Deep Learning Opportunities and Challenges in Drug Discovery and Biomarker Development

Xin Huang, Yan Sun

AbbVie

2019 Nonclinical Biostatistics Conference
May, 2019

Disclosure

The support of this presentation was provided by AbbVie.

AbbVie participated in the review and approval of the content.

Xin Huang and Yan Sun are employees of AbbVie Inc.

Disclaimer

- Not DL expert: statistician's perspective
- Not comprehensive: research question oriented, focus on Pharma R&D
- Not platform building nor deployment

Things expect to take away:

1. How DL is different from ML
2. When DL works better
3. What's statistician's role in DL

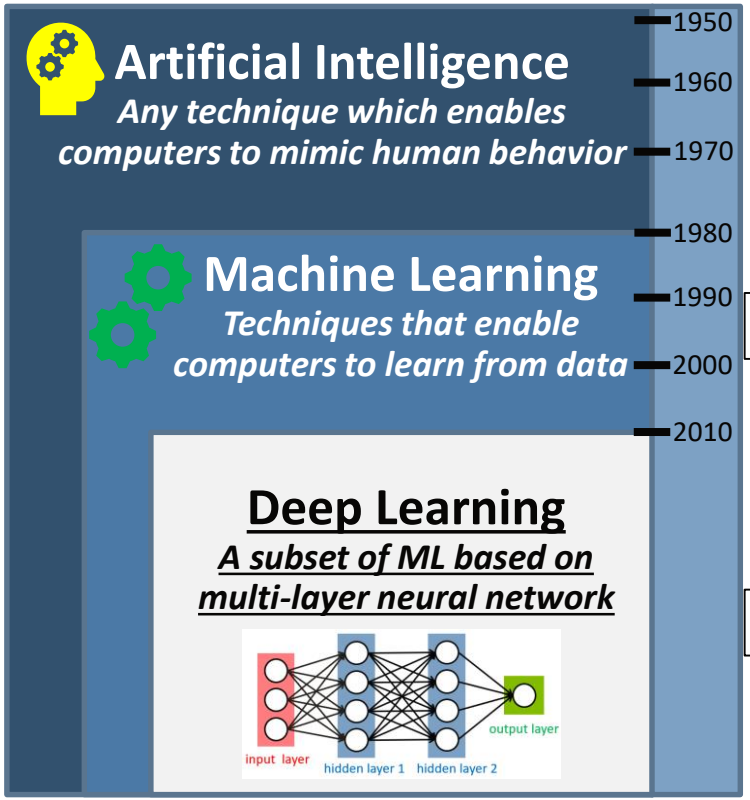
Things to think about:

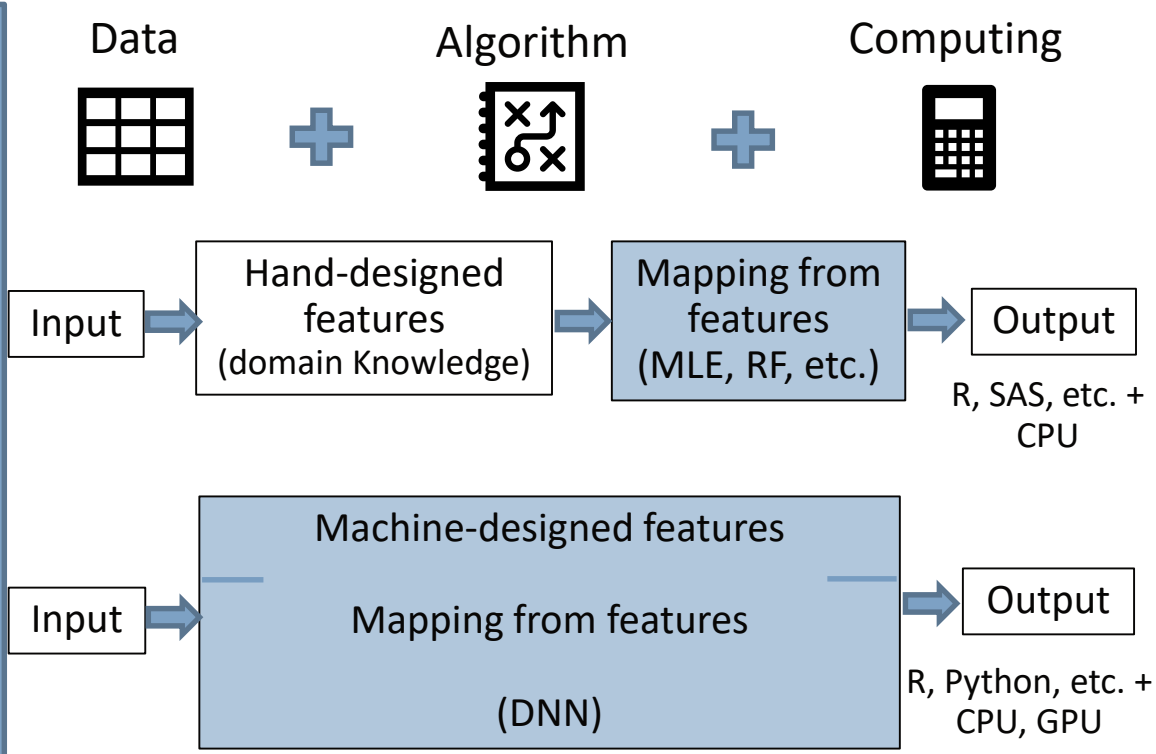
1. Where are the DL opportunities in my project
2. Whether DL can improve my current decision making

