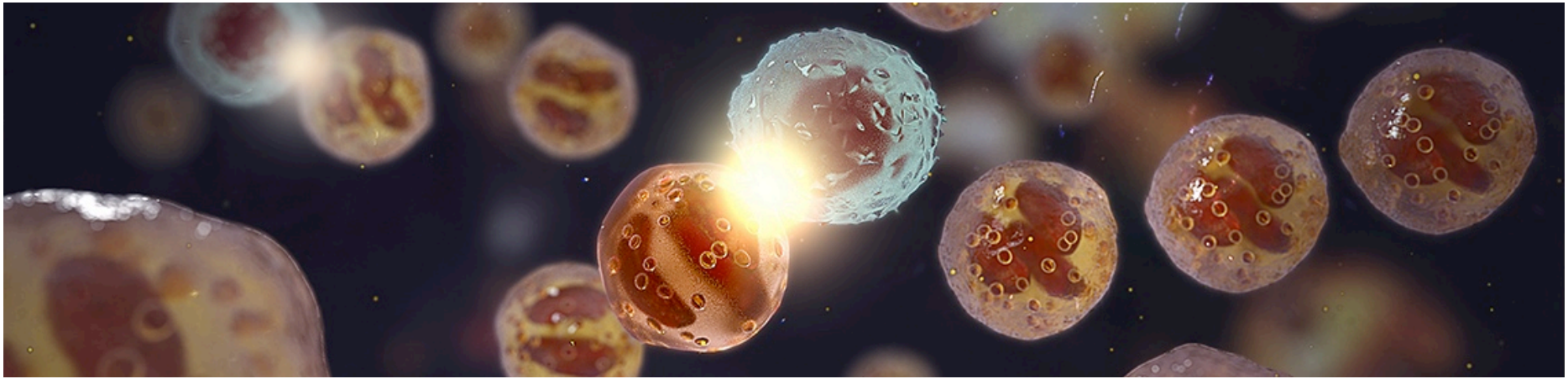
The Most Promising Areas of Machine Learning in Drug Development --with a focus on developing targeted therapies in oncology

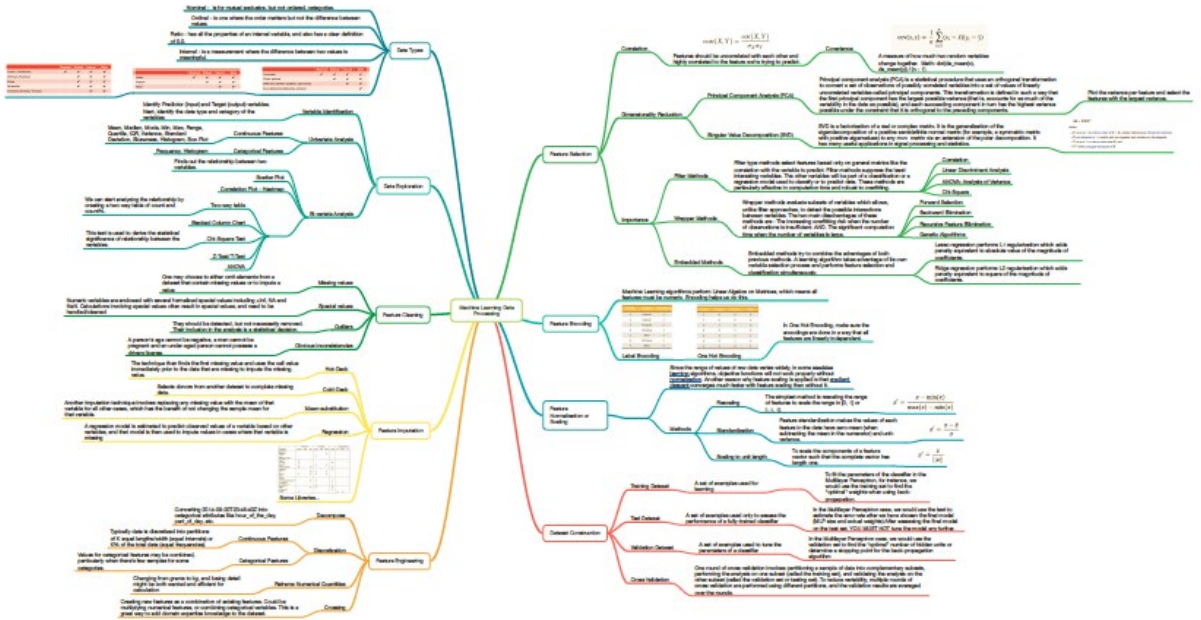
James Dunyak, Nidal Al-Huniti

26 FEB 2019



Machine Learning Algorithm Mind Map

Many packages are available to build and interpret machine learning algorithms: the challenge in ML is data, not tools



