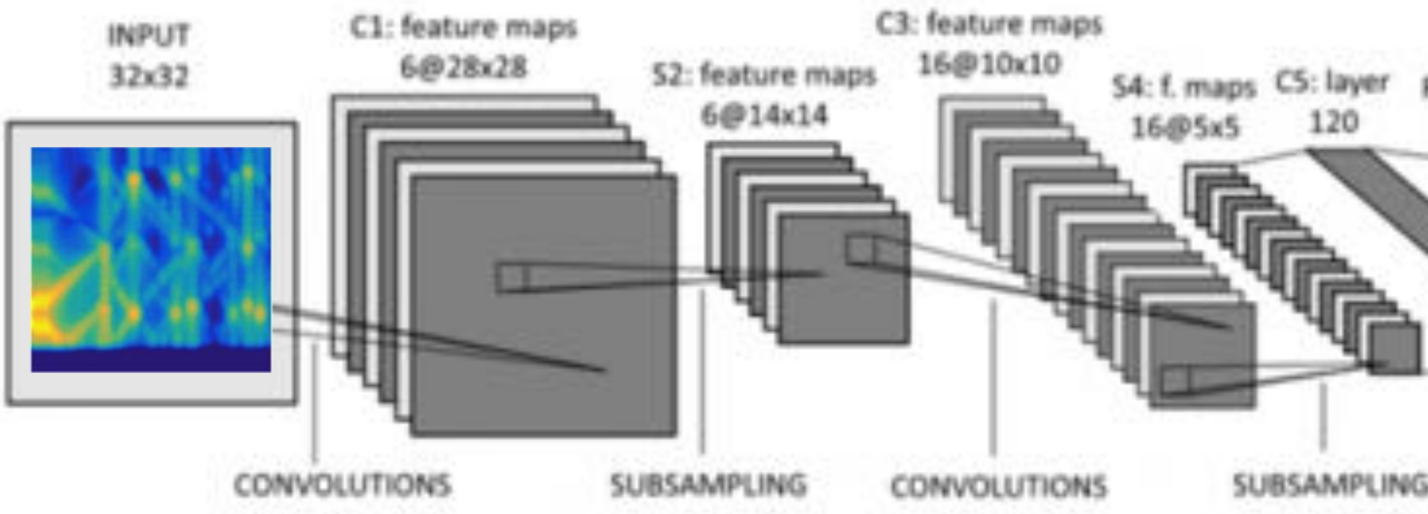
$$H = -\sum_{i=0}^{G-1} p(i) \log_2 [p(i)]$$