Outline

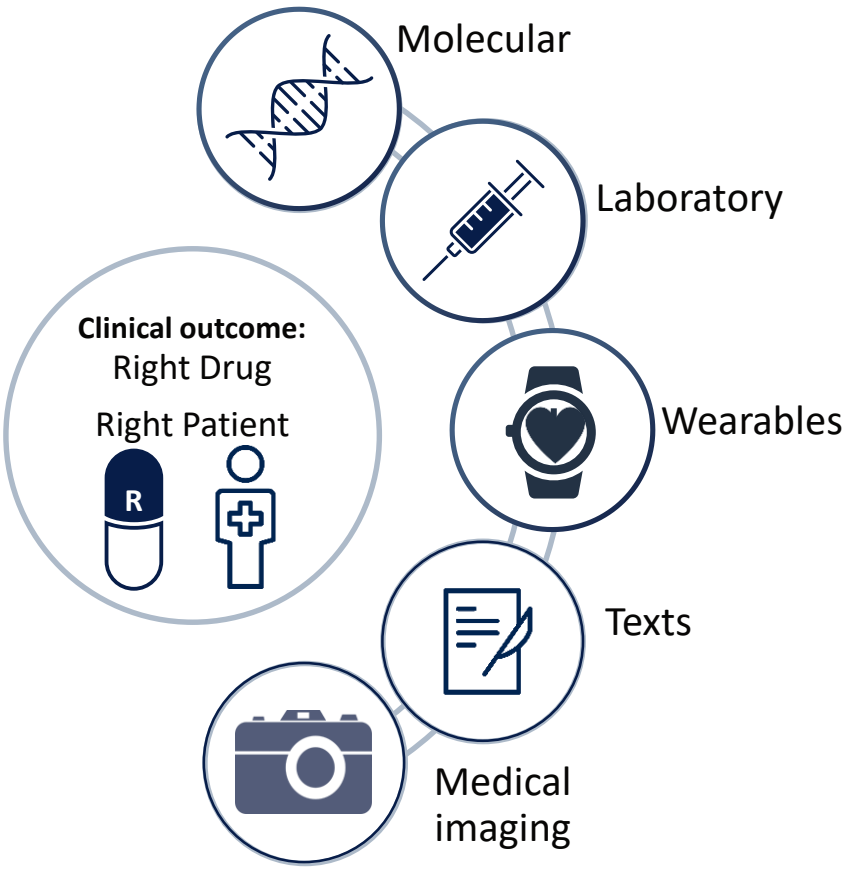
- Introduction
- A very brief introduction to Deep Learning
- Challenges and Opportunities in Pharmaceutical Development

What is Deep Learning?





When to use DL?



Data Example

Genomic
Proteomic

Renal
Blood parameters

Activity tracking
(spatial, temporal)

Physician notes
Social media

IHC, H&E, MRI

Unstructured Data
Feature not well defined

Challenges

Current stat tool cannot directly handle unstructured data

Prediction
(Patient Classification)

- Integrate all data types to achieve better accuracy

Estimation & Inference
(Efficacy, Effectiveness, Prognosis)

- Small sample size
- Inference after prediction

Interpretation

- Interpretation is often very difficult for neural network

How to build Deep Neural Network

- Build from scratch
  - Require large sample size: hundreds of thousands → millions of samples
    - Imagenet: 1000 classes with more than 10 million images
  - Require large amount of computing resources
    - Google Inception V4 used TensorFlow distributed machine learning system using 20 replicas, each running a NVidia Kepler GPU
  - Require deep understanding of neural network structure

- Stand on the shoulders of giants
  - Transferred learning borrows information from pre-trained DNN