And many others…

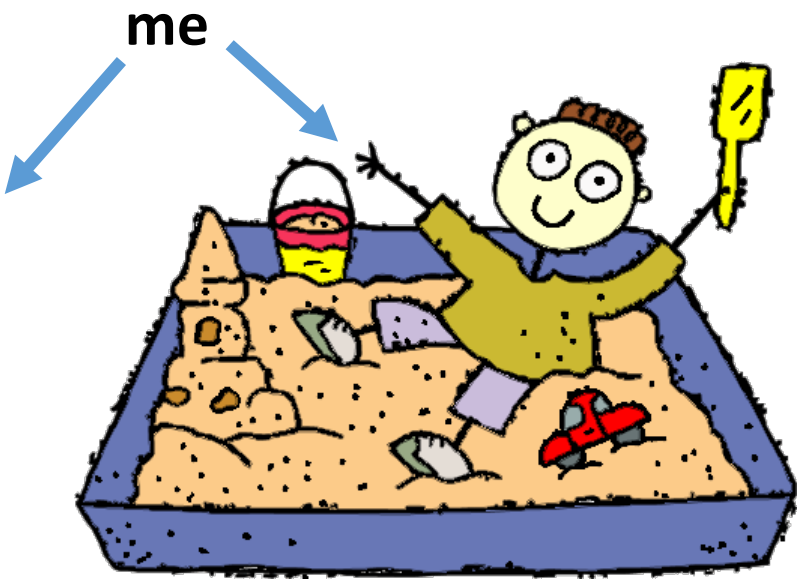
Many machine learning algorithms use the same underlying geometry resulting in similar performance when diligently applied.

Machine learning literature is vast and often novel with evocative names. Don't take it too seriously.

If you can imagine it, a machine learning algorithm can be invented to sort of do it

Fuzzy Probability Neural Networks, anyone?

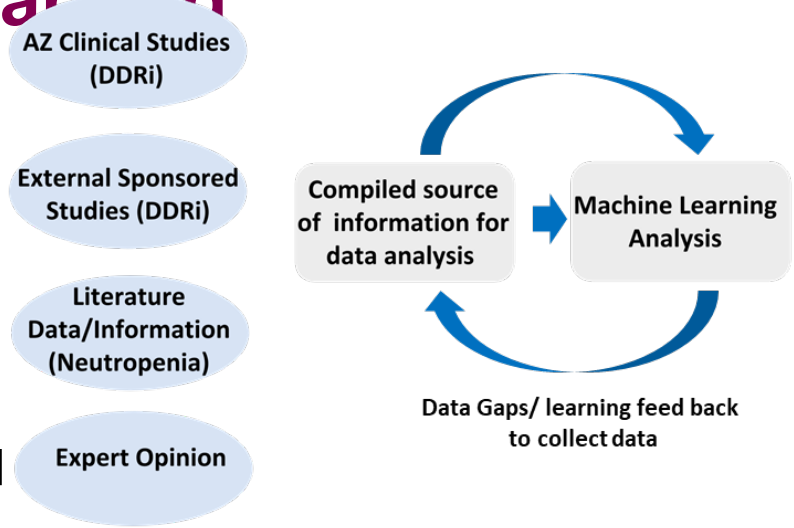
1. "Fuzzy Number Neural Networks," J.P. Dunyak and D. Wunsch, in *Fuzzy Sets and Systems*, 108, p. 49-58, November, 1999.
2. "Fuzzy Regression by Fuzzy Number Neural Networks," J. Dunyak and D. Wunsch. *Fuzzy Sets and Systems*, v. 112, n. 3, p. 371-380, June 2000.
3. "A Theory of Independent Fuzzy Probability for System Reliability," J. Dunyak, D. Wunsch, and I. Saad. *IEEE Transactions on Fuzzy Systems*, v.7, no. 3, p. 286-294, June 1999.



