# Getting Started in Computational Pathology

Tyna Hope, Ph.D., P.Eng

# Computational Pathology – a definition

- **An approach to diagnosis** that **incorporates multiple sources of raw data** (e.g., clinical electronic medical records, laboratory data including "-omics," and imaging [both radiology and pathology imaging]); **extracts** biologically and **clinically relevant information** from these data; **uses mathematic models** at the molecular, individual, and population levels **to generate diagnostic inferences** and predictions; and presents this clinically actionable knowledge to customers through dynamic and integrated reports and interfaces, **enabling physicians**, patients, laboratory personnel, and other health care system stakeholders **to make** the best possible **medical decisions**.
- More generally, using computation for the interpretation of multi-parameter data to improve health care.

Computational Pathology: A Path Ahead. Louis DN, Feldman M, Carter AB, et al. Arch Pathol Lab Med. 2015;140(1):41-50.

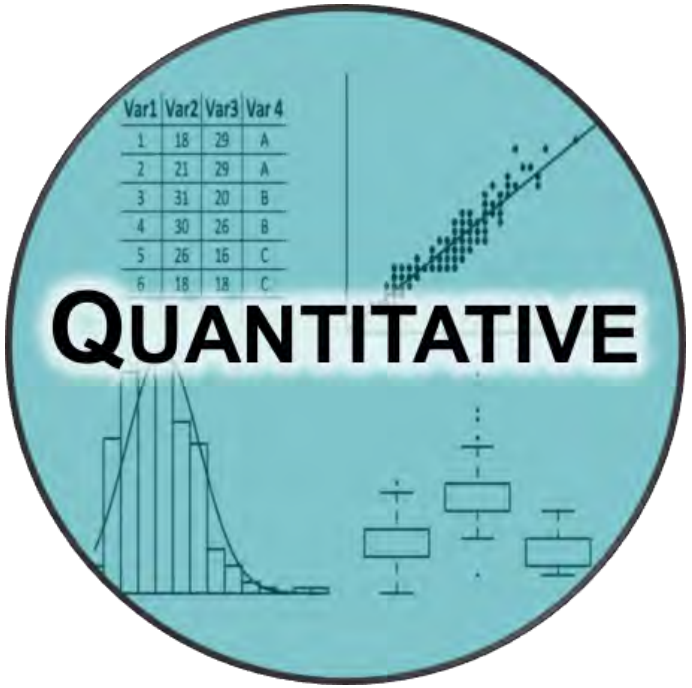
# Computational Pathology – a definition

- **An approach to diagnosis** that **incorporates multiple sources of raw data …; extracts …clinically relevant information ..; uses mathematic models …to generate diagnostic inferences …enabling physicians to make medical decisions.**

Computational Pathology: A Path Ahead. Louis DN, Feldman M, Carter AB, et al. Arch Pathol Lab Med. 2015;140(1):41-50.

# Computational Pathology – Why?

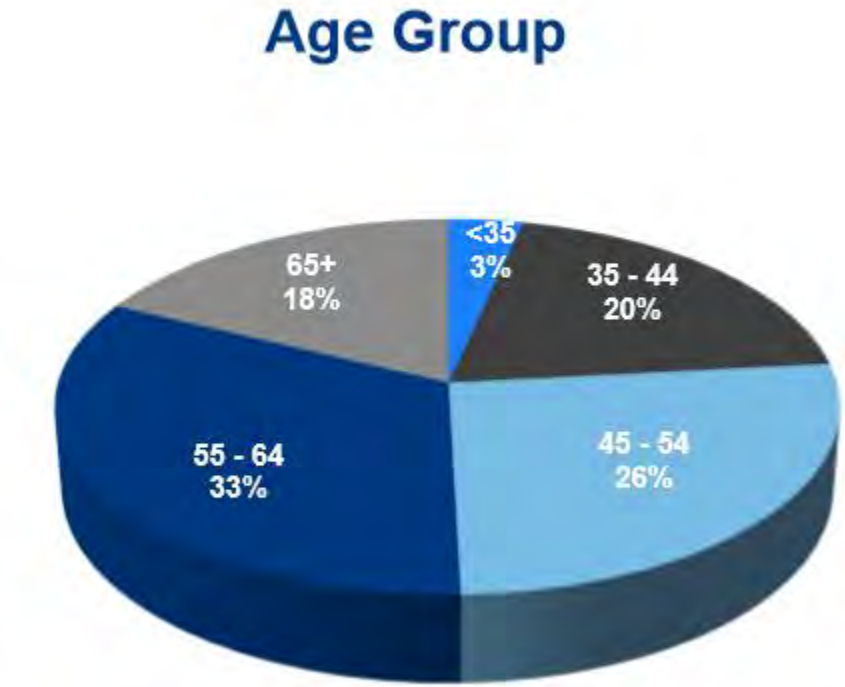
- **enabling physicians**, patients, laboratory personnel, and other health care system stakeholders **to make** the best possible **medical decisions**.
  - Reduce inter and intra observer variability
  - Quantitative rather than semi-quantitative measures