- **Energy**

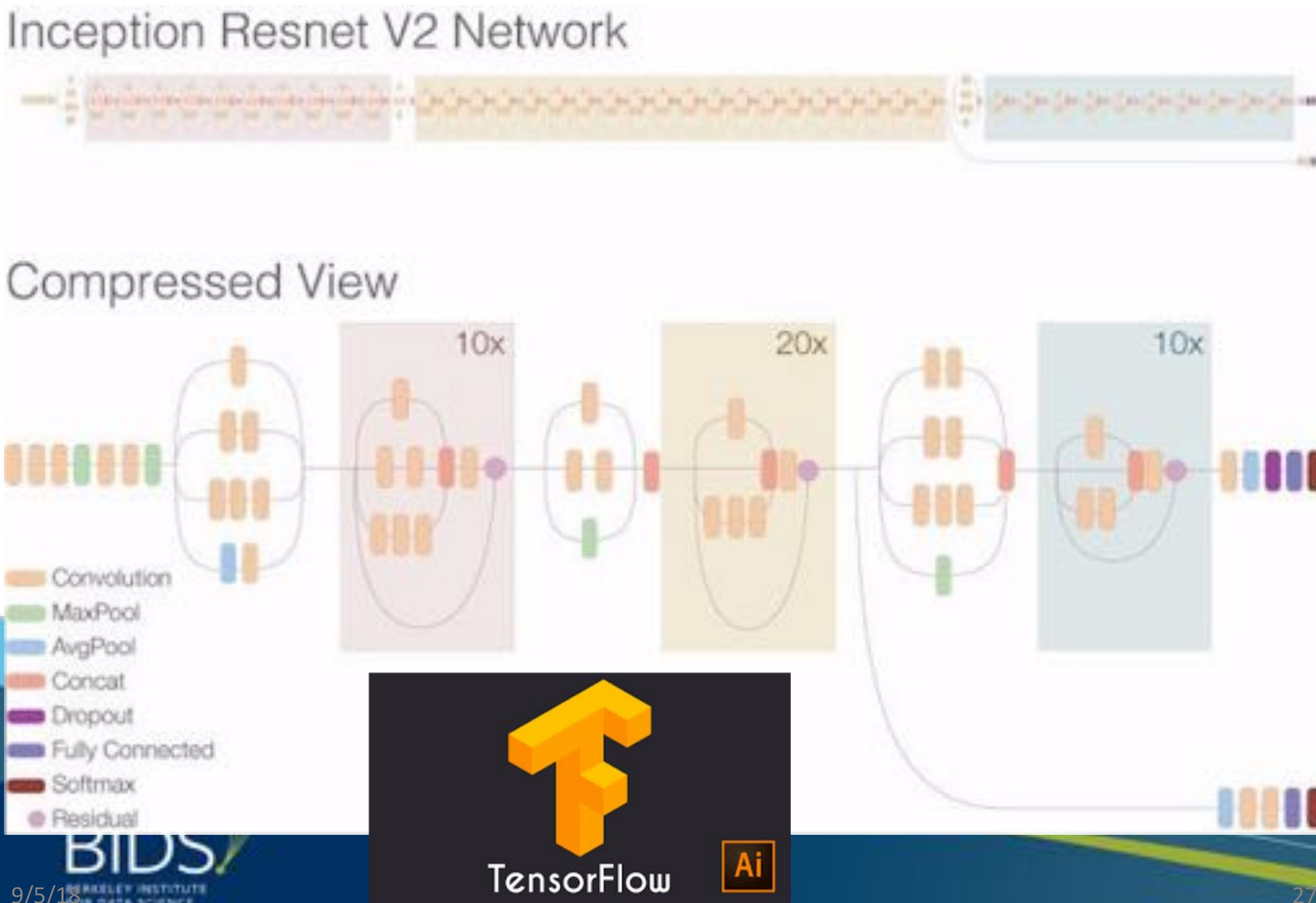
$$E = \sum_{i=0}^{G-1} [p(i)]^2$$

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

LeNet-5


INPUT
32x32
C1: feature maps
6@28x28
S2: feature maps
6@14x14
C3: feature maps
16@10x10
S4: f. maps
16@5x5
C5: layer
120
CONVOLUTIONS
SUBSAMPLING
CONVOLUTIONS
SUBSAMPLING

