

ML & Advanced Analytics For Biomedicine



CCTS 40500 / CCTS 20500 / BIOS 29208
Winter 2019



Today

- Probability Theory
- Bayesian Statistics
- Naive Bayes Classifier
- Bias vs Variance



So What Exactly Is Machine Learning?



- Learning from machines?
- Learning with help of computers?
- Modeling data?
- Regression?



- Learning from machines?
- Learning with help of computers?
- Modeling data?
- Regression?

data – intelligent automated algorithmic analysis – actionable insights



How Is ML Different From ...

- Artificial Intelligence
- Statistics
- Data Mining
- Deep Learning



How Is ML Different From ...

- Artificial Intelligence
- Statistics
- Data Mining
- Deep Learning

- “*Machine learning is essentially a form of applied statistics*”
- “*Machine learning is glorified statistics*”
- “*Machine learning is statistics scaled up to big data*”
- “*The short answer is that there is no difference*”
- “*Machine learning is for Computer Science majors who couldn't pass a Statistics course.*”
- “*Machine learning is Statistics minus any checking of models and assumptions.*”
- “*I don't know what Machine Learning will look like in ten years, but whatever it is I'm sure Statisticians will be whining that they did it earlier and better.*”

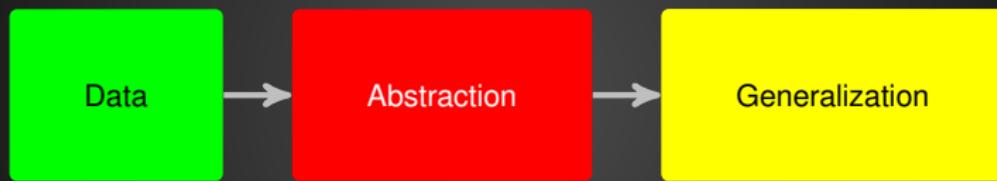


How Is ML Different From ...

- Data Science
- Big Data Analytics



How Do We Teach Machines?





ML Categories

Machine Learning

Supervised

Unsupervised



ML Categories

Machine Learning

Classification

Prediction

Modeling



ML Categories

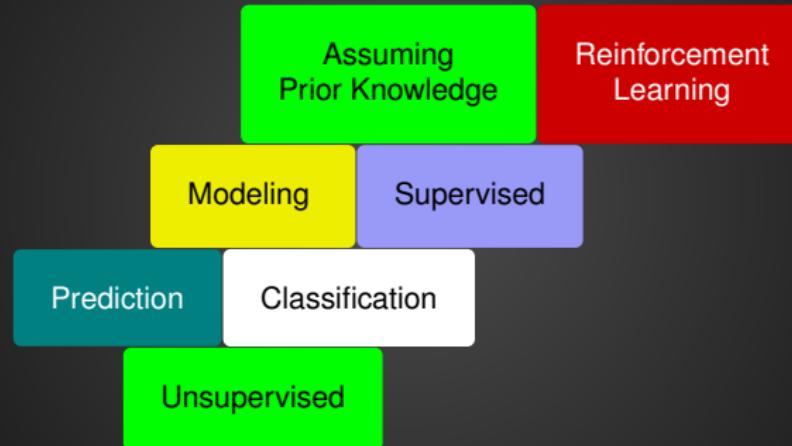
Machine Learning

Assuming
Prior Knowledge

Zero Knowledge
Learning



ML Categories





- Introduction to ML

<http://www.r2d3.us/visual-intro-to-machine-learning-part-1/>

- Bias Variance Tradeoff

<http://www.r2d3.us/visual-intro-to-machine-learning-part-2/>



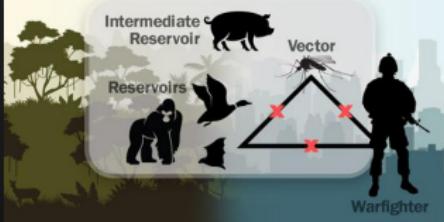
ML: Breaking The Status Quo

Taking On The Grand Challenges

Funded Projects



Zero Knowledge Discovery



Data Driven Discovery of Models (D3M)

Defense Advanced Research Proj. Agency (DARPA I2O)

- 2017-2021

Preventing Emerging Pathogenic Threats (PREEMPT)

Defense Advanced Research Proj. Agency (DARPA DSO)

- 2018-2021

Physics Augmented AI (PAI)