# Why?

- Efficient and effective Personalized Medicine

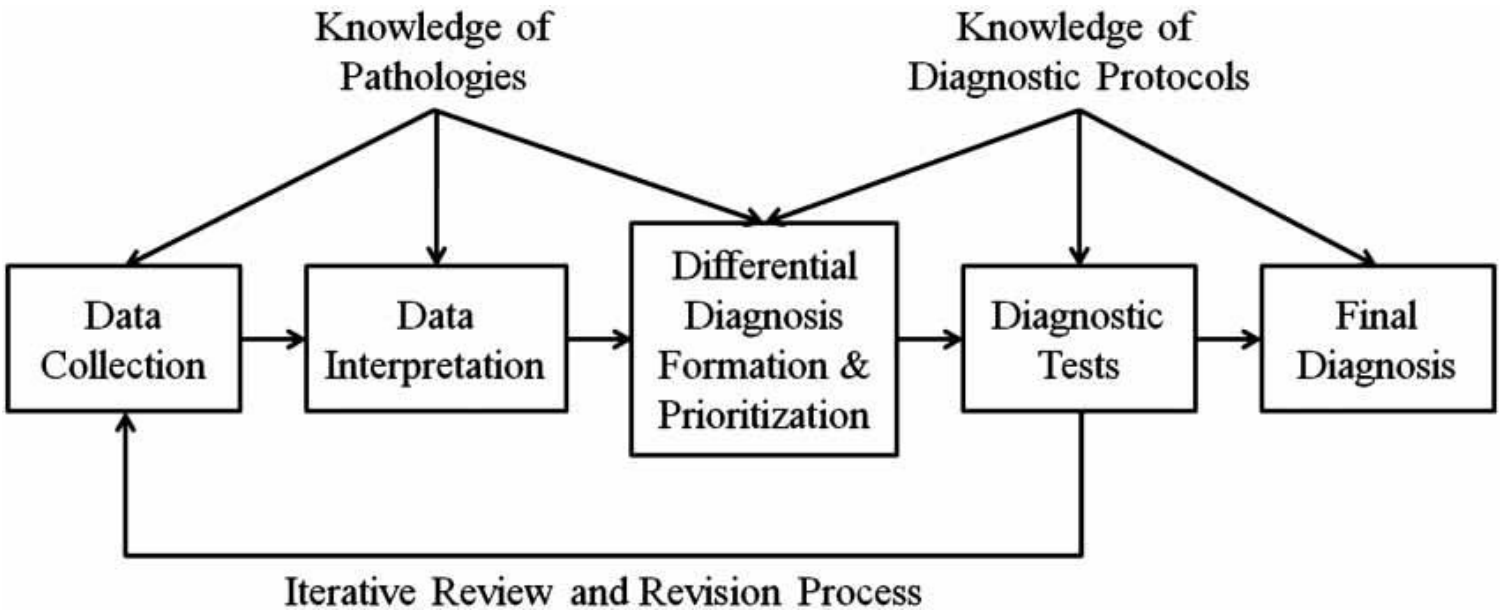





Pathologists in Canada 2018 from CMA

# Computational Pathology – How?

- **extracts** biologically and **clinically relevant information** from these data; **uses mathematic models** at the molecular, individual, and population levels **to generate diagnostic inferences** and predictions;




A Suggested Model of the Medical Diagnostic Process

https://www.researchgate.net/publication/239796277_Organizational_Diagnosis_An_Evidence-based_Approach

# Models

- Mathematical representations of data showing the relationships between inputs or between inputs and outputs
- Statistics: a branch of mathematics dealing with ***data*** collection, organization, analysis, interpretation, and presentation (Wikipedia)
- Machine Learning: ***algorithms*** and statistical models that computer systems use to perform a task without using explicit instructions, using patterns and inference instead (Wikipedia)
- To investigate further: The Close Relationship between applied Statistics and Machine Learning (blog: machinelearningmastery.com)

# Statistics vs Machine Learning

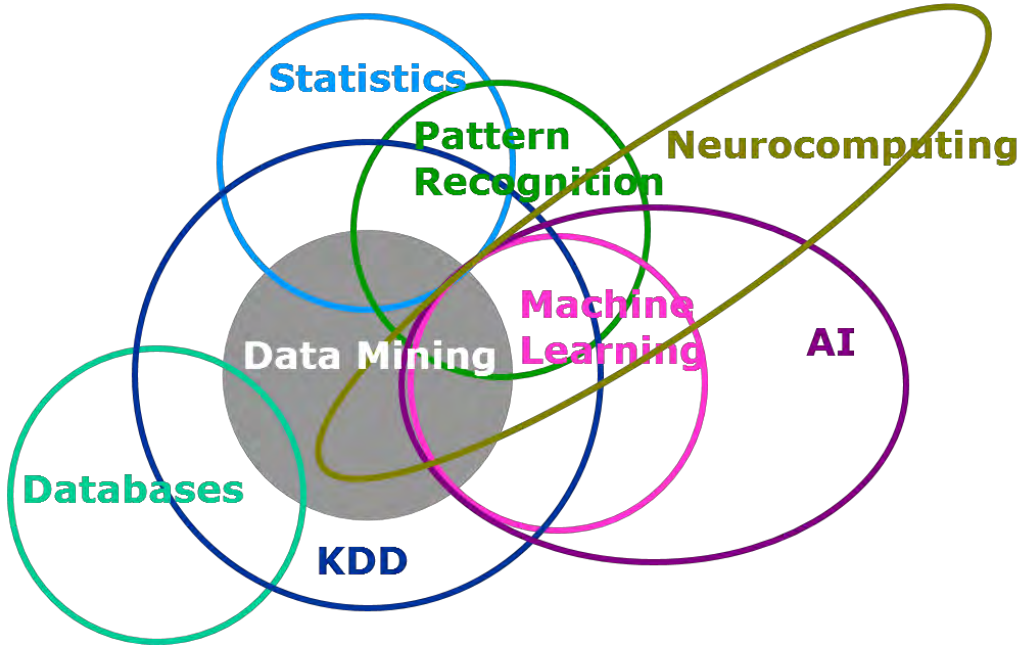
Statistics draws population inferences from a sample, and machine learning finds generalizable predictive patterns.

- Inference creates a mathematical model of the data-generation process to formalize understanding
- Prediction aims at forecasting unobserved outcomes or future behavior

Statistics versus machine learning, Danilo Bzdok, Naomi Altman & Martin Krzywinski, Nature Methods volume 15, pages 233–234 (2018)