Deeper CNN: Inception Resnet v2

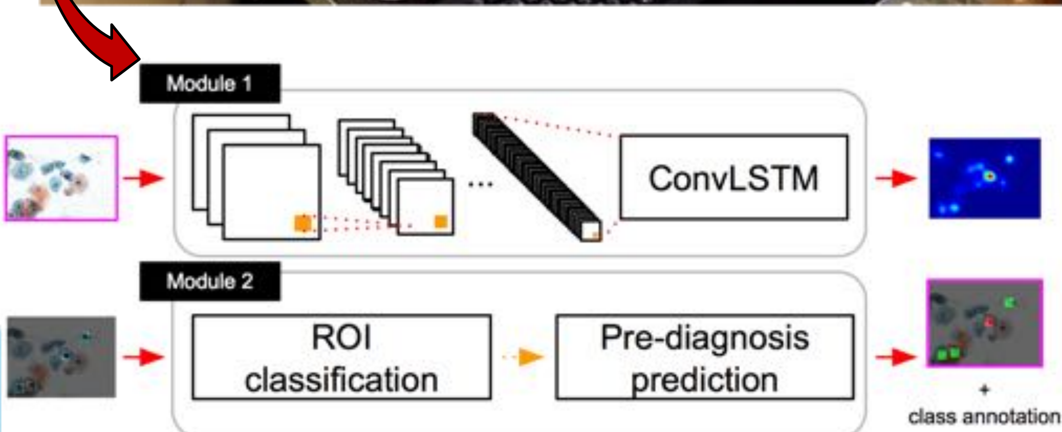


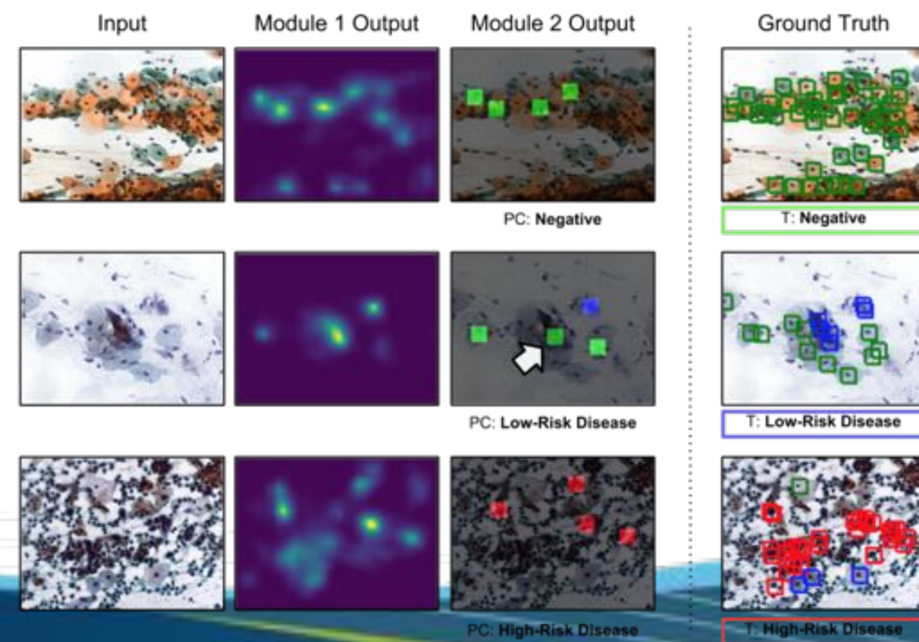
Similarity

- **IF Image = multidimensional vector, THEN similarity = distance!**

1. Euclidean
2. Infinity
3. Cosine
4. Pearson
5. Chi-Square
6. Kullback-Liebler Divergence
7. Jeffrey Divergence
8. Kolmogorov-Smirnov Divergence
9. Cramer-von Mises Divergence
10. Cityblock Distance

Saliency-driven system for cell image analysis

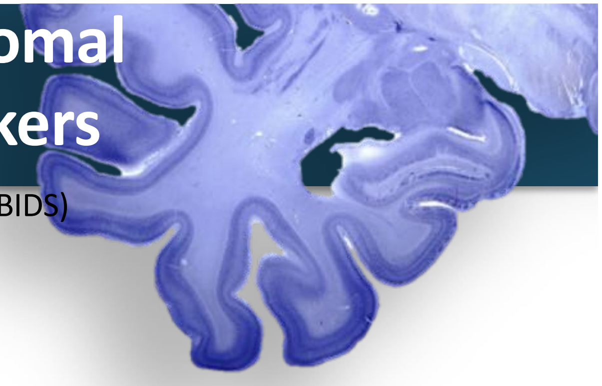




BERKELEY
Institute for Data Science

Collaboration across communities

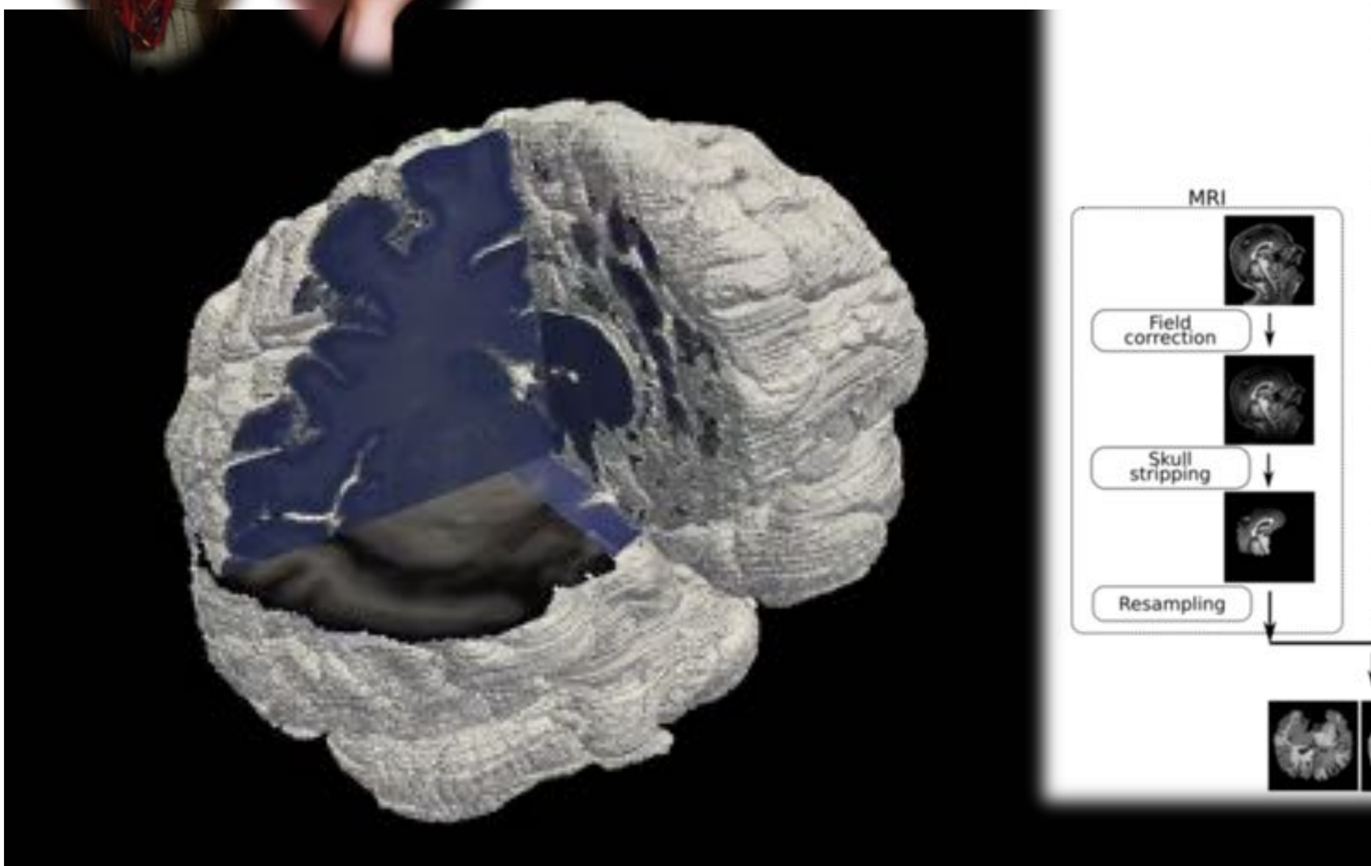
Pathway toward diagnosing and monitoring prodromal Alzheimer's Disease using novel imaging biomarkers