Assistant professor seeking tenure

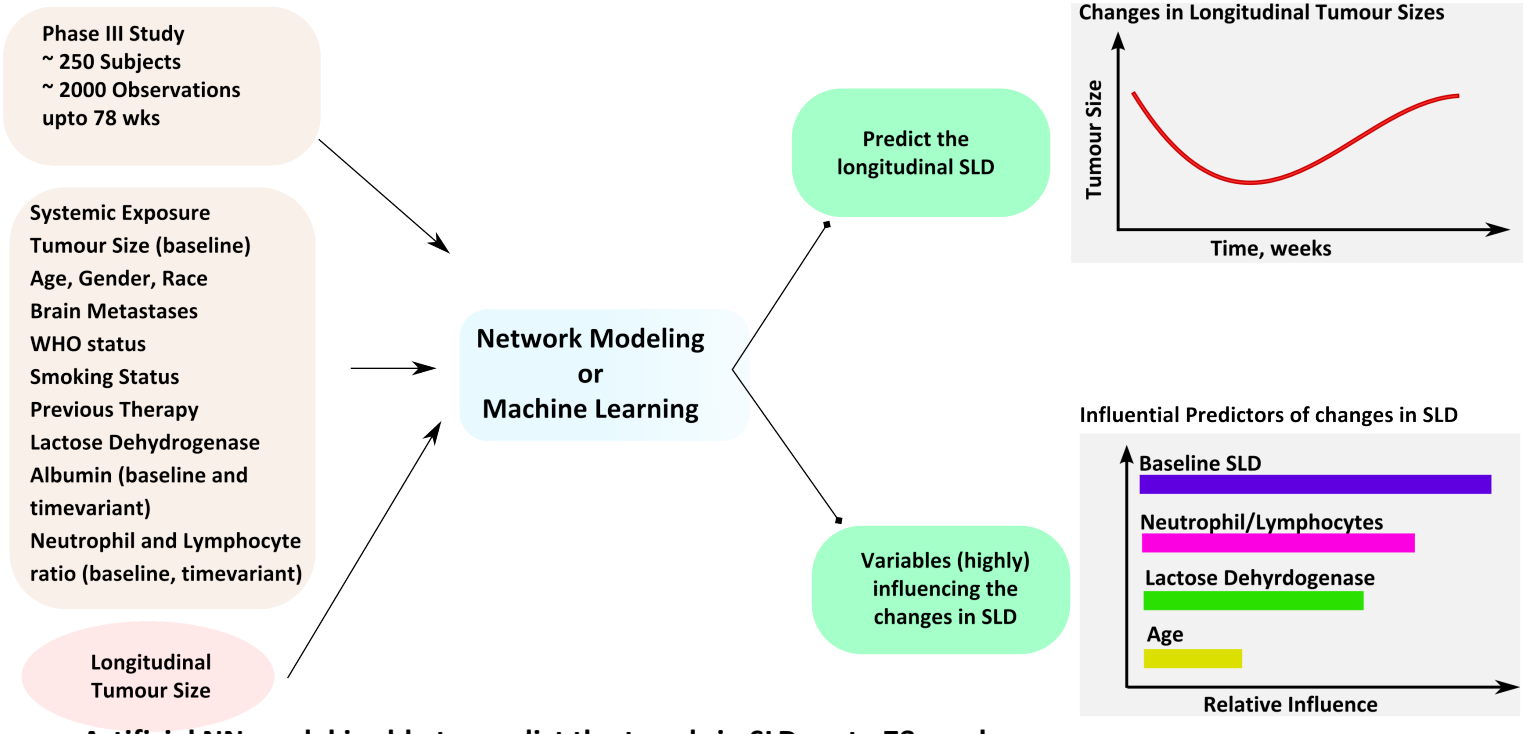
Unsupervised learning: Integrating knowledge and discovery using deep learning

- **Capturing clinical insights and knowledge by encoding drug-pathway interaction for learning network input**
  - Identify disease-specific pathways
  - Capture drug interaction (up and down regulation) with pathways and networks
  - Encode interaction of drug with identified pathways for deep learning input
- **Use deep learning to discover new relationships based on drug-pathway interactions**



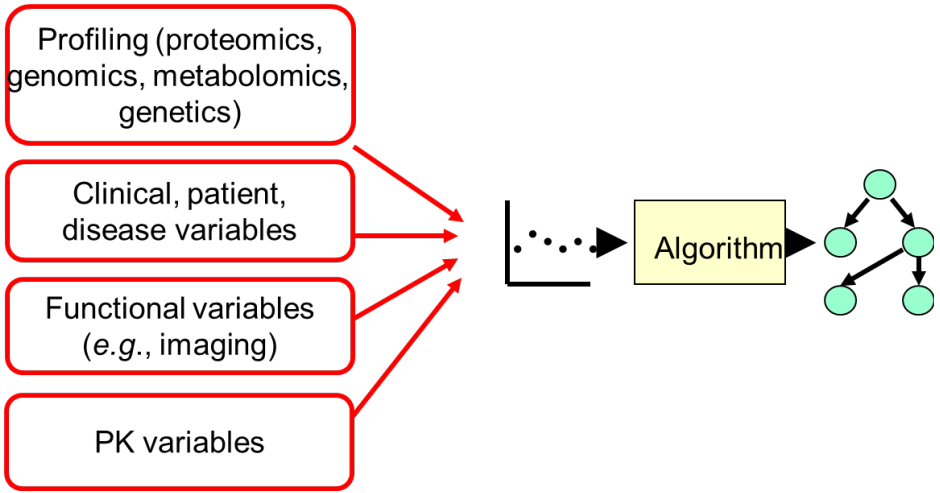
Deep learning using a biomolecular network-based representation can reflect, illustrate and learn the relationships among drugs, disease-related genes, therapeutic targets, and disease-specific signaling pathways as a system.