- Software: R, Python
  - Keras package is available in both R and Python

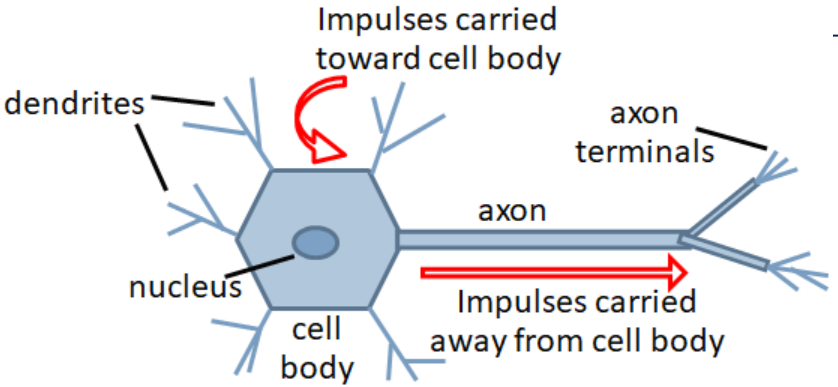
Outline

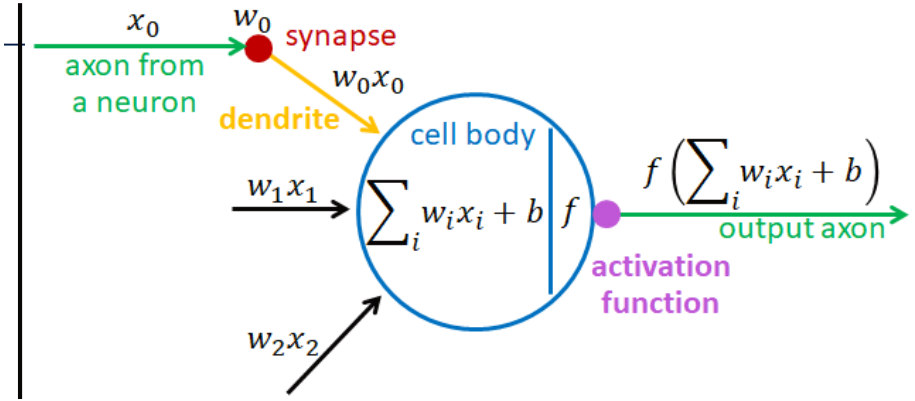
- Introduction

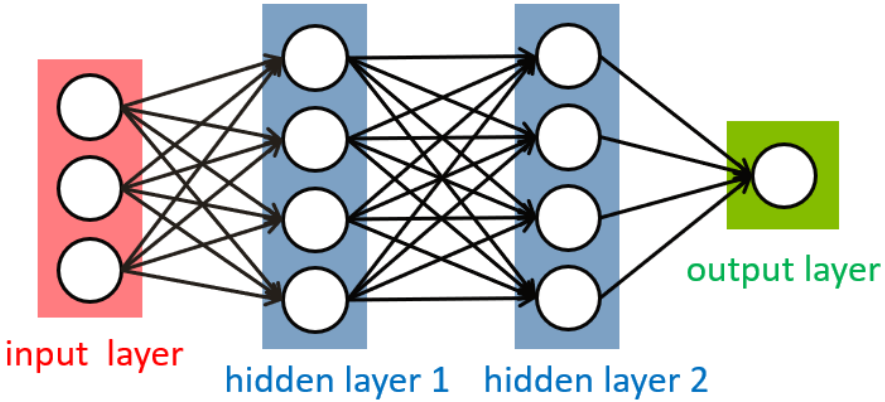
- A very brief introduction to Deep Learning

- Challenges and Opportunities in Pharmaceutical Development

Basics of Neural Networks



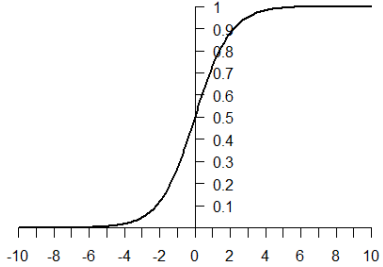




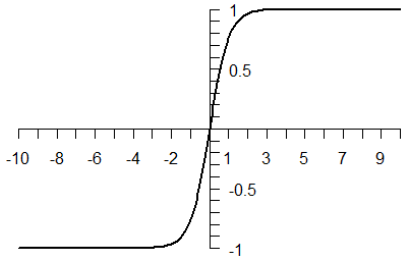
- Activation function
  - E.g., sigmoid function
- Loss function:
  - E.g., cross-entropy
- Optimization:
  - E.g., gradient descent with back propagation

Basics of Neural Networks

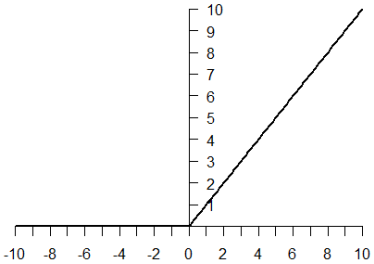
- Activation Function



Sigmoid: $f(x) = 1/1 + e^{-x}$

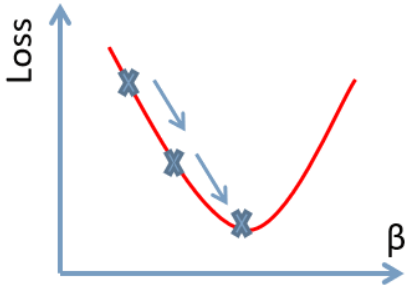


tanh: $f(x) = 2\sigma(2x) - 1$



ReLu: $f(x) = \max(0, x)$

- Loss function:
  - Cross-entropy for categorical outcome: $-\sum_i y_i \log(\hat{p}_i)$
  - L2 loss for continuous outcome: $\sum_i (\hat{y}_i - y_i)^2$
- Gradient Descent:



Basics of Neural Networks

- Regularizations and Model Improvement
  - Parameter Norm Penalties
    - L1, L2 Norm: Loss + Penalty