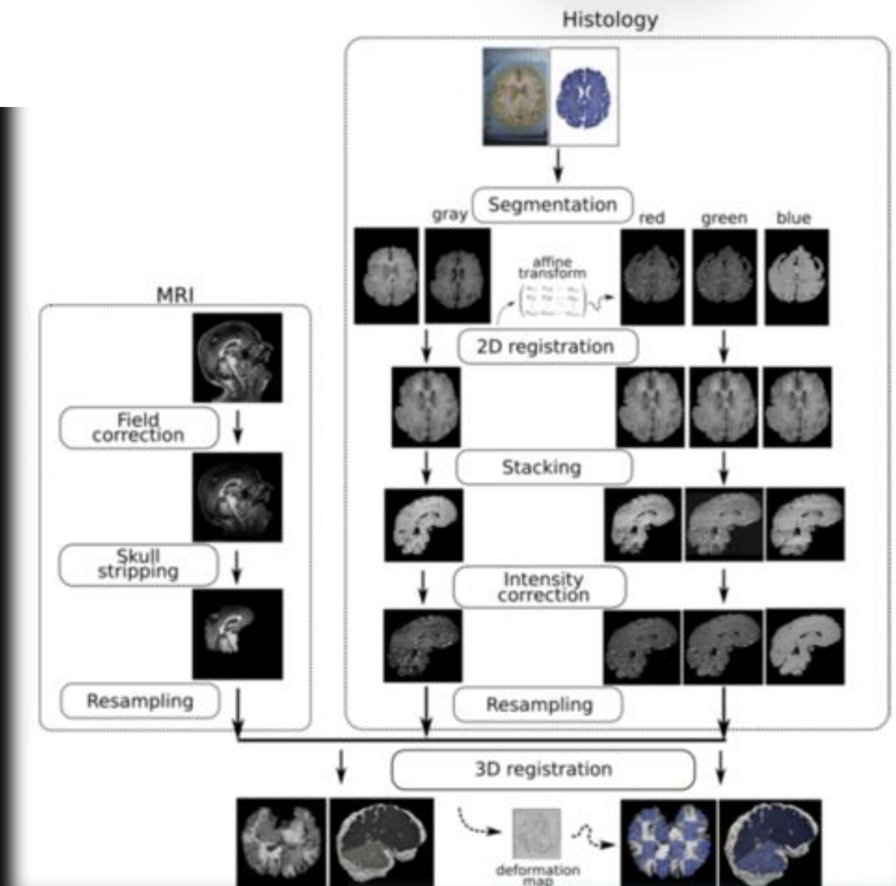


Lea T. Grinberg, M.D, Ph.D (UCSF), Maryana Alegro, Ph.D (UCSF/LBNL) and Dani Ushizima, Ph.D (LBNL/BIDS)

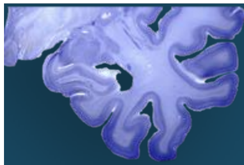
University of California
San Francisco
UCSF
Department of Neurology
Memory and Aging Center

Tissue





Alegro, Amaro, Heisen, Loring, Alho, Zollei, Ushizima, Grinberg, "Multimodal whole brain registrations: MRI and high resolution histology", IEEE CVPR **2016**.