Quantitative Clinical hrarmacology example: Artificial Neural Networks to predict the changes in tumour sizes



- **Artificial NN model is able to predict the trends in SLD up to 78 weeks.**
- **Baseline SLD, NLR at weeks 6 and 18, LDH at baseline, and age at baseline were identified as the most influential predictors for changes in longitudinal SLD.**

Platform trials will require integration and modeling of many rich clinical data sources



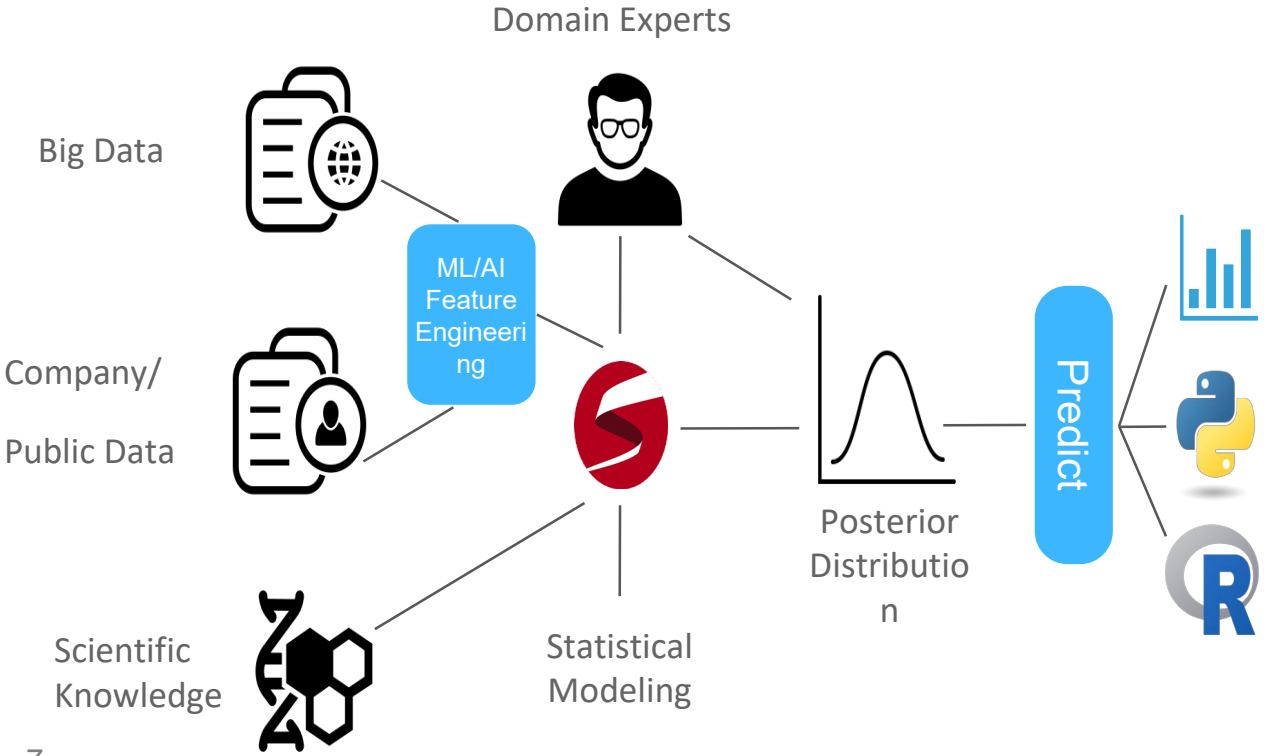
The Quantitative Clinical Pharmacology focus and challenge:
Put "clinical" in machine learning

How do we move machine learning and other advanced modeling methods from "hypotheses generation" to making actionable development decisions based on real clinical data?

Based on my personal experience, leaving the machine learning sandbox when using clinical data is difficult and high risk.