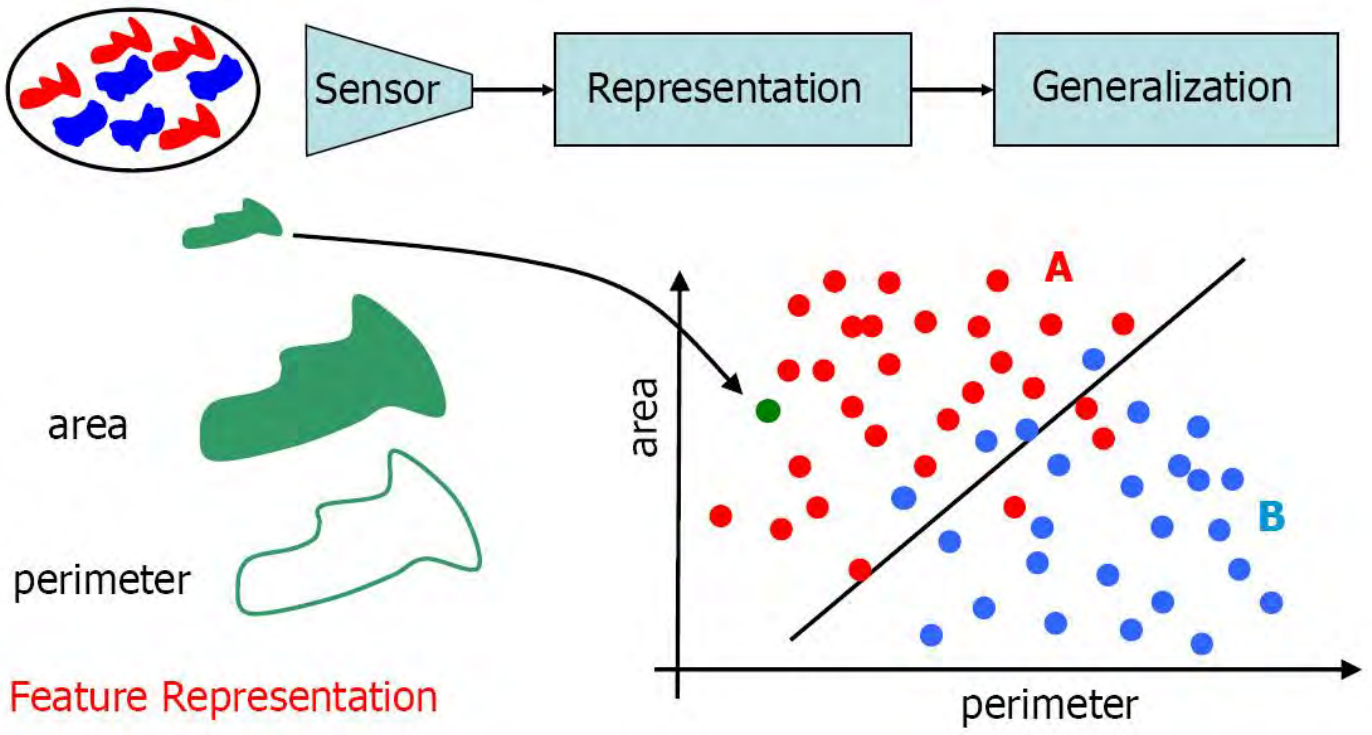
# Statistics and Machine Learning

- There is a strong argument for using both
- Structured data (tables)
- Unstructured data ( images, speech)

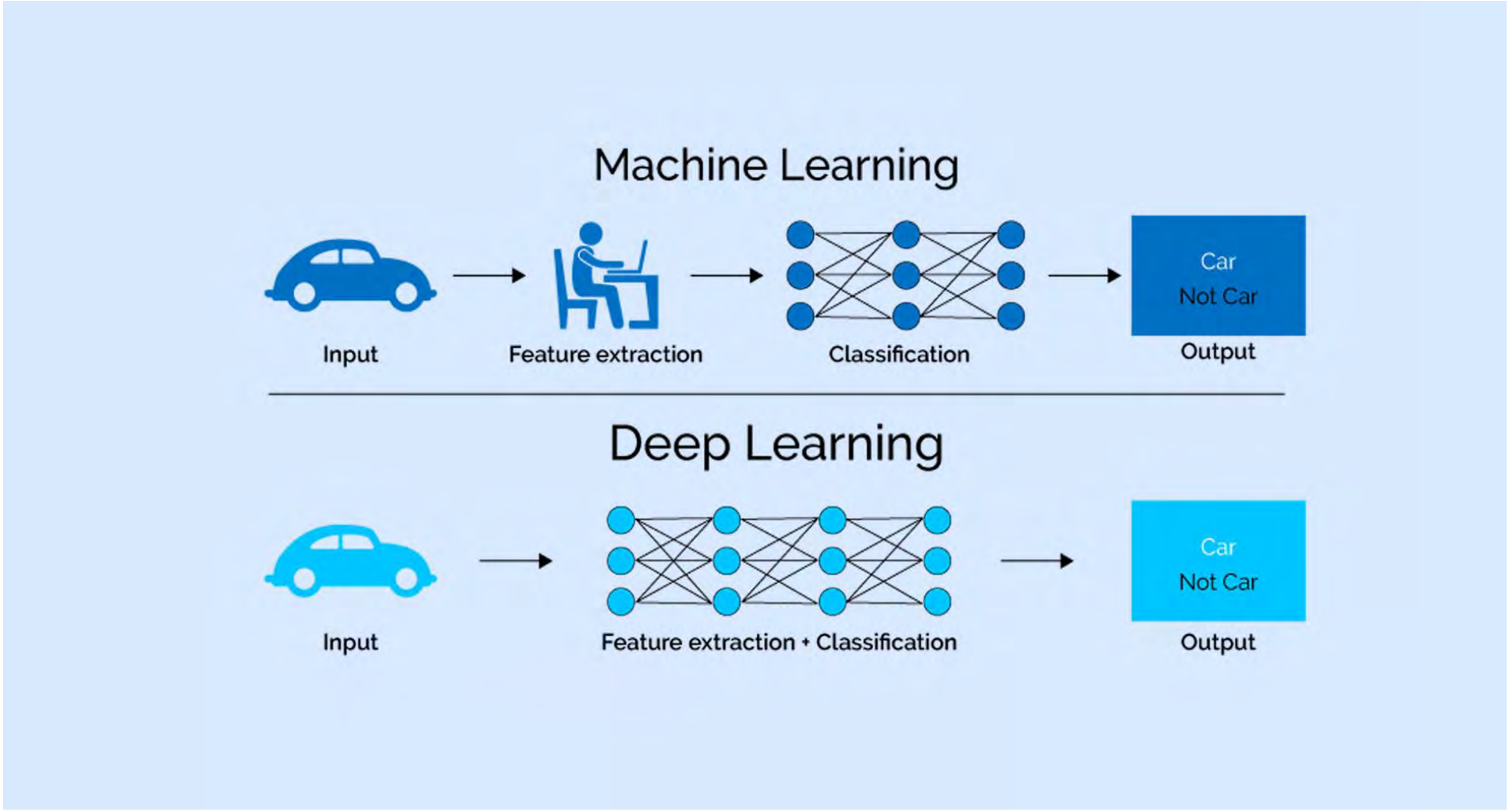



https://towardsdatascience.com/exploring-the-meaning-of-ai-data-science-and-machine-learning-with-the-latest-wikipedia-5fea5f0a2d46