$$Cost\,function = Loss + \frac{\lambda}{2m} * \sum \|w\|^2$$

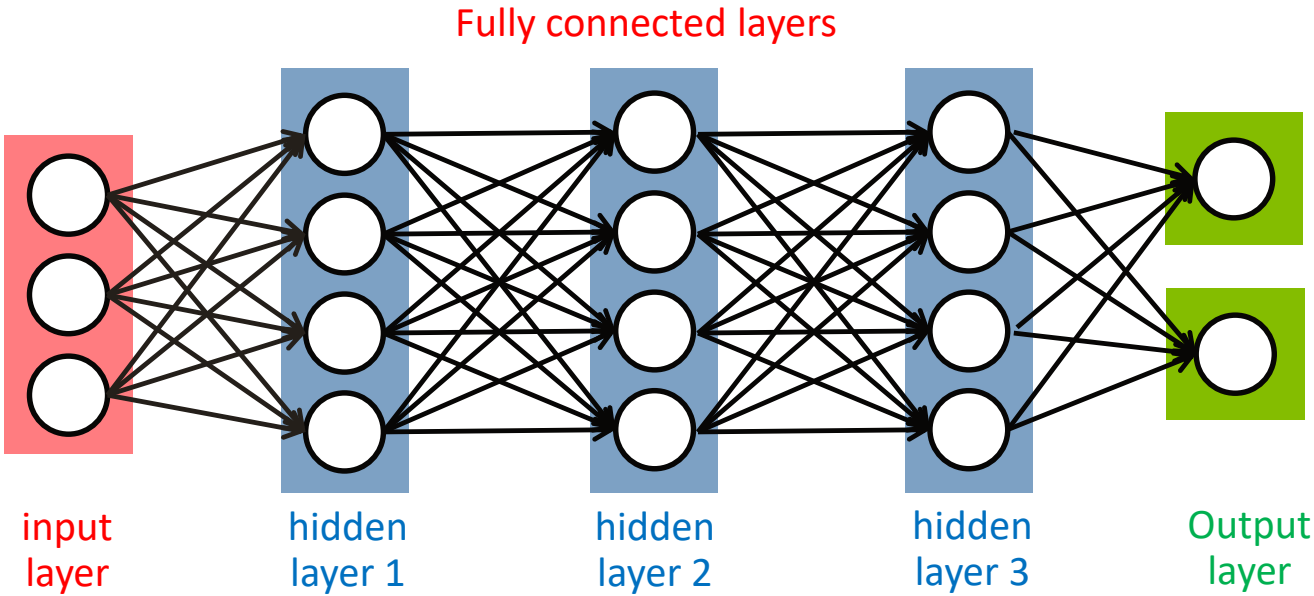
$$Cost\,function = Loss + \frac{\lambda}{2m} * \sum \|w\|$$

  - Dropout
    - At every iteration, randomly select some nodes and remove them along with all of their incoming and outgoing connections
  - Dataset augmentation
    - Increase the dataset by generating new data from existing data
    - Examples:
      - Rotating image
      - Scaling image
  - Early Stopping – most commonly used form of regularization in DL
    - Use validation set to determine the number of steps (epochs) when no improvements over the last number of epochs (user defined)

Convolutional Neural Network

- One of the most popular classes of Deep Learning
  - What people often mean when saying Deep Learning.

- What's wrong with "traditional" neural network?



- We capture all pixel level information.
- We treat the value of different pixels as a vector, like what we are used to do.
- Do we miss anything?
  - Spatial structure

Deep Convolutional Neural Network

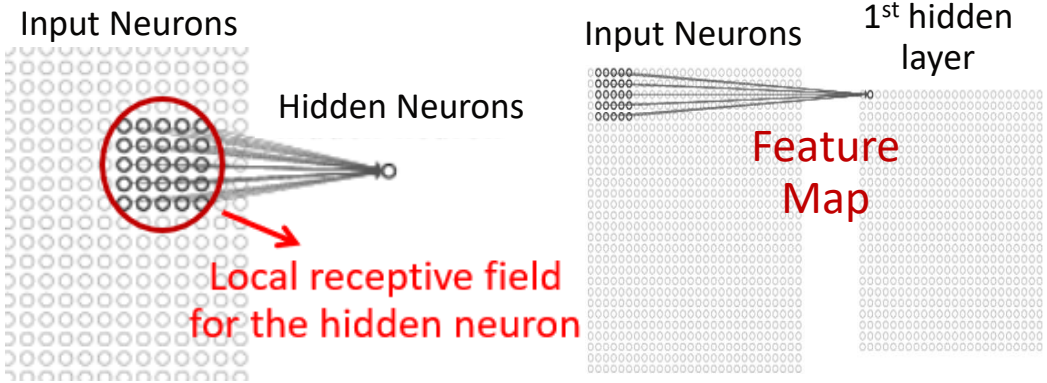
- CNN
  - Takes advantage of the spatial structure
  - Widely used for image recognition

  - 3 basic ideas:
    - Local receptive fields
    - Shared weights
    - Pooling

Deep Convolutional Neural Network

Local receptive fields

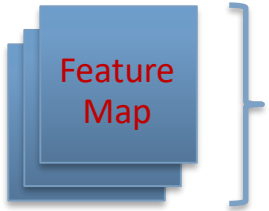
Treat pixel values as a matrix, the same as the original image.



- Each neuron in the first hidden layer will be connected to a small region of the input neurons
- Each connection learns a weight and a bias
- Slide the local receptive field across the entire image
  - Different stride length can be used

Shared weights and bias

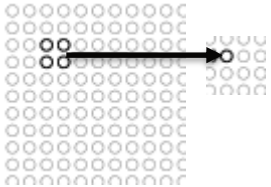
The weights and bias (connection) remain the same when sliding the field.