Imaging biomarkers and multimodal data

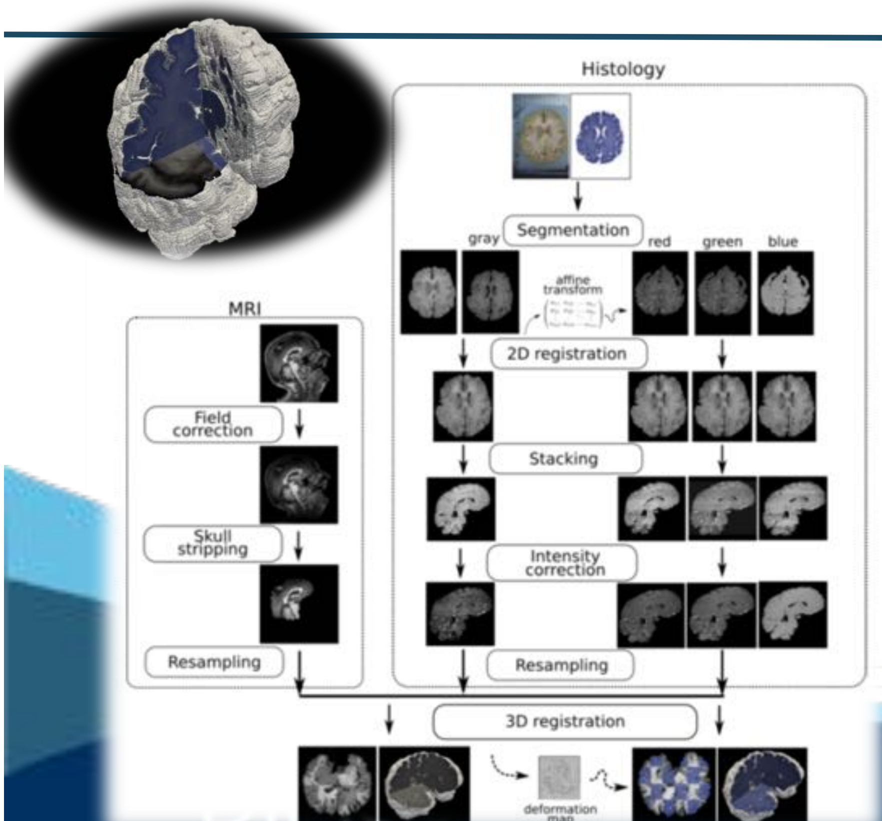
LBNL PI: D. Ushizima (CRD), UCSF PI: L. Grinberg

Scientific Achievement

- Image registration using data from different sources;
- Automated classification of microscopic structures enabled by hybrid data models.

Significance and impact

- Software tools to enhance data through multimodal imaging, e.g., increased clinical MRI resolution with histology;
- Develop biomarkers with potential to detect and track Alzheimer's progression in early stages;



Research details

- Exploit machine learning and computer vision algorithms from ASCR ECRP toward broader applications, such as NIH data from UCSF Memory and Aging Center;
- Testing ability to categorize thousands of samples, considering data from different instruments such as CT, MRI and fluorescent microscopy;
- Collaboration with 2 UC campuses: UCB and UCSF;
- HPC usage: XSEDE Jetstream, NERSC Cori.

. Alegro, Theofilas, Nguy, Castruita, Seeley, Heinsen, Ushizima, Grinberg. "Automating Cell Detection and Classification in Human Brain Fluorescent Microscopy Images Using Dictionary Learning and Sparse Coding", Journal of Neuroscience Methods, 2017.

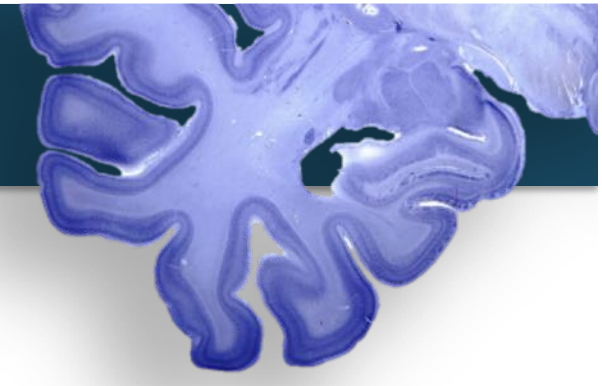
. Alegro, Amaro, Heisen, Loring, Alho, Zollei, Ushizima, Grinberg, "Multimodal whole brain registrations: MRI and high resolution histology", IEEE CVPR 2016.