Defense Advanced Research Proj. Agency (DARPA DSO)

- 2018-2020

Crimes of Prediction

Neubauer Collegium for Culture & Society

- 2017-2019



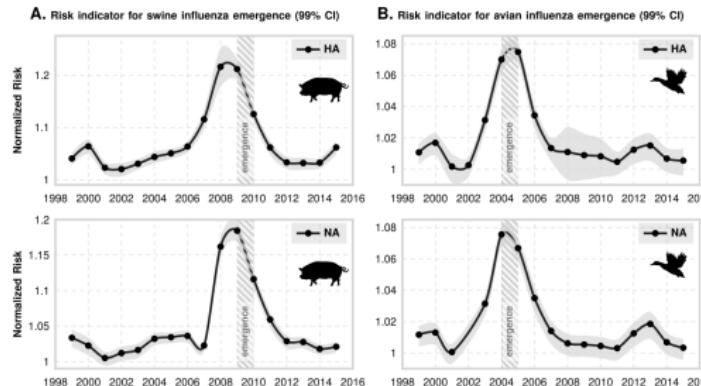


ML: Breaking The Status Quo

Taking On The Grand Challenges

Biosurveillance For Predicting Zoonotic Emergence

J.Dhanoa, B. Manicassamy and I. Chattopadhyay , "Algorithmic Bio-surveillance For Precise Spatio-temporal Prediction of Zoonotic Emergence", <https://arxiv.org/abs/1801.07807>



Predicting
the
next pandemic:
**Where
When
Who**

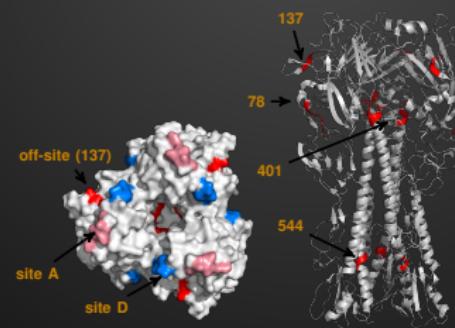
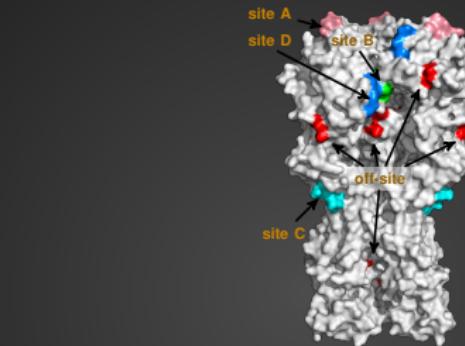


THE UNIVERSITY OF
CHICAGO BIOLOGICAL SCIENCES



ML: Breaking The Status Quo

Taking On The Grand Challenges

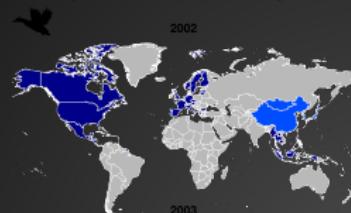




ML: Breaking The Status Quo

Taking On The Grand Challenges

A. Geospatial Risk of Avian Influenza Emergence



2003



2004



2005



B. Geospatial Risk of Swine Influenza Emergence



2007



2008



2009



Min.

Normalized Emergence Risk

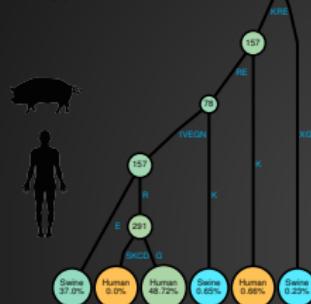
Max.