# Classical ML Applied to Structured Data



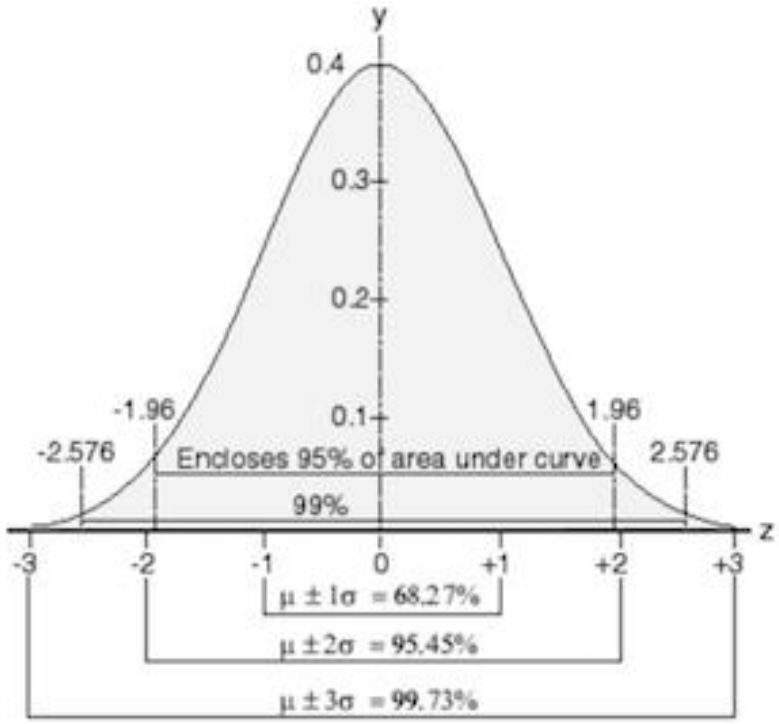
# Deep Learning Applied to Unstructured Data




http://didarc.com/en/news/applying-deep-learning-to-facial-recognition-cameras

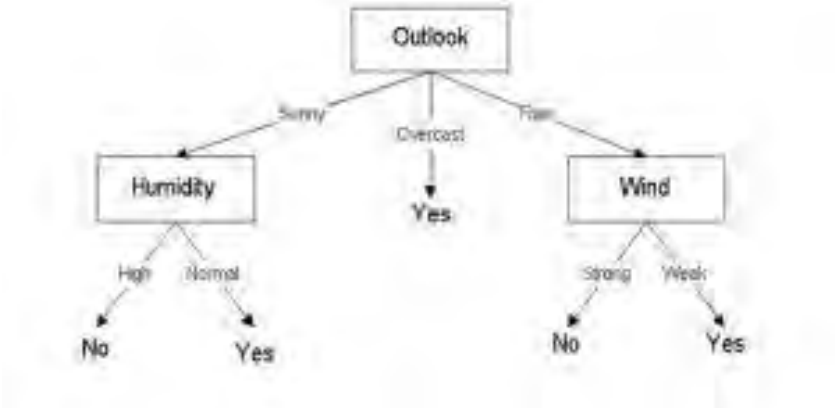
# Making Models

- Statistical methods with structured data
  - Survival model
  - Hypothesis testing
  - Summary statistics
  - Bayes model
  - Power calculations
  - Linear regression

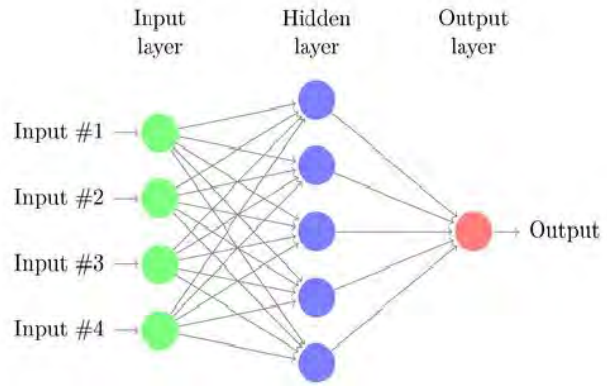


# Making Models

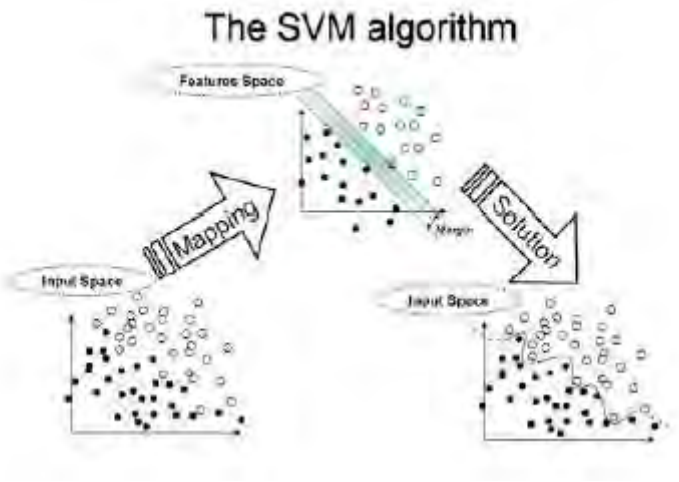
- Statistical methods & classical ML with structured data
  - Support vector machines
  - Decision trees
  - Ensemble methods
  - Logistic regression
  - Multilayer perceptron