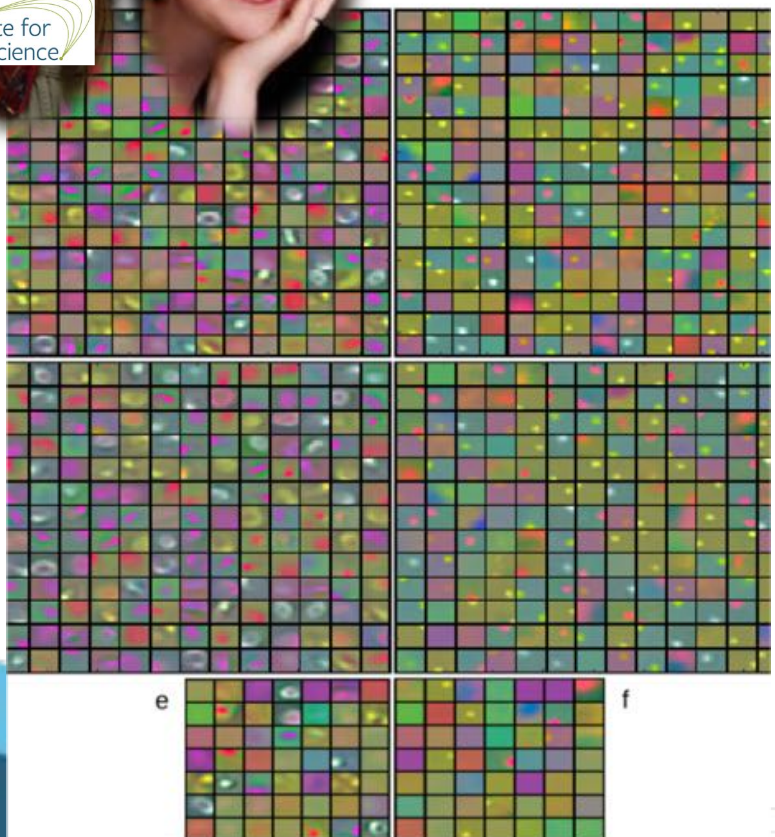
Human brain fluorescent cell microscopy

University of California
San Francisco
UCSF
Department of Neurology
Memory and Aging Center