Advancing precision medicine from integrating genomics and real-world clinical phenotype evidence using ML and Bayesian modeling

Integration of genomics, clinical trial outcomes, and real-world clinical phenotype evidence are expected to drive to reveal the full landscape of human cancer



Objectives

- Identify predictive biomarkers for disease subtyping
- Relate molecular genomic features to clinical phenotypes
- Create a generic computational framework
- Develop genomics knowledge base
  - Omics data matched with clinical outcomes
- Develop a high-performance feature selection method
- Identify predictive biomarkers for cancer subtyping, using ML methods
- Relate patient features to clinical phenotypes from EHR
- Validate methods

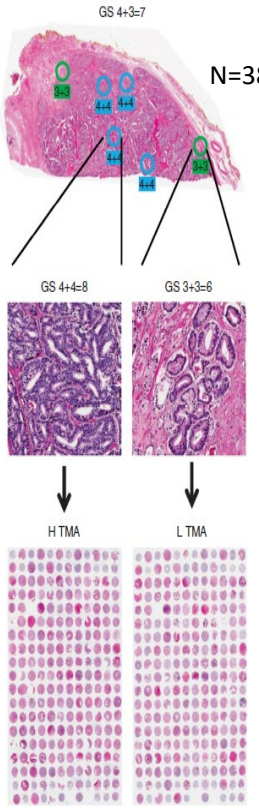
What is machine learning really good for, anyway?

- Machine learning is **GREAT** for:
  - Conference posters
  - Journal articles (often with irreproducible results, including mine)
  - Slide decks
  - Corporate demos and dog-and-pony-shows
  - Corporate awards and AZ postcards
  - Getting funding in academic and corporate settings
  - Getting tenure (certainly worked for me!)
- Machine learning is **CHALLENGING** when:
  - "Real world" **generalization** is required
  - New data from other sources must be analyzed
  - The algorithm will actually be used
  - Success in the real world can be measured
- I have personally participated in about $80M of failed machine learning projects, including one which destroyed the company, spanning mid 1980s to 2015
- I have also participated in successful machine learning projects
- **How did successful and failed projects differ?**

I can never tell whether I am the dog or the pony

Moving from ML algorithm development to real-world application

Goal: Develop an assay algorithm to stratify patient risk in prostate cancer based on needle biopsies



Step 1: Biological criteria

Step 2: Technical criteria

Step 3: Performance-based criteria

Initial list: 160 biomarker candidates

Prioritised list: 120 candidates
1. Biological relevance for cancer
2. *In silico* analysis
3. Human protein atlas
4. Monoclonal Ab availability

Technical validation: 62 candidates
DAB: specificity, signal intensity
IF: specificity, signal intensity, dynamic range, epitope stability, staining uniformity

Biomarker selection on H&L TMAs cohort
39 Ab candidates tested
Univariate performance, multivariate performance

Final selection
12 Markers identified
Biological relevance, published literature, univariate and multivariate analysis, technical considerations

Machine learning/ regularization/ data driven algorithm step, including automated image decomposition This would provide "robustness."

11 candidate biomarker selection from previous study *(Shipitsin et al.; ref. 22)*

- *Uniquely developed quantitative multiplex proteomics imaging (QMPI) platform used*
- *Selected markers had univariate performance for both aggressiveness and lethal outcome*
- *Marker performances were resistant to sampling variability*

Multicenter biomarker assay model development study
Development cohort size (381)

- *Using our QMPI platform, marker assay results compared with prostatectomy pathology to calculate odds ratio*
- *Biomarker frequency determined in top multivariate assay models*
- *Identification of 8-biomarker assay model with predictive value*

Clinical validation of biomarker assay model
Validation cohort size (276)