https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html



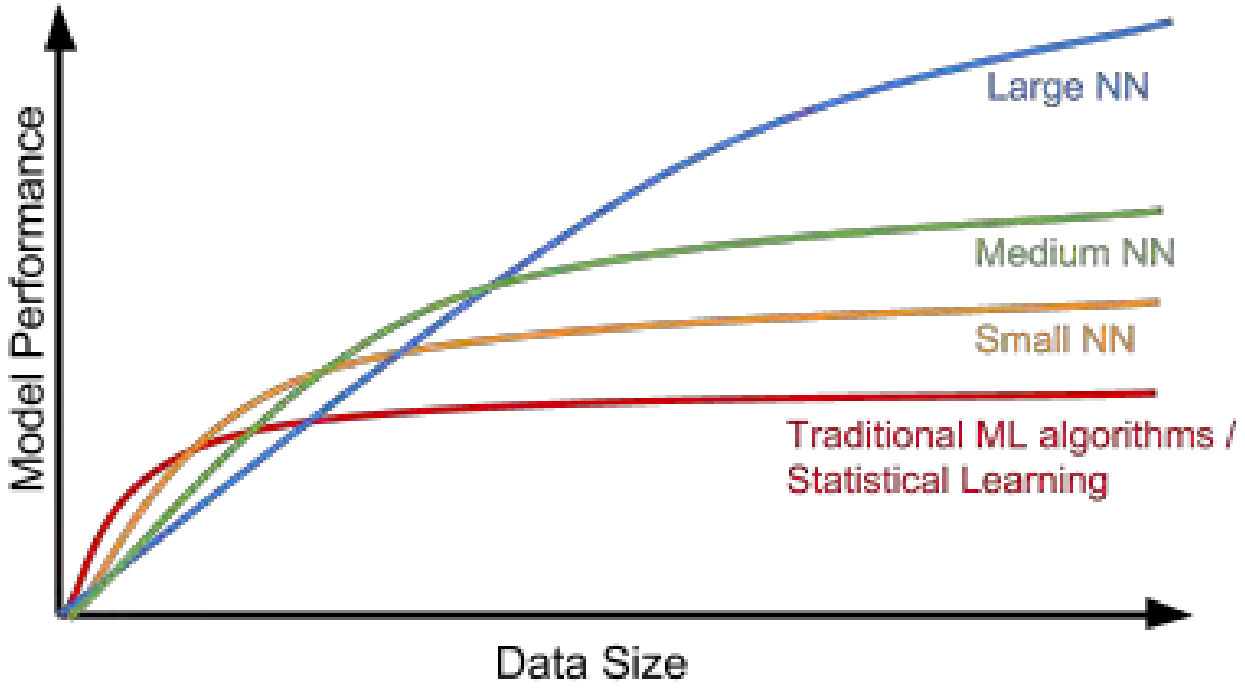

https://www.researchgate.net/figure/Graphical-representation-of-an-MLP-or-multi-layer-NN_fig3_279973874




https://www.datasciencecentral.com/profiles/blogs/collection-of-svm-libraries-by-language

# Making Models-Deep Learning

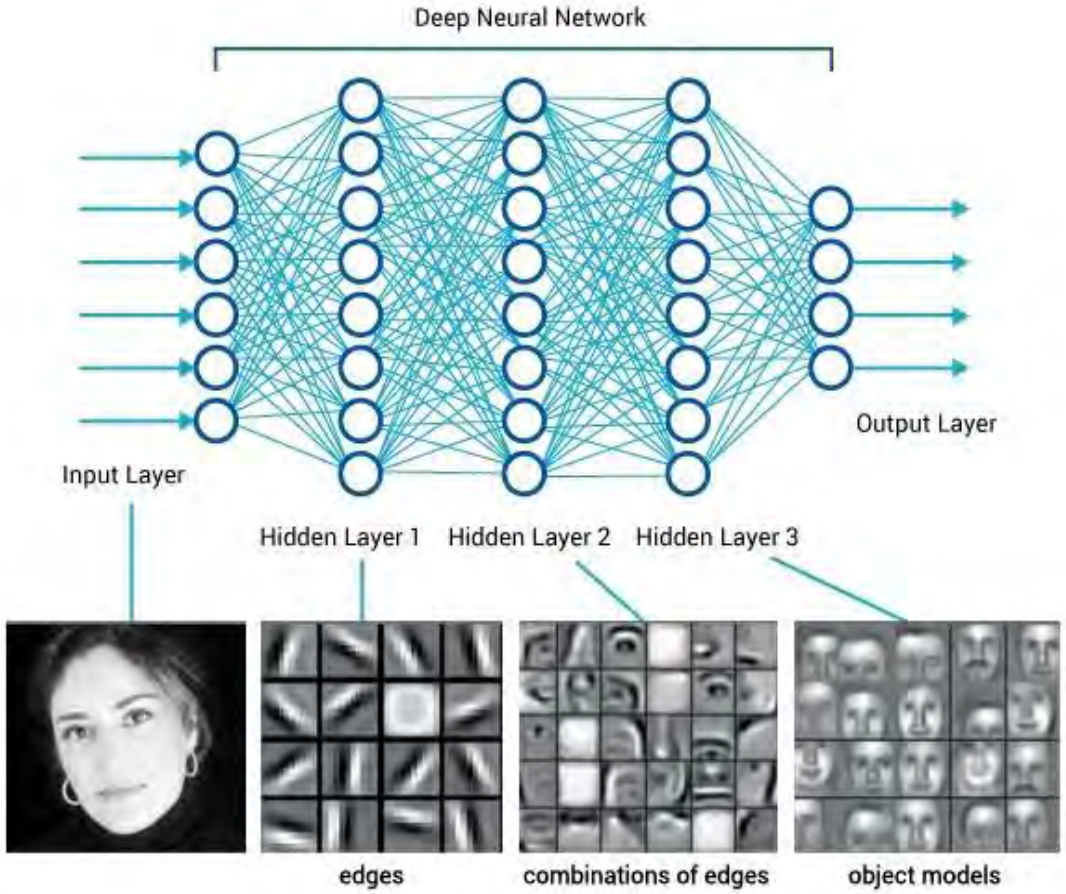
- Can be applied to structured and unstructured data
- The biggest benefit has been seen in the application to unstructured data




Andrew Ng – Coursera Deep Learning Specialization

# Deep Learning with Unstructured Data

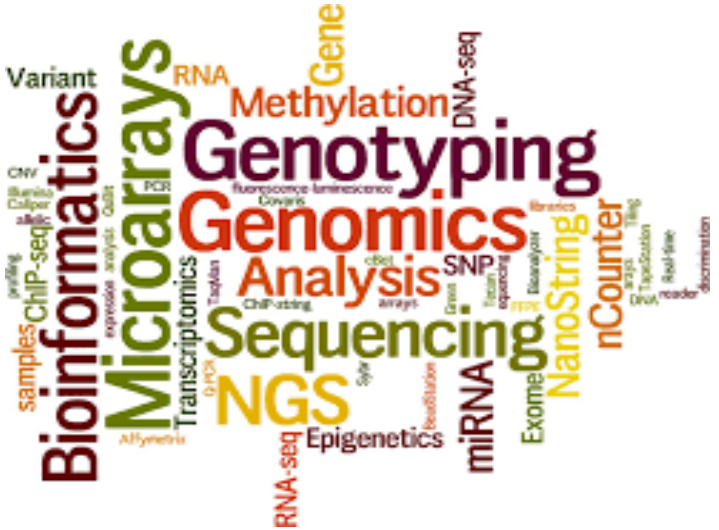
- **Input data** (aka features) are **no longer hand crafted** from complex signal spaces such as images.
- NN are very deep & during learning interesting aspects are discovered.
- Moved the effort from crafting of features to data & the design of NN




https://www.eqtr.com/ideas/posts/2018/06/embedding-layers-in-deep-learning/

# Computational Pathology – the data

- **An approach to diagnosis that incorporates multiple sources of raw data (e.g., clinical electronic medical records, laboratory data including "-omics," and imaging [both radiology and pathology imaging]); …**