Convolutional Layer

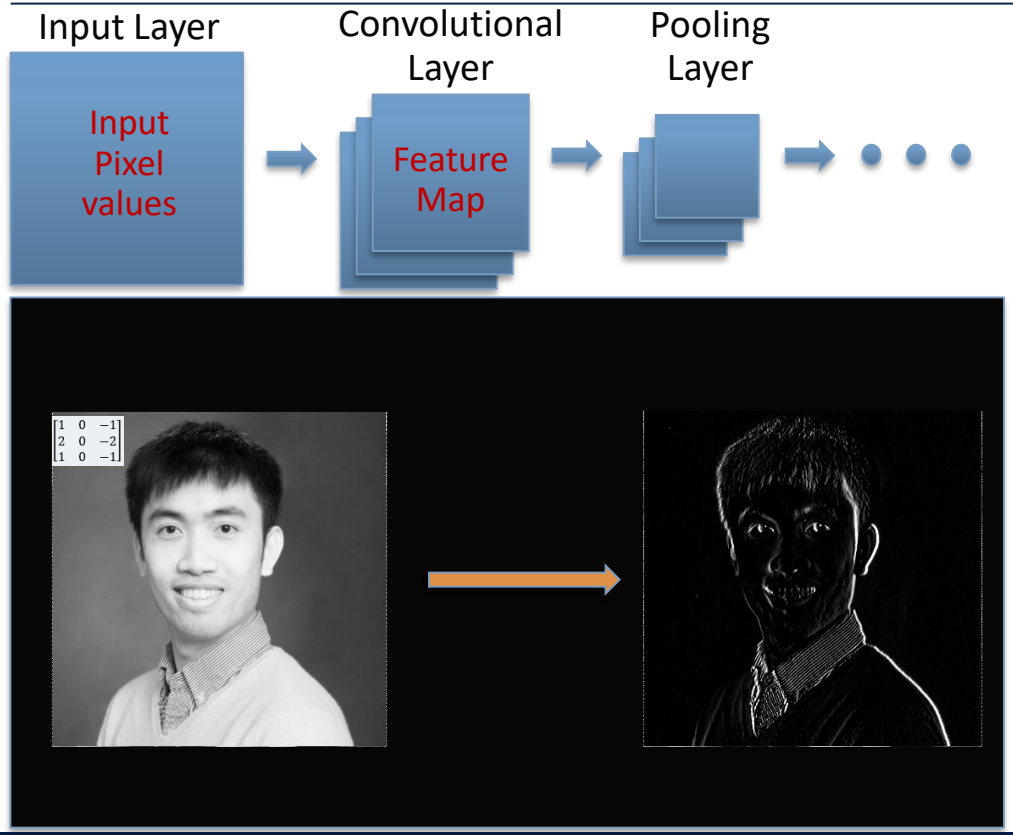
Pooling Layers

The pooling layer condenses feature map to a "smaller" version.

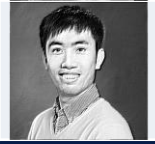


e.g., condense 4 neurons to 1 neuron by picking the maximum

Deep Convolutional Neural Network

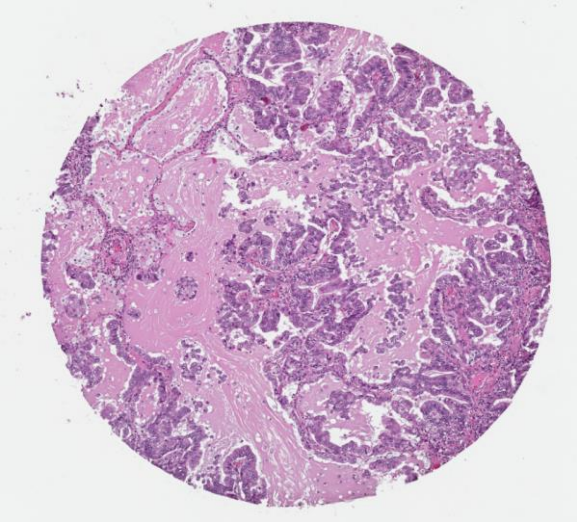


Why does it work?

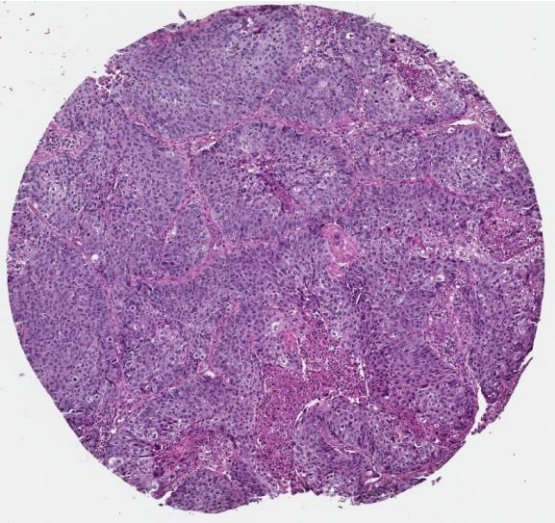
Operation	Filter	Convolved Image
Identity	$\begin{bmatrix}0 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$	
Edge detection	$\begin{bmatrix}1 & 0 & -1\\2 & 0 & -2\\1 & 0 & -1\end{bmatrix}$	
	$\begin{bmatrix}1 & 2 & 1\\0 & 0 & 0\\-1 & -2 & -1\end{bmatrix}$	
	$\begin{bmatrix}-1 & -1 & -1\\-1 & 8 & -1\\-1 & -1 & -1\end{bmatrix}$	
Sharpen	$\begin{bmatrix}0 & -1 & 0\\-1 & 5 & -1\\0 & -1 & 0\end{bmatrix}$	
Emboss	$\begin{bmatrix}-2 & -1 & 0\\-1 & 1 & 1\\0 & 1 & 2\end{bmatrix}$	