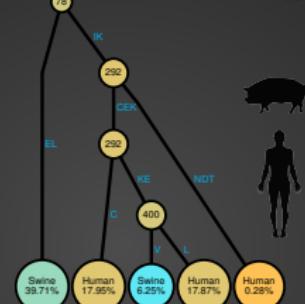
ML: Breaking The Status Quo

Taking On The Grand Challenges

A. 2007



B. 2009



E. Algorithm Steps

I. Query IRD
for species-specific
protein sequences

II. Generate
Decision Trees using
Iterated Feature Deletion

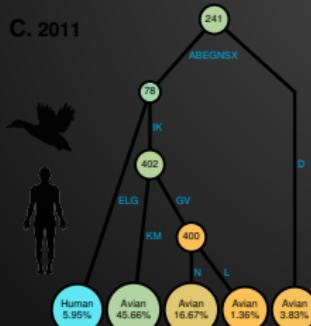
III. Identify
Minimal
Feature Set

IV. Train
Random Forest Classifier
with
Minimal Feature Set

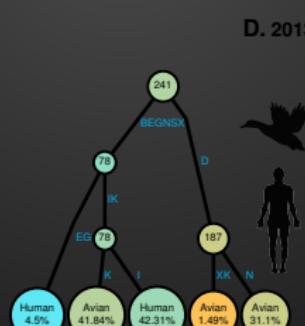
Log
In-sample Accuracy
For Each Year

Generate
Risk Indicator Curve

C. 2011



D. 2013





ML: Breaking The Status Quo

Taking On The Grand Challenges

Predicting Neuropsychiatric Diagnoses From Sparse Electronic Health Records

Truven Health Analytics MarketScan® Commercial Claims and Encounters
Database for the years 2003 to 2012

- 4.6 billion inpatient and outpatient service claims
 - 150 Million individuals



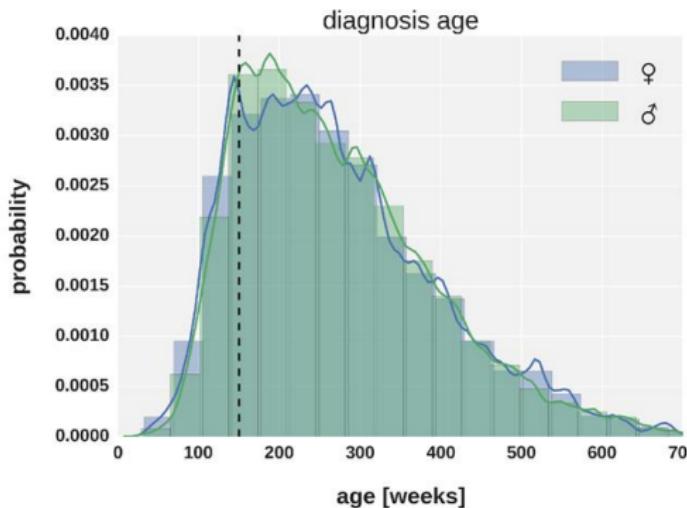
THE UNIVERSITY OF
CHICAGO BIOLOGICAL SCIENCES



ML: Breaking The Status Quo

Taking On The Grand Challenges

Autism Diagnoses Age Distribution



⑩ Can we predict ASD or CDD diagnoses from sparse medical history?

⑩ Can we get etiological insights?