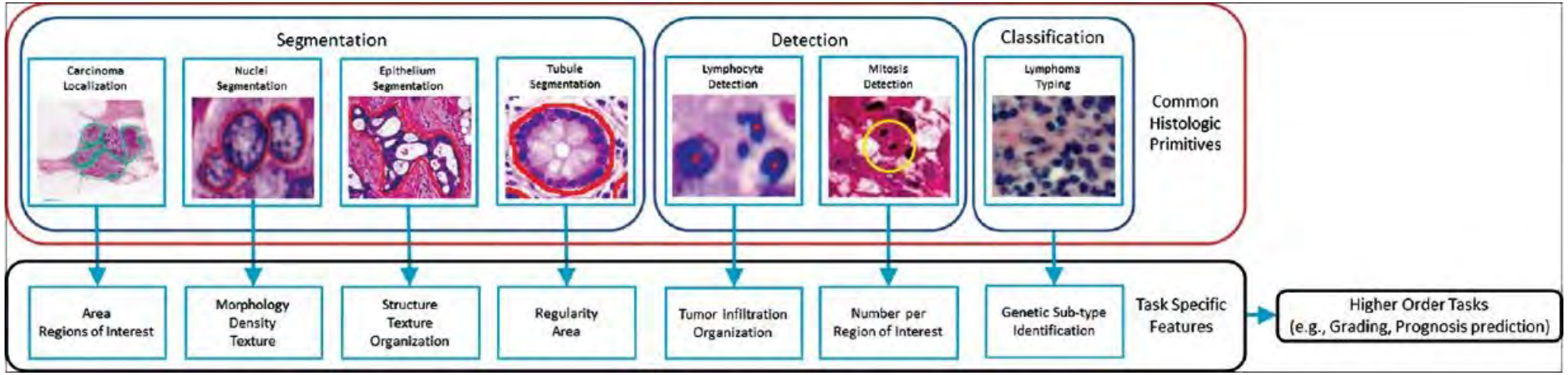


# Digital Pathology as a component

- There is a wealth of information on the slide
- Software tools make the extraction of new information tractable, routine information faster and less variable

# Beyond Image Sharing

- Many pathology departments already scan slides for education & consult
- What else can be done with the images?
  - Segment
  - Detect
  - Quantify
  - Classify
  - Reconstruct
  - Map
  - Predict

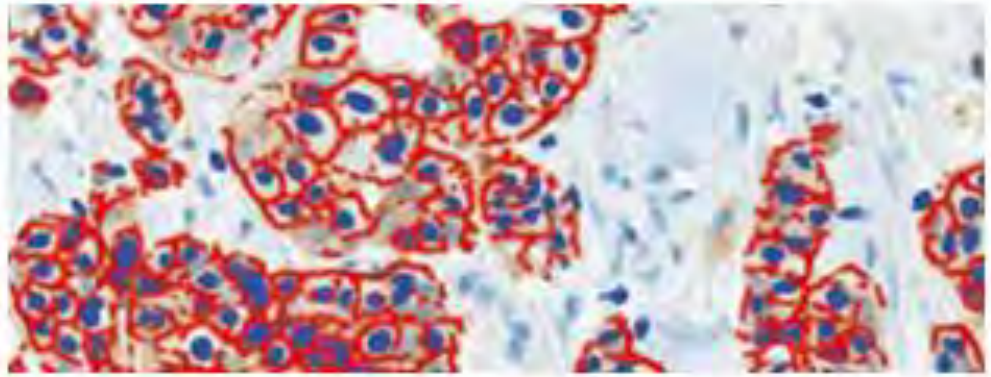



Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J Pathol Inform [serial online] 2016

# Clinical Tools

Image analysis  software is intended to provide quantified or semi-quantified information to provide repeatable data to identify patients who are most likely to respond to a treatment option

Some examples are Ventana's Er, Pr, Her2, p53, Ki67 image analysis software



The Companion Algorithm HER2 (4B5) image analysis software

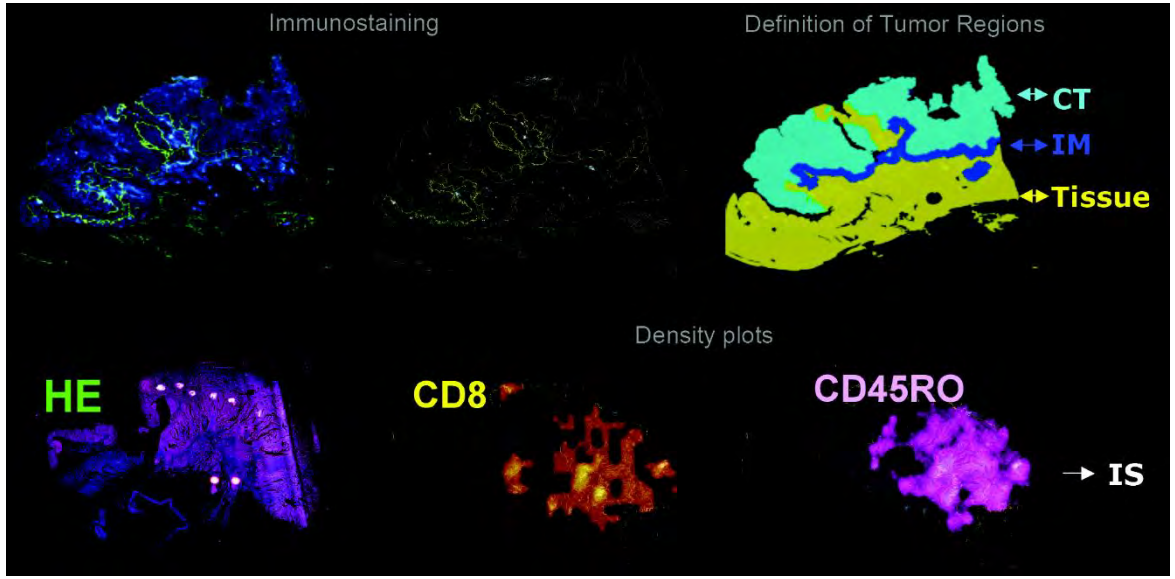
*The PATHWAY HER2 (4B5) assay is FDA approved.