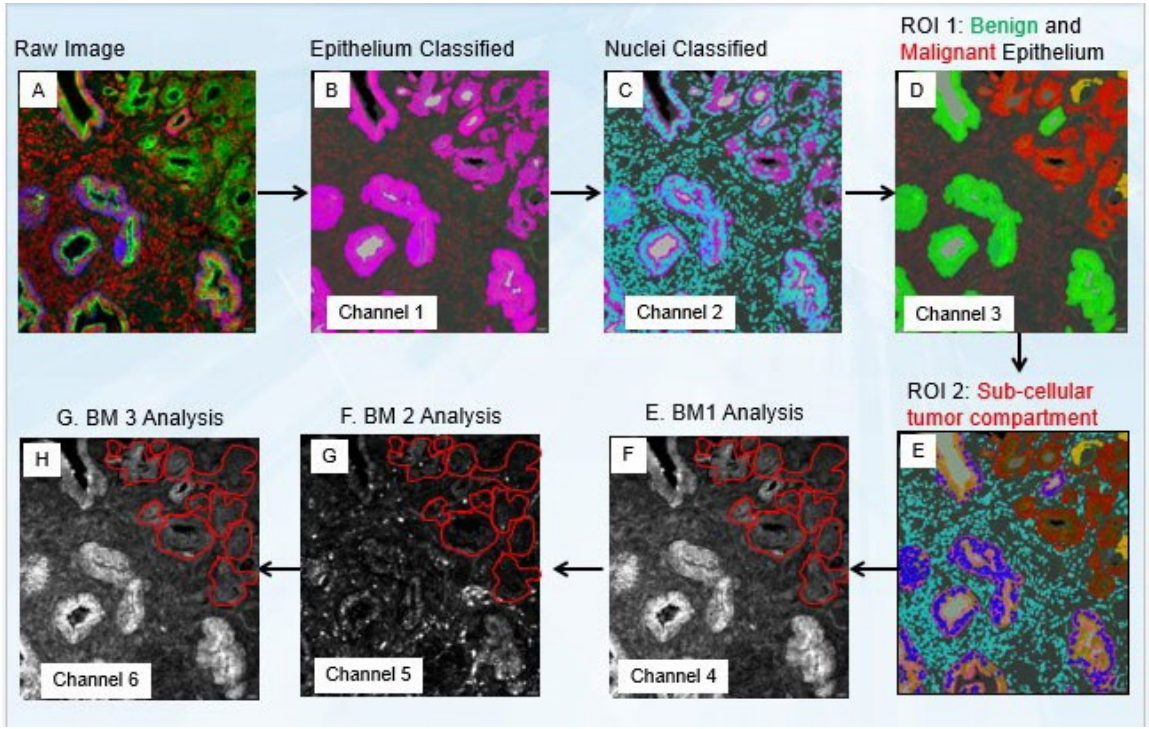
- *Validation of 8-biomarker assay model with identification of defined endpoints of "favorable" vs. "nonfavorable" as well as "GS6" vs. "non-GS6" pathology in patients*
- *Biomarker assay "risk-scores" comparision with current risk stratification systems (NCCN and D' Amico) with positive predictive value*
- *Calculation of added benefit using net reclassification index*

Development and Clinical Validation of an In Situ Biopsy-Based Multimarker Assay for Risk Stratification in Prostate Cancer
Peter Blume-Jensen et al Clinical Cancer Research, 2015

Moving from ML algorithm development to real-world application

Goal: Develop an assay algorithm to stratify patient risk in prostate cancer based on needle biopsies

Florescent labeling and ML for automatic image processing



The approach worked great in its FIRST clinical trial.

Development and Clinical Validation of an In Situ Biopsy-Based Multimarker Assay for Risk Stratification in Prostate Cancer
Peter Blume-Jensen et al Clinical Cancer Research, 2015

Moving from ML algorithm development to real-world application: What was challenging

Goal: Develop an assay algorithm to stratify patient risk in prostate cancer based on needle biopsies

Machine Learning components

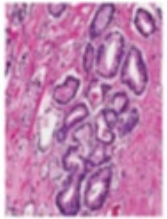
Sample mounting, staining of multiple slides, fixing fluorescence → Scanning → Automated image decomposition and signal level estimate using Definiens → Application of signature → Result

- Sample age (archival vs. freshly fixed)
- FFPE process variation
- Prolong Gold manufacturing batch changed
- Stability of Vectra Intelligent Slide Analysis Systems
- Definiens algorithm reproducibility
- Signal processing methods were updated post-hoc to try to "fix" interpretation of signature components, with new validation required.

Specification:
"Protects dyes from fading during imaging"
= no spec at all!

Moving from ML algorithm development to real-world application: What went right

Goal: A Five-Gene Hedgehog Signature for Patient Preselection Tool for Hedgehog Inhibitor Therapy in Medulloblastoma



Micro-dissection of FFPE slide

RNA extraction 5 signature + control genes

Machine Learning component

TaqMan-based RT-PCR assays gene expression profiling

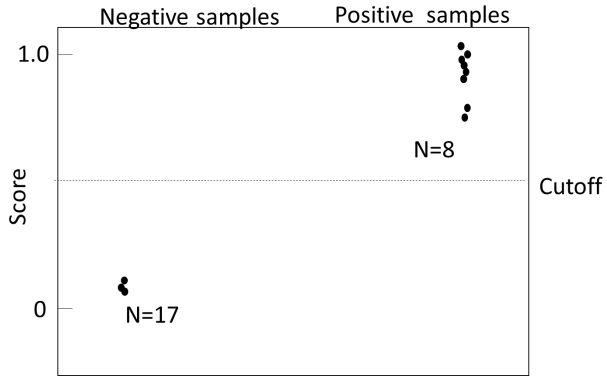
Application of signature

Machine Learning component

Result: patient is /is not candidate for treatment

FFPE process variation still poorly explored

ABI PCR system spec and calibration not sufficient for this use. We needed to send a specific technician and calibration plate from Cambridge MA to Carlsbad CA to successfully transfer assay and have ML algorithm work.