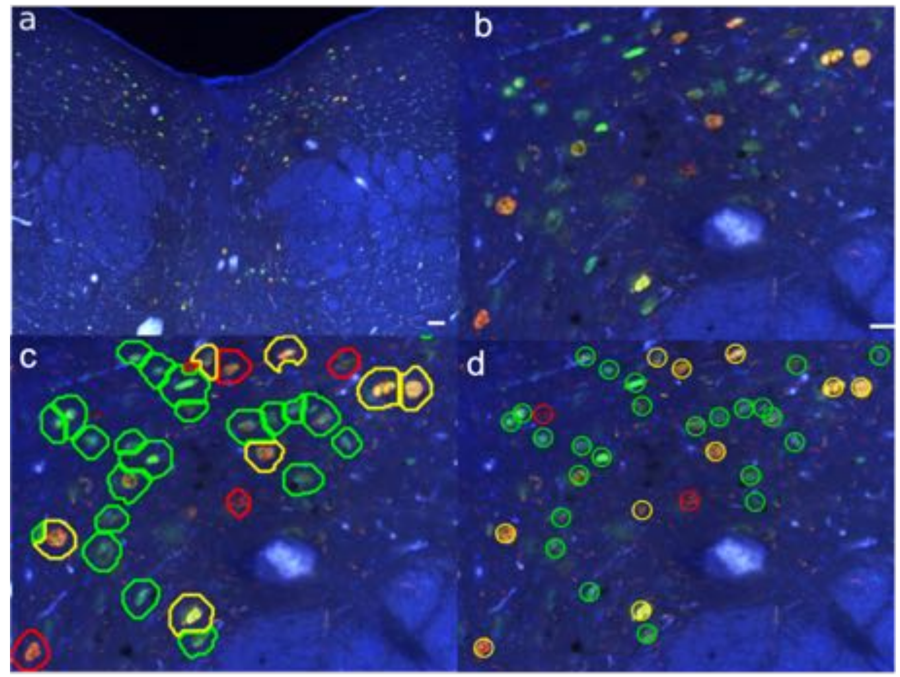
BERKELEY
Institute for Data Science

Neurons





Dictionary learning



Neuron detection

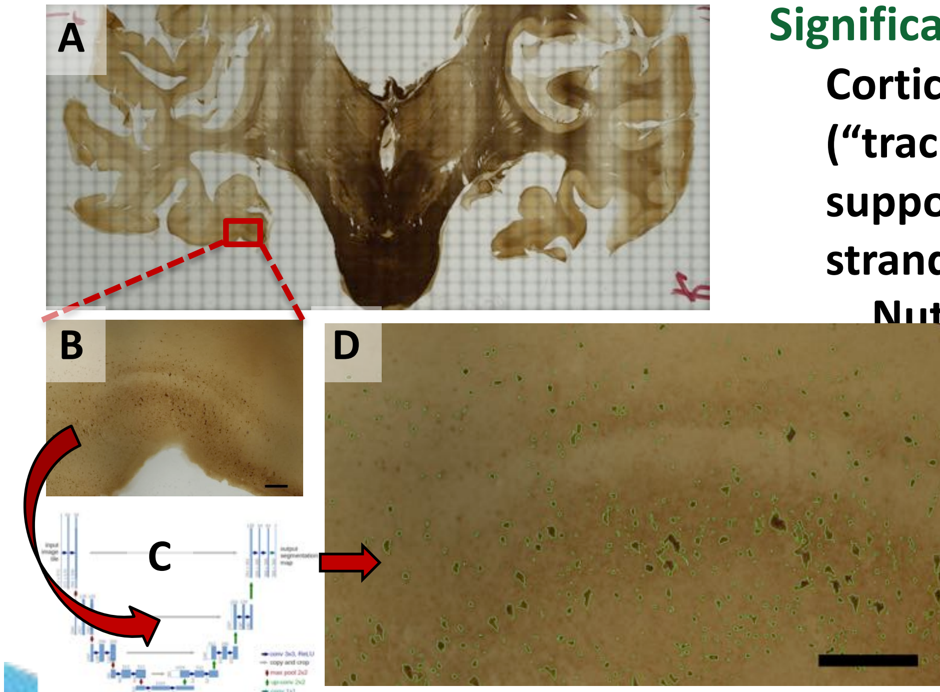
Alegro, Theofilas, Nguy, Castruita, Seeley, Heinsen, Ushizima, Grinberg, “Automating Cell Detection and Classification in Human Brain Fluorescent Microscopy Images Using Dictionary Learning and Sparse Coding”, Journal of Neuroscience Methods, 2017.

Deep Learning for Neurodegenerative Diseases

with UCSF: Lea Grinberg and Maryana Alegro

Scientific Achievement

Machine learning for billion-pixel brain pathology scan detects and segment tangles with concentrates of Tau protein in the human brain, a key indicator of neurodegeneration.



(A) Hippocampal slice at full resolution with scanned images preprocessed and tiled; (B) Magnification of ROI; (C) A Unet variant is used to locate and segment the Tau inclusions. (D) Result of the convolutional neural network segmentation for previous image (scale bar is 300 microns in B and D).

Significance and Impact

Cortical transport is organized in orderly strands ("tracks") through which nutrients travel along. Tau protein supports these tracks, but when it collapses into twisted strands (tangles), they fall apart and disintegrate. Nutrients and other essential supplies can no longer move through the cells, which eventually die.