# Clinical Tools

- Challenges around digital pathology in some markets
  - FDA gave approval to use Phillips Imaging system for diagnosis in 2017 https://www.archivesofpathology.org/doi/pdf/10.5858/arpa.2017-0496-CP
  - While DPS use in Canada has largely been at the discretion of pathologists, HC approval has been in place for Omnyx and Leica since 2013 https://www.archivesofpathology.org/doi/pdf/10.5858/arpa.2013-0289-ED
- With digital pathology not available for use in the large US market until recently, there had been reduced incentive for quantitative analysis tools suitable for the clinic

# Out-Sourced Tests

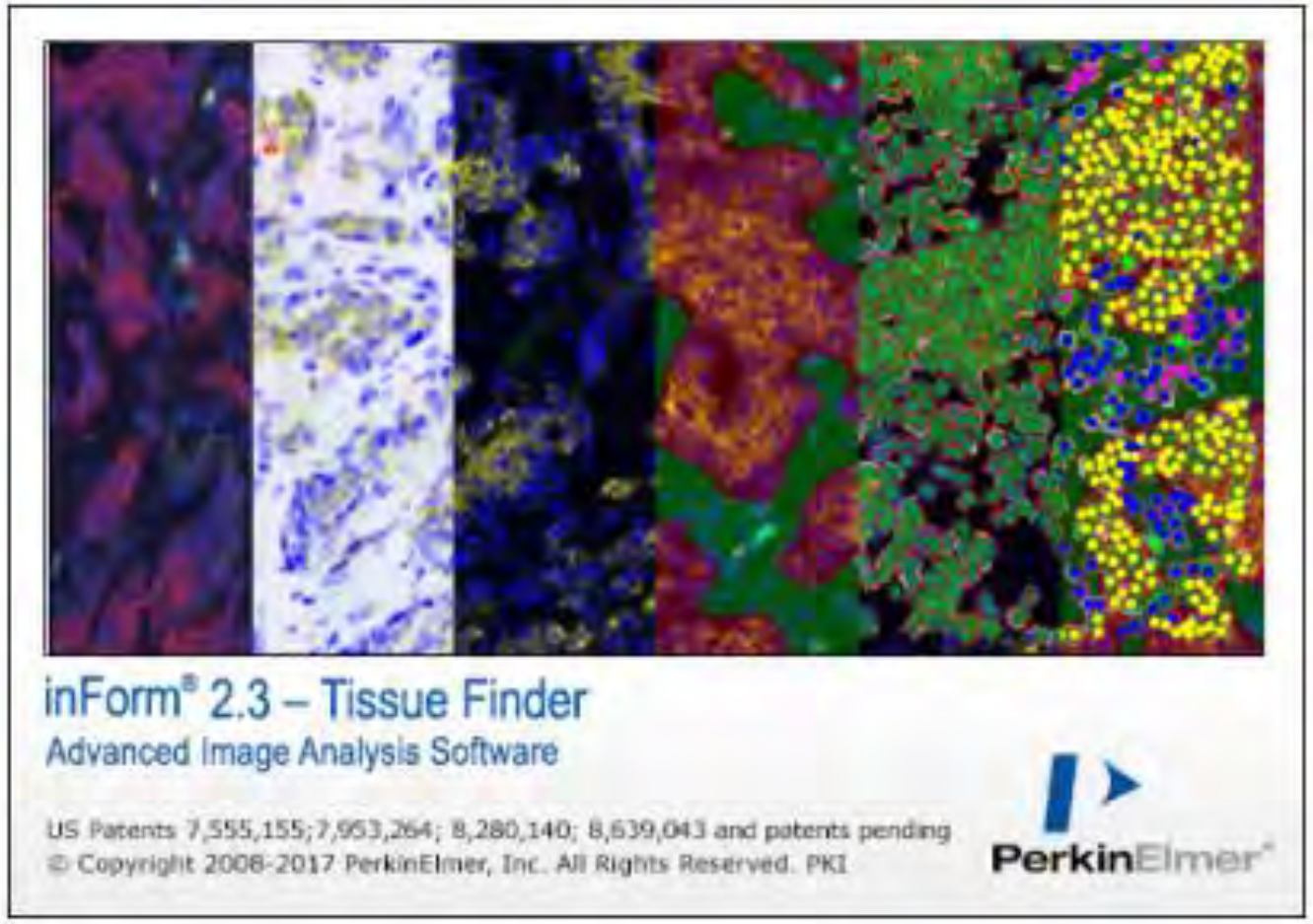
- Haliodx (commercially available predictive model)
  - Immunoscore is an in vitro diagnostic test predicting the risk of relapse in early stage colon cancer patients, by measuring the host immune response at the tumor site using image analysis. (CD3+, CD8+ on consecutive slides)