Lung Cancer Subtyping

- Deep Learning based digital pathology image analysis for lung cancer subtyping



Adenocarcinoma



Squamous

Image source: H&E image of tissue microarray (AbbVie procured NSCLC samples from vendors)

Lung Cancer Subtyping: Data and Platform

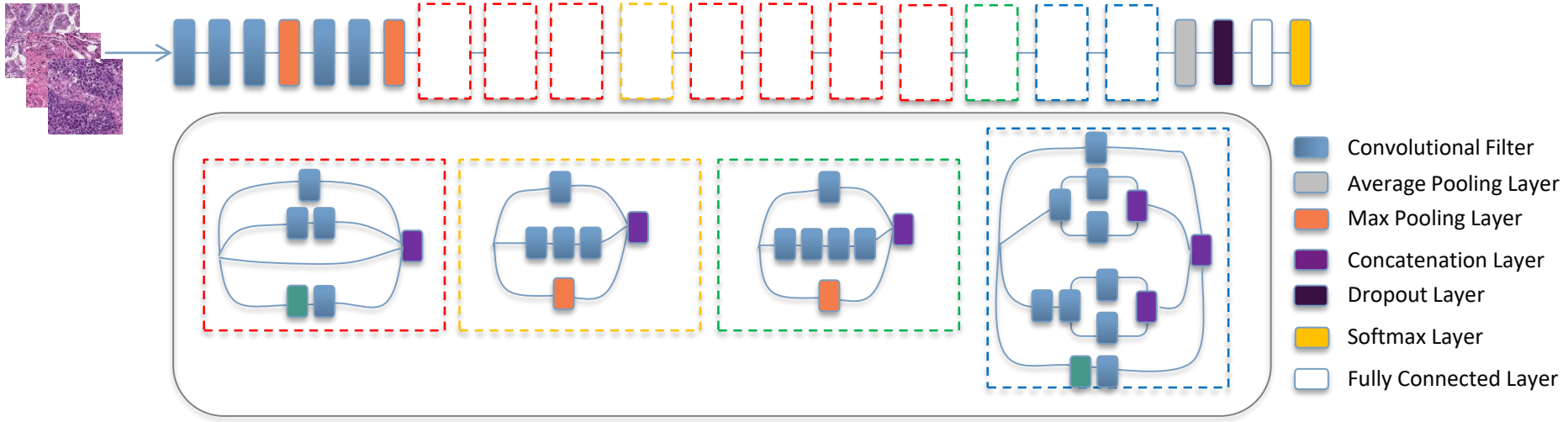
- H&E images
  - 30 images for training, 7 images for testing
  - Each image is labeled as adenocarcinoma, squamous or normal

- Data pre-processing
  - Pathologists help identify tumor/non-tumor region within each slide
  - Magnification selection: 10x
  - Patch cutting: 299*299
  - Data augmentation:
    - Generate more training data through transformation of the image

- Platform:
  - R + Keras (a high-level neural networks API)
  - GPU server (one Tesla V100 from Nvidia)
    - 30X faster than a CPU with 10 cores

Lung Cancer Subtyping: Transferred Learning

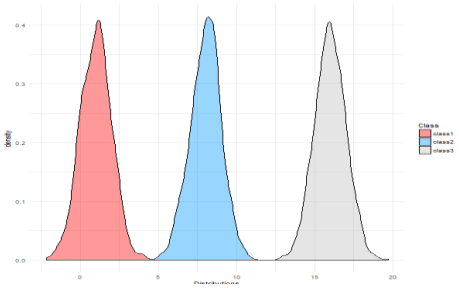
- Transferred learning:
  - Utilizes training results from a pre-trained neural network
  - Our current model utilizes the feature extraction part of a pre-trained Inception-v3 architecture, but reconstructs the classification layers specific to our dataset



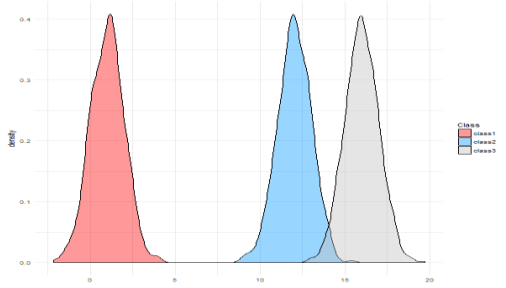
Inception-v3 Modules

Lung Cancer Subtyping: Training Strategy

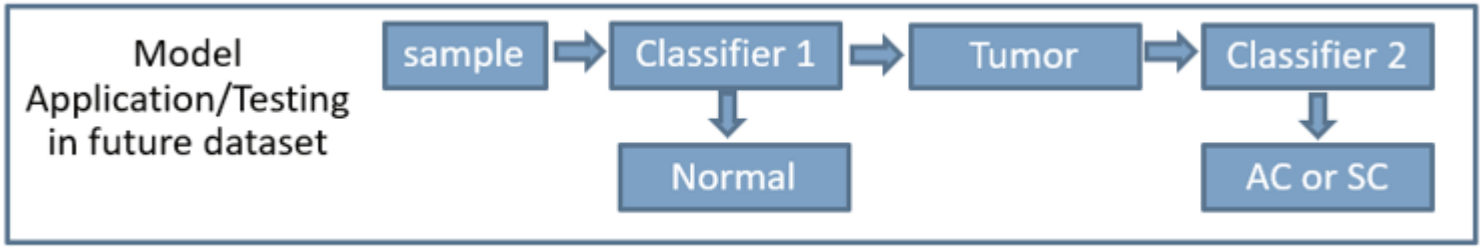
2+2-class training
vs.
3-class training



vs.