What went right? The machine learning algorithm was **DESIGNED** as part of a system

- Clear intended use
- Clear risk-based requirements
- Followed a design process:
  - POC/ feasibility
  - Intended use and requirements
  - Design (**including input data requirements**)
  - Design verification
- **ML design did not end at the POC dog-and-pony show!**

Shou et al. Clinical Cancer Research 2015

← This paper is actually reproducible!

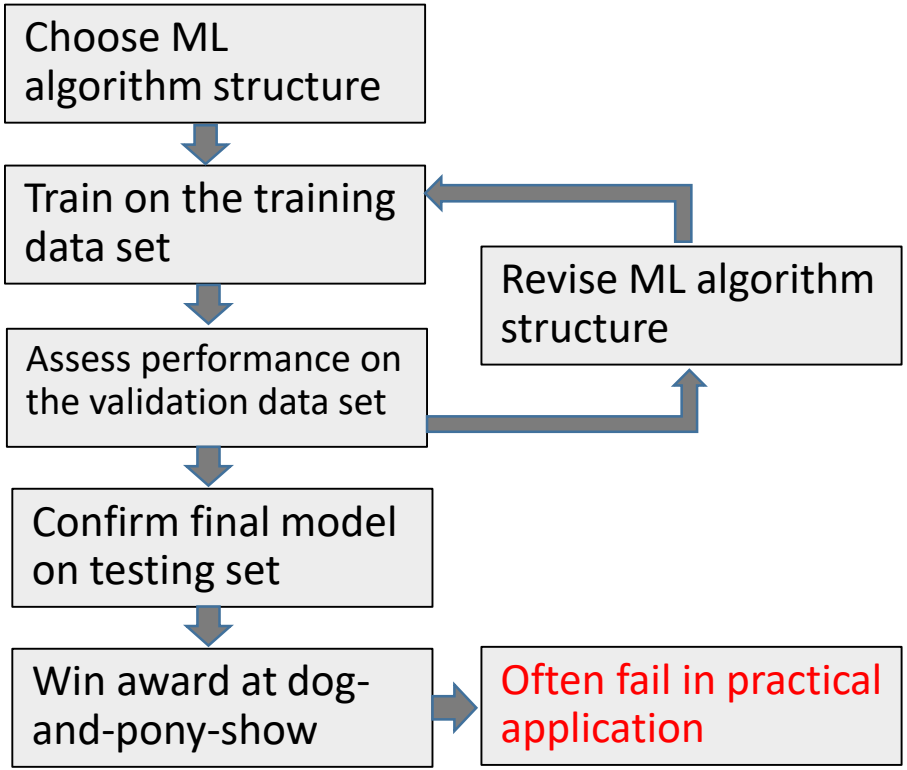
Issues with lack of model interpretation

- No interpretation = limited insight
- Common data and model issues are difficult to address using standard methods
- Confounding
- Measurement artifacts

Machine learning algorithms **ALWAYS** look behind the curtain, whether we want them to or not.

And they don't tell us what they find.

The standard approach: randomly divide data into training, validation, testing sets



Why do these machine learning algorithms often fail in application?

- Input data is not typical of the training/validation/testing data

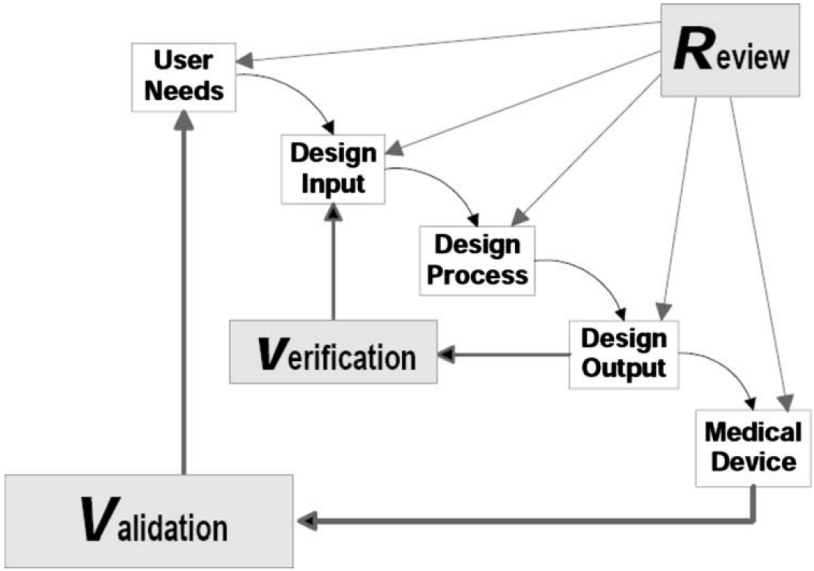
  The model fails to "generalize."