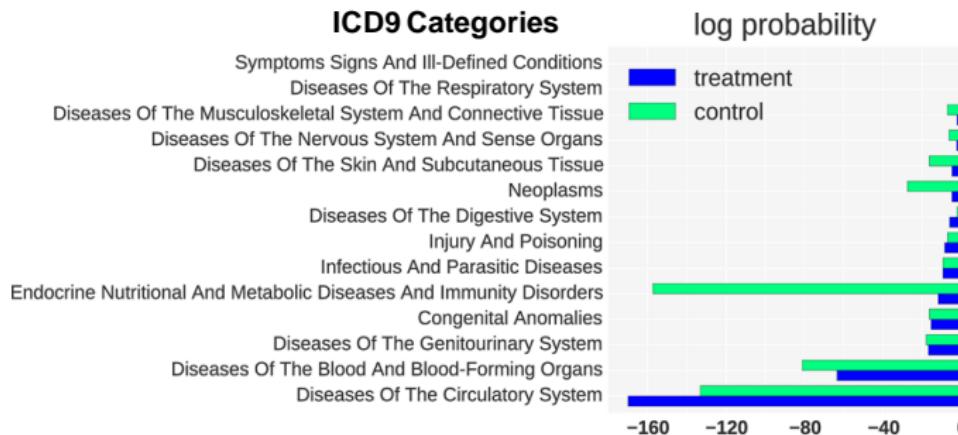




ML: Breaking The Status Quo

Taking On The Grand Challenges

Specific Disease Groups are Overrepresented

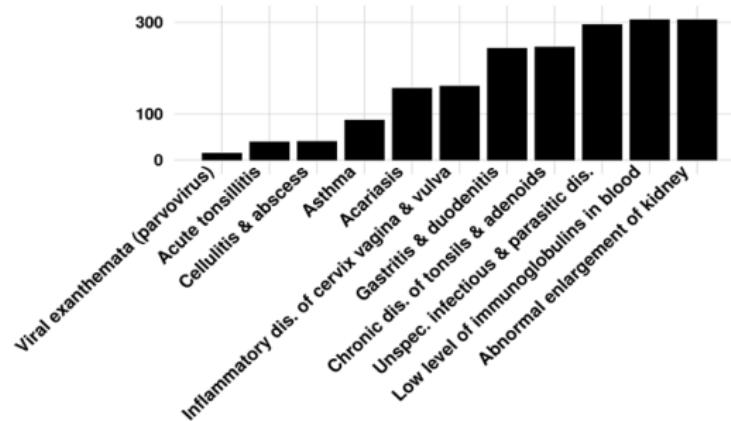




ML: Breaking The Status Quo

Taking On The Grand Challenges

Log-Odds Ratio Specific Disease Groups are Overrepresented

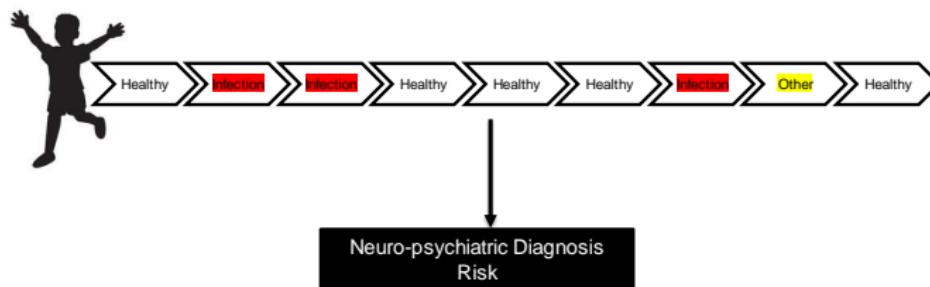




ML: Breaking The Status Quo

Taking On The Grand Challenges

Models from Population Data Predictions at the Individual Level

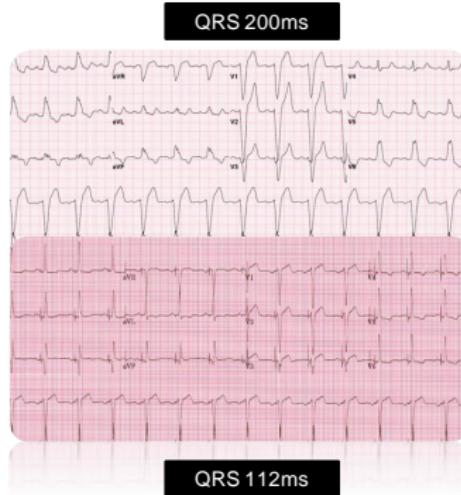
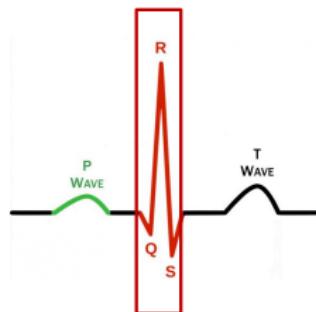




ML: Breaking The Status Quo

Taking On The Grand Challenges

Accurate Prediction Of Left Bundle Branch Blocks From Surface EKG

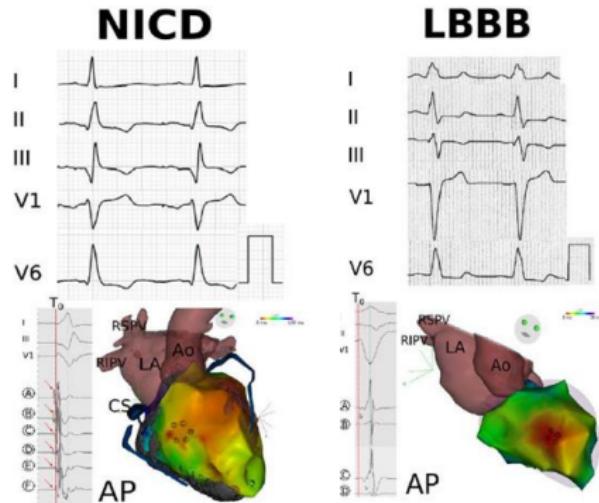




ML: Breaking The Status Quo

Taking On The Grand Challenges

What's a left bundle?