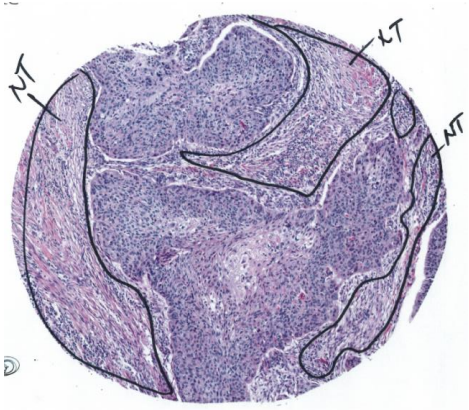
Our Strategy



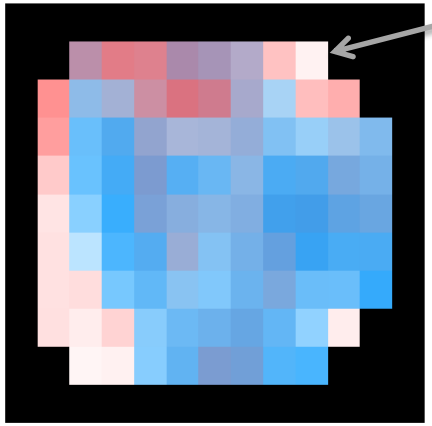
Lung Cancer Subtyping: Results

Overall Testing results	Predicted AC	Predicted SCC
Actual AC	3	1
Actual SCC	0	3

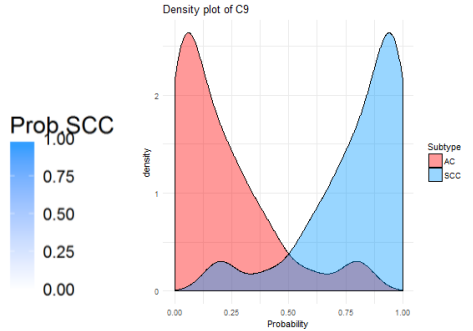
One Testing Result Example



Ground truth: SCC



masked by probability of Non-Tumor (white)



SCC

Predicted patch probability distribution

Outline

- Introduction
- A very brief introduction to Deep Learning
- Challenges and Opportunities in Pharmaceutical Development

Recap: Challenges in DL/ML for Pharma R&D

Data:

- Small data in pharma setting
- Potential bias in data
- Privacy and data sharing

Algorithm

- Current DL model mostly optimized for prediction accuracy
  (The optimization is problem specific: FDR control, Estimation accuracy, unbiased estimate, minimized variance)
- How to integrate analysis using both structure and unstructured data
- Interpretation

Computing

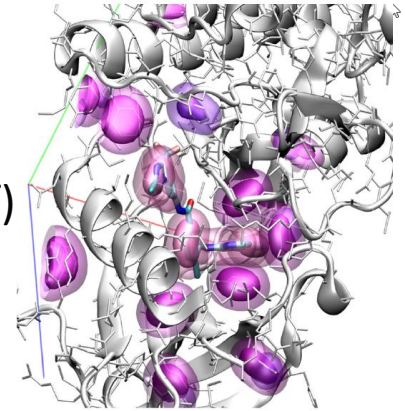
- Money and time in training project specific models

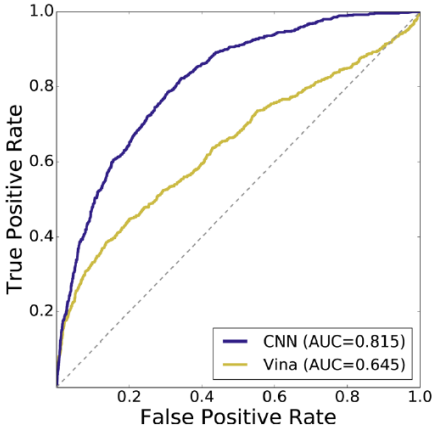
Drug Discovery: protein-ligand binding prediction, Drug synergy screening

Question:
- Which compound binds to the target of disease the best? (over 1.5 million candidates)
- Which combination of the compounds has the synergy effect for which indication?

Statistical Challenge:
- should optimize for FDR control

(best rank the most important compounds, cutoff)
- increase the probability of success
- how to quantify the uncertainty
  (docking score, synergy score)





Source: DOI: 10.1021/acs.jcim.6b00740

Oncology: PD-L1 expression scoring, TME for IO drugs - predictive biomarker

Question: what population response to this drug best? (And Why?)

DL opportunity for H&E and IHC imaging (All AbbVie ADC)

Current practice: IHC scoring, cons is low concordance between pathologist
Using genomics data: bulk tumor, lack of spatial information

Statistical Challenges:

- fit-for-purpose
- analytical validation
- clinical validation

Diagnostic biomarker Example: IDx LLC. Diabetic retinopathy diagnostic system (sens=87.2%, sep=90.7%) for 1st such system FDA clear for 510k (doi:10.1038/s41746-018-0040-6)

Immunology: wearable device – surrogate biomarker

Question: Is the patient's symptom improved after treatment?