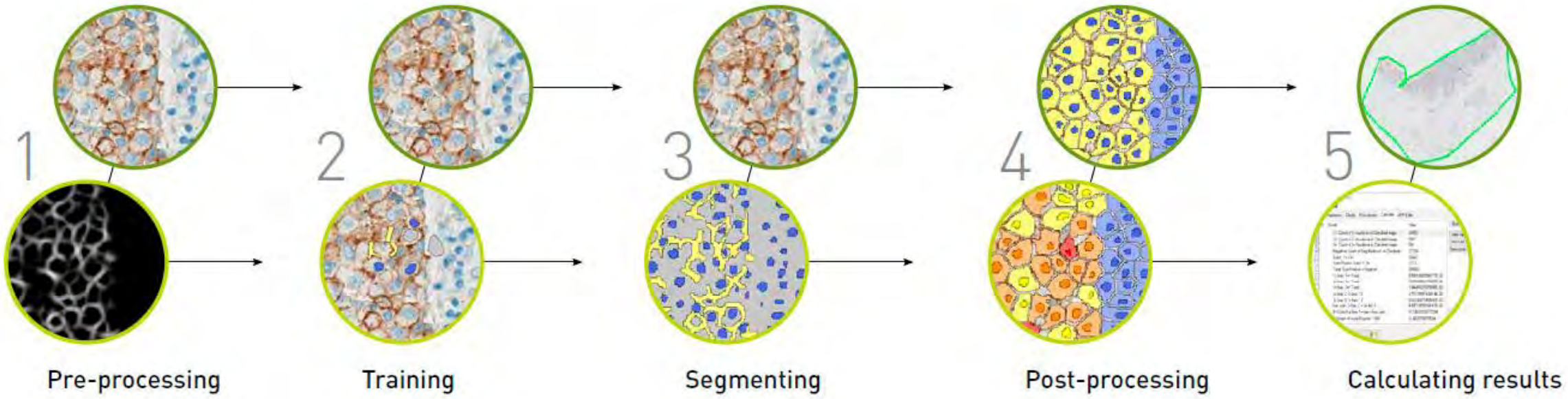

http://www.immunoscore.org/research.shtml

# Research Tools

- There are many platforms that allow the researcher to segment images and quantify results
  - inForm
  - Tissue Mark
  - Visiopharm
  - Definiens
  - Indica Labs



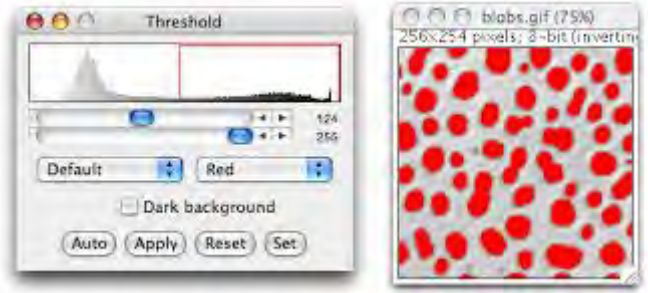
# Visiopharm – a Typical Workflow

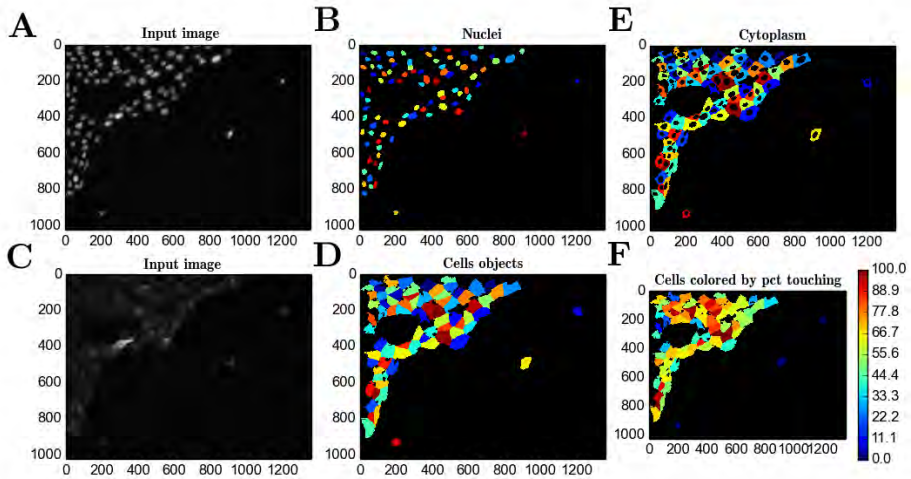


# DIY*ish*

Commercial Image Analysis platforms may not suit your needs or may be too expensive.

- Some “bioimage informatics” platforms
  - ImageJ (NIH)
  - Cell Profiler (Broad Institute)
  - Cran R (R Foundation)
  - icy (France-BioImaging)



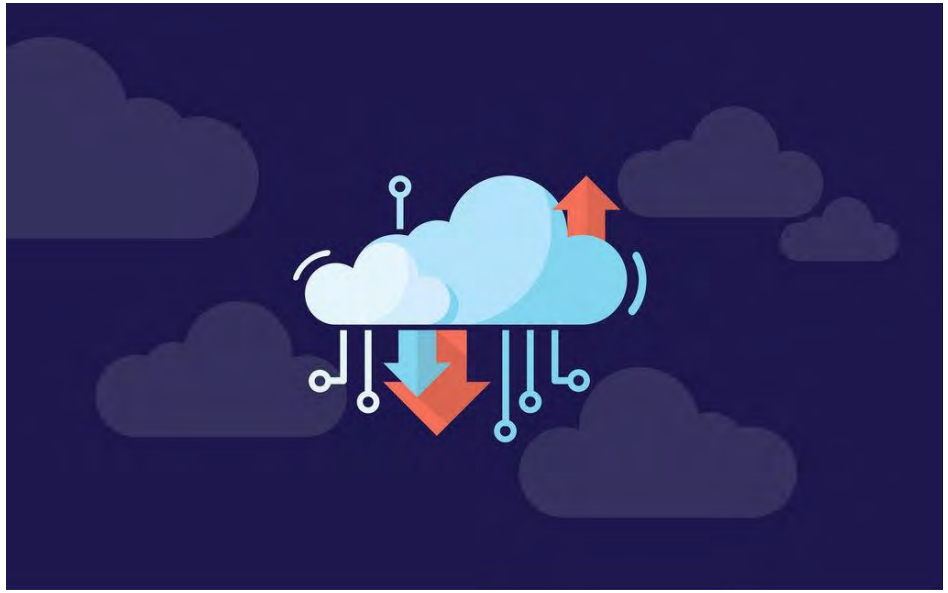

You may (will) need to do some tweaking to get what you need out of them

# DIY- Code

- For the code comfortable or those who wish to partner with engineering/ comp. sci./ statisticians
  - python libraries
    - Scikit-learn
    - Pandas
    - Keras, tensorflow
    - XGBoost…
    - opencv
  - Cran R – has thousands of packages
  - OpenSlide- to be able to deal with the large images


# DIY – Computational resources

- This work takes computing power
- Desktop - often good enough for proof of concept
  - May take hours to process a single image – may need to tile images
  - Often a hierarchical approach helps
  - GPUs
- Multiple Computers
  - Cloud services (AWS, Google Cloud)



https://hackernoon.com/trends-and-challenges-in-cloud-computing-with-deep-learning-33c23e9201a9

# Some Additional Resources

- http://tissuepathology.com
- https://scikit-learn.org/stable/index.html
- https://cran.r-project.org/web/views/MachineLearning.html
- https://thepathologist.com/inside-the-lab/the-promise-of-computational-pathology-part-ii
- http://www.jpathinformatics.org/
- tyna.hope@gmail.com