ACC/AHA/HRS Criteria for LBBB (2009)

- QRS \geq 120 ms
- Notched or slurred R-wave in leads I, aVL, V5, and V6
- Absent q wave in I, V5, V6
- R peak time > 60 ms in V5 and V6
- ST and T waves usually opposite to QRS

In LBBB: Passive Activation of the Left Bundle

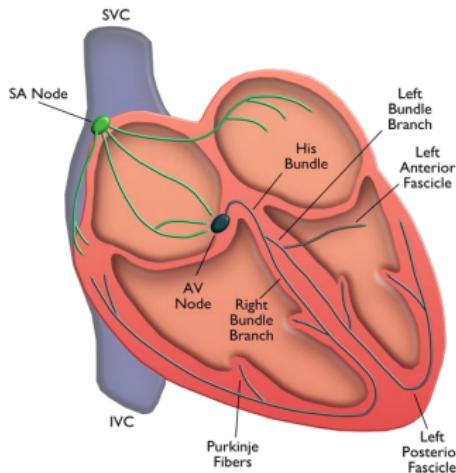




ML: Breaking The Status Quo

Taking On The Grand Challenges

Objectives



- His Bundle Pacing in Lieu of LV Lead for CRT
- Determine which cases will correct with HBP
- Determine level and extent of conduction system block
- Make this determination with surface leads instead of invasive EP studies





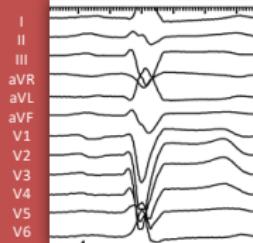
ML: Breaking The Status Quo

Taking On The Grand Challenges

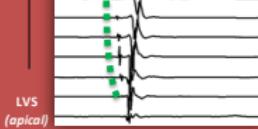
CRT Lab UChicago

Normal His-Purkinje Activation

Sinus Rhythm QRS 135

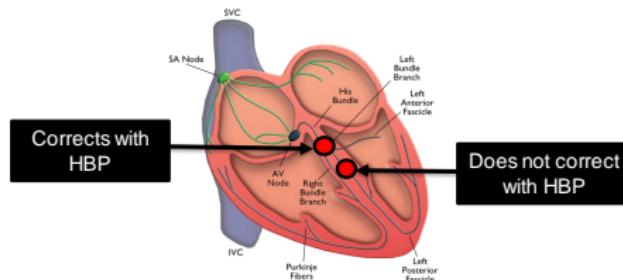


LVS (basal)



LVS
(apical)

- ❑ No surface ECG criteria could predict QRS correction.
- ❑ Wide QRS with intact Purkinje activation cannot be corrected.



	Intact Purkinje Activation	Delta LV His to Mid-Septum (ms)	Latest LV to QRS end (ms)	Mid-Septal QLV (%)
Correction	0%	47±21	58±13	67%
No Correction	75%	-19±34	74±17	51%
p-value	0.0001	<0.0001	0.0132	0.0001

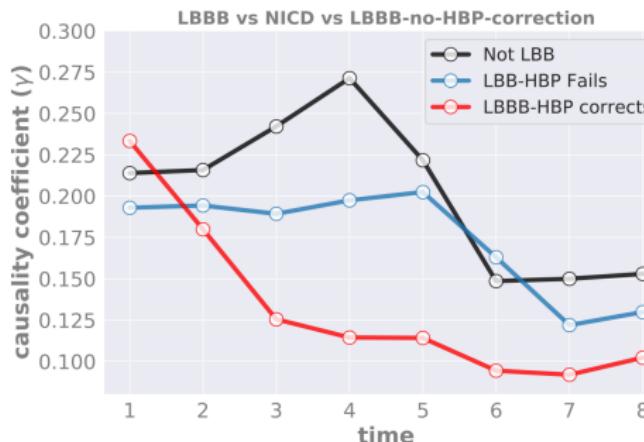


ML: Breaking The Status Quo

Taking On The Grand Challenges

Deep Granger Nets: Identifying Phenotype From Surface EKG

- I. Chatopadhyay, Causality Networks. [CoRR abs/1406.6651](https://arxiv.org/abs/1406.6651) (2014)



- ~96% accuracy in identifying three classes
- Etiological insight
- Small cohort tested.
Larger samples being prepared



THE UNIVERSITY OF
CHICAGO BIOLOGICAL SCIENCES