Wearable Actigraphy Sensor

DL: activity recognition, predict behavior frequency, etc

Surrogate biomarker/endpoints statistical criteria:

- association with the biological clinical endpoints
- Potentially early response detection

Statistical challenge: optimize the Surrogate biomarker (summary measures) for most associated with the biological clinical endpoints (outcome)

Neurosciences: Alzheimer's Disease, diagnostic biomarkers, enrichment/prognostic biomarkers for enrichment clinical trial design

Question: which group of patient with fast progression from MCI/early AD to AD?

Multimodal, multiple biomarker available: CSF, tau, amyloid beta, PET, MRI, etc
How to combine structure data + unstructured data for enrichment/prognosis biomarkers, to enroll fast progression patients

Statistical criteria for such a optimal prognostic (composite) biomarker, focus on evaluation of the point estimate (i.e., progression) in the selected population: unbiased and minimum variance vs. sample size, tied with POC of the enrichment design

Post-marketing: Safety monitoring

Question: Does our drug has significant AE vs. the competitor on the market?

NLP, AE detection from doctor notes (RWE, treatment emergent AE event rate comparison between drugs) or social media (drug related AE event early detection)

Challenge:
Statistical inference, causality

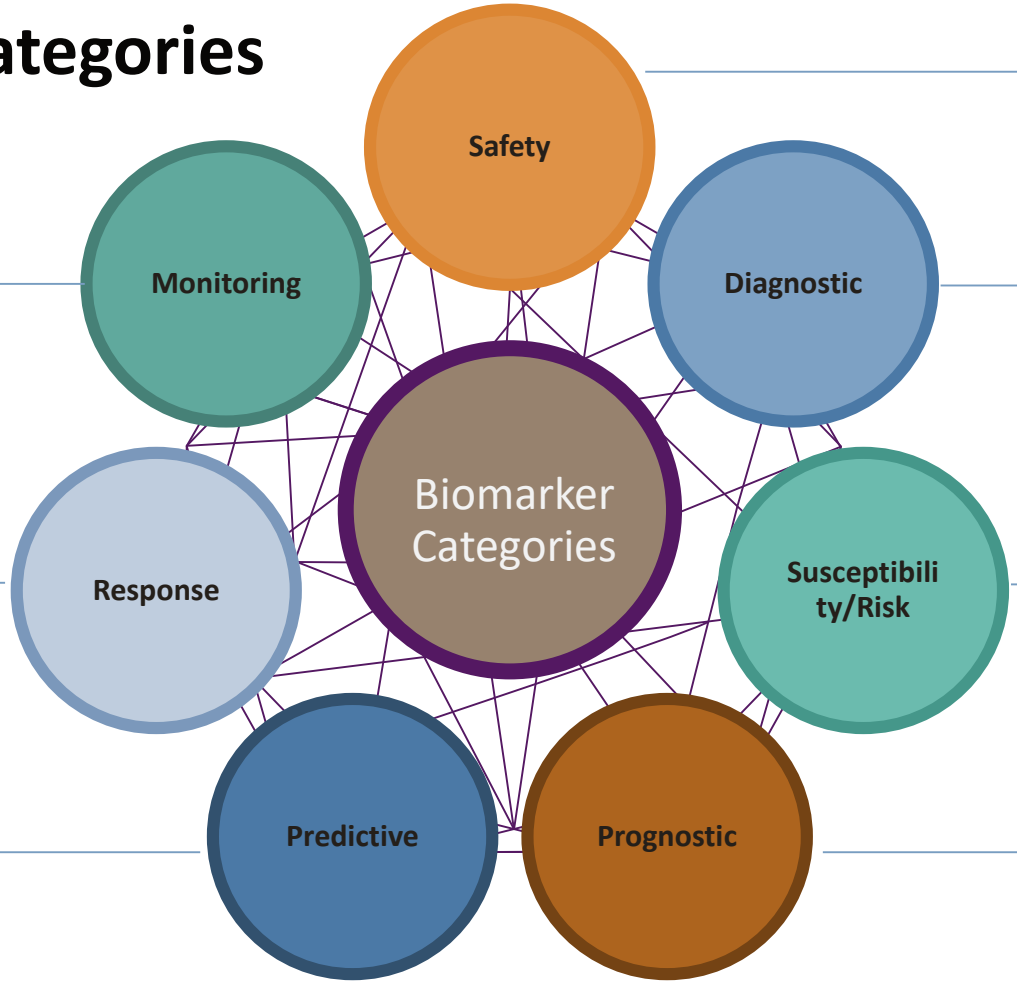
How to borrow strength of DL for “structure” data
How to combine structure and unstructured data

Option 1: Automatic feature engineering (extracting deep features from unstructured data and combine with structure data for classical ML or statistical inference)

Option 2: Convert structure data into unstructured data (need to find a way to convert your data into image/free text)


abbvie

Biomarker categories



Predict likelihood of AE

Assess status of disease/condition after exposed to drug

detect or confirm presence of disease/condition

PD marker, show biological response after exposed to drug

Potential of developing a disease/condition

Identify patients who are likely to response to the treatment

Enrich for patients with unfavorable prognosis, increase study power