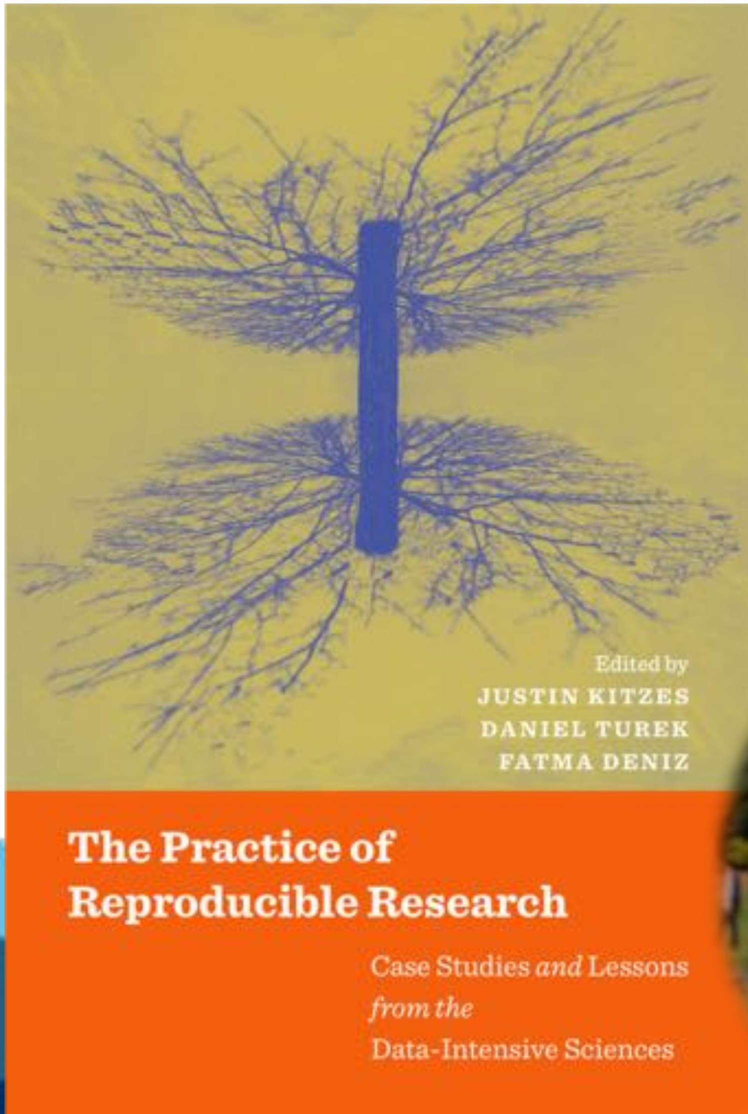
Research Details

- Demonstrated that high-res histopathology imaging can now be automatically processed using computational methods.
- Detected tangles using new convolutional neural network UNet to quantify morphology and spatial distribution of Tau.
- Showed that how to use these regions to create Tau-heatmaps by combining multimodal representations of the brain using MRI, PET (in-vivo) and histology (postmortem).

Alegro, Chen, Castruita, Satrawada, Heinsen, **Ushiziima**, Tsun, Grinberg, "Deep learning for billion-pixel digital pathology analysis: application in mapping Tau protein in the human brain", *Deep Learning in Biomedicine. UCSF, San Francisco, USA* (2018).

BIDS Berkeley Institute for Data Science
University of California San Francisco UCSF Department of Neurology Memory and Aging Center
U.S. Department of Veterans Affairs
BrightFocus™ Foundation Cure in Mind. Cure in Sight.

BIDS book from dozens of fellows



Portable Learning Environments for Hands-On Computational Instruction

Using Container- and Cloud-Based Technology to Teach Data Science

Chris Holdgraf
Berkeley Institute for Data Science,
University Of California, Berkeley
190 Doe Library
Berkeley, CA 94720
choldgraf@berkeley.edu

Aaron Culich
Research IT, University Of California,
Berkeley
250-26 Warren Hall
Berkeley, CA 94720
aculich@berkeley.edu

Ariel Rokem
University of Washington eScience
Institute
The WRF Data Science Studio
Seattle, WA 98105
arokem@uw.edu

Fatma Deniz
Berkeley Institute fo
University Of Ca
190 D
Berke
fatma

Maryana Alegro
University of California, San
Francisco
Mission Bay
an Francisco, CA 94158
ryana.alegr

Dani Ushizima
Lawrence Berkeley National Lab and
Berkeley Institute for Data Science,
University Of California, Berkeley
dushizima@lbl.gov

2016

ImageXD

Image processing across domains

2019

BERKELEY Institute for Data Science

https://imagexd.org

Education

BLACK GIRLS CODE

BIDS has supported the BGC **instructors training** by allowing the use of its space for two years in preparation to Robot Expo. Also, BIDS fellows and staff volunteered in benefit of hundreds of girls in the Bay Area.

Stacey Dorton, Beth Reid, Dani Ushizima

Education

software carpentry

BIDS has supported the Software Carpentry workshops and instructors training. BIDS fellows and staff have volunteered in benefit of hundreds of aspiring Data Scientists in the Bay Area.

Dani, Scott, Jane, Ann, Dima, Nelle, ??

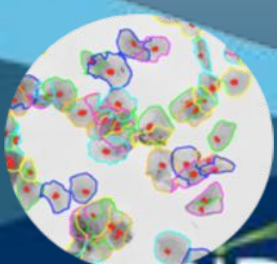
BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

TechWomen - U.S. Dept of State Bureau of Education

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

CRIC: Center for Recognition and Inspection of Cells

http://bit.ly/centercric



GORDON AND BETTY MOORE FOUNDATION

ALFRED P. SLOAN FOUNDATION 1934

Science Without Borders

Microsoft

CRIC

Dr. Claudia Carneiro (Pathology), me, Prof. Andrea Bianchi (CS/Physics), Prof. Fatima Medeiros (Engineering)