



Building Blocks for Al

Marco Antonio Gutierrez

Informatics Division & Biomedical Informatics Laboratory Heart Institute, University of Sao Paulo Medical School São Paulo, BRAZIL





Heart Institute (InCor) University of Sao Paulo Medical School, Brazil

- 500 beds
- 260,000 outpatients/yr
- 12,000 inpatients/yr
- 4,500 surgeries/yr

- 2,000,000 lab tests/yr
- 200,000 Image diagnosis/yr
- 12,000 CATH procedures/yr
- Graphic methods 140,000/yr





Clinics Hospital



University of Sao Paulo Medical School, Brazil

 Largest hospital in Latin America

7 Institutes

- 3.500 outpatients/day
- 1.000.000 lab tests/month
- 3.000 beds AMB InRad

A vew eds





Today: Many disparate data types, streams...



Future: Integrated computing and data







Today Data Sources

- Legacy Systems
 - Print to Text or
 Delimited String
- All EMR Initiated Data (Stored Procedures)
- Medical Imaging Devices (DICOM/HL7

- Physiological Monitors (HL7)
- Ventilators (HL7)
- Lab/Analysis Equipment (HL7)
- Smart Pumps (HL7)
- Home Monitoring
- Hospital Sensors







Newer Data Sources

- External Streaming Device Data
- Wearables
- Home Devices
- Geographic Information System (GIS) Data
- Omic Data

- Open Data
 - http://dados.gov.br
 - http://data.gov.uk
 - http://www.data.gov
- Adverse Drug Event
 - http://www.researchae.com
- Internet of Things (IoT)
 - Telematics
 - 5G

Sa, J.H. ; Rebelo, M. S. ; Brentani, A.; Grisi, S. J.F.E. ; IWAYA, L. H. ; Simplicio, M. A. ; Carvalho, T.; Gutierrez, M.A. Intl. Journal of Medical Informatics, v. 94, p. 91-99, 2016.





EMR: Future data sources

- Patient-Generated Data
 - Home Devices (Scale, Vital Signs, Glucose)
 - Exercise & Diet (Fit Bit, Jawbone, Nike)
- Combining Phenotype Data with Genotype Data
- Patient Threat Analysis

- Edge and Vertices Analysis
 - Patient caregivers and outcomes

Heimar, M. ; Massad, E. ; Gutierrez, M. A. ; Rodrigues, R. J. ; Sigulem, D. Global Health Informatics: How Information Technology Can Change Our Lives in a Globalized World. 1st. Ed. London, Elsevier, 2017.



Al @ Heart Institute (InCor)



BigData & InCor ~0.5 PetaByte (0.5 10¹⁵) and counting 0.5 *1.000.000.000.000



- ✓ 1.4 million patients
- ✓ 40 million LA exams
- ✓ 2 million Image exams
- ✓ > 1 million EKG

Modality		# Exams
Chest X-Ray	CR	840.159
ECO/Ultrasound	US	528.916
EKG	ECG	317.614
Tomography	СТ	235.355
X-Ray Angiograph	XA	174.470
Nuclear Medicine	NM	101.403
Magnetic Resonance	MR	73.107
Endoscopy	ES	14.696
PET	PT	3.674
	Total	2.289.394





- I. Biomedical signal processing (e.g. EKG analysis);
- II. Medical Image processing (e.g. Cardiac CineMRI & Chest X-Ray Analysis);
- III. NLP to extract non-structured information from ePR (e.g. medical report or Corona Virus outbreak)
- IV. Integration of biomarkers & clinical data to produce predictive risk algorithms to assist the management of CVD.





EKG Analysis





Initial Dataset: 20K EKGs Extended Dataset: 200K EKGs

TPR	TNR	FPR	FNR	PPV	F1 Score	Accuracy	AUC Score
0.94	0.71	0.29	0.06	0.90	0.92	0.88	0.82



CineMRI Automated Analysis







	DICE index	APD (mm)	Good Contours (%)
Epicardium	$0,93 \pm 0,06$	$0,72 \pm 0,62$	$98,59 \pm 4,28$
Endocardium	$0,86 \pm 0,13$	$1,19 \pm 1,29$	$94,98 \pm 14,04$



Chest X-Ray Automated Classification







Any Al initiative needs GPU power





<u>**5 h**</u> (NV 1050ti) -> <u>**5 min**</u> (M100)

<u>**1** h</u> (NV 1050ti) -> <u>**1** min</u> (M100)



InCor Data Science Enterprise Resource













Big Data in Life Sciences	 Sequencer advances – 4x data in 50% less time .5TB/device/day 4D molecular imaging produces 2TB/device/day Fragmented software ecosystem, lots of open source 	
Burdens of Data Management	 Store, manage, share, ingest and move PBs of research & clinical data Need to reliably 'snapshot' pipelines with archive to tiered storage 	10 ¹⁸
Innovation Drives Change	 Rapid iteration of algorithms far outpace IT, requiring flexibility, agility Most applications do not fully leverage available infrastructure 	Enabling extreme- scale computing on massive data sets
Converged Infrastructure	 Workloads converging between local and cloud-based HPC/Big Data Advanced orchestration required to maximize throughput & efficiency 	







- Managing peta/exabytes
- How to organize it
- How to reorganize it
- How to share it?

- Tools for visualization & presentation
- Building and executing models
- Documenting experiments
- Curation and long-term preservation

Adapted from: The Fourth Paradigm Data-intensive Scientific Discovery Edited by Tony Hey, Stewart Tansley, and Krist in Tolle, Microsoft Corp, 2009 "Not everything that counts can be counted, not everything that can be counted, counts." Albert Eisntein

VIOLET KASHI PHOTOGRAUM





Thank you!

Email: marco.gutierrez@incor.usp.br