- Algorithm result is used in an unexpected way
  - Most ML algorithms reproduce an input/output map
  - This map may not be that relevant to the question at hand
- Quality training data is not enough

We know how to fix this: engineer ML algorithms
(but you may not like the answer)

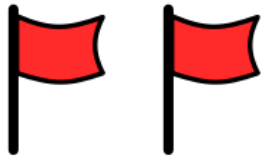
Take responsibility for the final application

- Stage 0: Feasibility
  - Play in the sandbox, but don't stop there!
- Stage 1: Identify critical design issues
  - Intended use, risks, requirements
- Stage 2: Design to requirements
  - An algorithm is unlikely to meet a requirement not addressed during design
- Stage 3: Verify requirements are met
- Stage 4: Continue algorithm validation through its lifetime
  - New data, new uses, new risks



Famous CDRH waterfall plot (actually taken from Canadian medical authorities)

Understand critical design issues

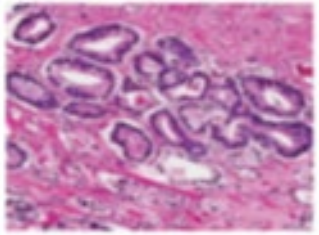


Stage 1: Identify critical design issues

- Intended use –Not just “the algorithm models liver toxicity”
  - What input data will be used, and what level of generalization is required?
    - Data from new patients, clinical trials, clinical sites, different labs, different laboratory instruments, different reagent lots, ….
    - Data collected over the next year, five years, ten years, forever?
  - How will the output be used?
    - Kill projects, influence team decisions, generate hypothesis to motivate other studies, select patients, …
- Risks –What risks are associated with algorithm errors?
    - To the project, the company, future patients, public health, ….
    - No meaningful risks = few benefits
- Requirements to meet intended use in context of risk
  - Give appropriate answers for data from new patients, new laboratory sites, ….
  - Have an error rate of less than ….
  - Robust to ….

Red flags:

- Preclinical biomarkers
- All innovative assays
- Biomarkers requiring testing at central labs –multiple trials?
- Samples collected for another purpose (ie. FFPEs, images,…)



References

Xi. Li,G. Qin,Q. Yang,L. Chen,L. Xie, “Biomolecular Network-Based Synergistic Drug Combination Discovery, BioMed Research International, Volume 2016,

L. Huang, F. Li, J. Sheng et al., “DrugComboRanker: drug combinationdiscoverybasedon target network analysis,” Bioinformatics,vol. 30, no. 12, pp. I228–I236, 2014.

P.Li, C.Huang, Y. Fuet al., “Large-scale explorationandanalysis of drug combinations,” Bioinformatics, vol. 31, no. 12, pp. 2007–2016, 2015.

S. Chandrasekaran, M. Cokol−Cakmak, N. Sahin et al., “Chemogenomics and orthology−based design of antibiotic combination therapies,” Molecular Systems Biology, vol. 12, no. 5, p. 872, 2016.

G. Jin, H. Zhao, X. Zhou, and S. T. C. Wong, “An enhanced Petri-Net model to predict synergistic effects of pairwise drug combinations from gene microarray data,” Bioinformatics, vol. 27, no. 13, pp. i310–i316, 2011.

Z. Wu, X.-M. Zhao, and L. Chen, “A systems biology approach to identify effective cocktail drugs,” BMC Systems Biology, vol. 4, no. 2, article 7, 2010.

J. Yang, H. Tang, Y. Li et al., “DIGRE: drug-induced genomic residual effect model for successful prediction of multidrug effects,” CPT: Pharmacometrics and Systems Pharmacology, vol. 4, no. 2, pp. 91–97, 2015.

Y. Sun, Z. Sheng, C.Ma et al., “Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer,” Nature Communications, vol. 6, article 9481, 2015.

X.-M. Zhao, M. Iskar, G. Zeller, M. Kuhn, V. Van Noort, and P. Bork, “Prediction of drug combinations by integrating molecular and pharmacological data,” PLoS Computational Biology, vol. 7, no. 12, Article IDe1002323, 2011.

Xu, Y., Dai, Z., Chen, F., Gao, S., Pei, J., & Lai, L. (2015). Deep l Earning or drug-induced liver injury. Journal of Chemical Information and Modeling, 55 (10), 2085-2093.

Chen et al. Learning a hierarchical representation of the yeast transcriptomic machinery using an autoencoder model, BMC Bioinformatics 2016, 17(Suppl 1):9