Image Analysis for Pathologists: Basic Information Needed to Make Sense of the Machine Learning/Artificial IntelligenceRevolution

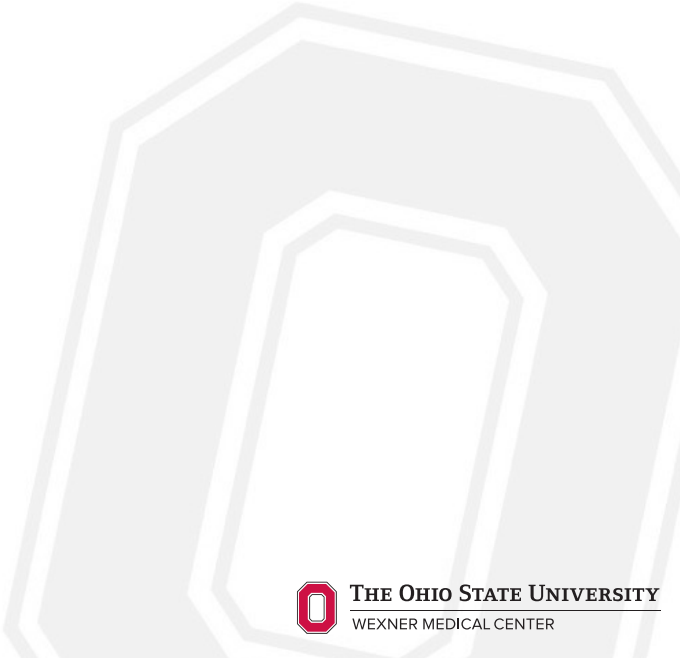
José Javier Otero, M.D., Ph.D

Director of Neuropathology

The Ohio State University

Disclaimers

I have no conflicts of interest and this work was not sponsored by any private companies.



Future Concepts in Neuropathology

Ohio Serpent Mound (Adena Culture)

Ohio, est 1000 BCE- 500 AD

Amphitheatre of El Jem

Tunisia, est ~238 AD

Architecture is not advancing.
Neuropathology is Advancing!!

Ohio, est 1000 BCE- 500 AD

Tunisia, est ~238 AD

Ohio today is ~ Tunisia in 238 AD

Pathology is Changing Culturally and Philosophically

- **Old Joke:** Why did the pathologist cross the road?
- To get to the other slide!
- **New Joke:** What do you get when you cross a mosquito to a rock climber?
- Nothing! You can't cross a vector with a scaler!

Philosophical Shift: Movement away from artistic interpretation of histological patient data to objective, quantifiable data for clinical decision-making.

- ***Modern Pathology has, often without realizing it, embraced applied mathematics in the form of biostatistics and bioinformatics.***

The Future is Now

"As traditional morphologic techniques reach their inevitable limits, incorporation of new methods permits expansion of the diagnostic frontier…

Cyto- and molecular genetics will continue to contribute not only valuable insights about astrocytic neoplastic transformation and malignant progression, but also novel diagnostic approaches.

PCR strategies, ***combined with morphologic techniques*** and image processing, should better detect tumors at risk for progression.

Scott Vandenberg, Current Diagnostic Concepts of Astrocytic Tumors, JNEN, 1992, VOL 51, 6, pp664-657 1992.

Pre-requisite for Digital Pathology

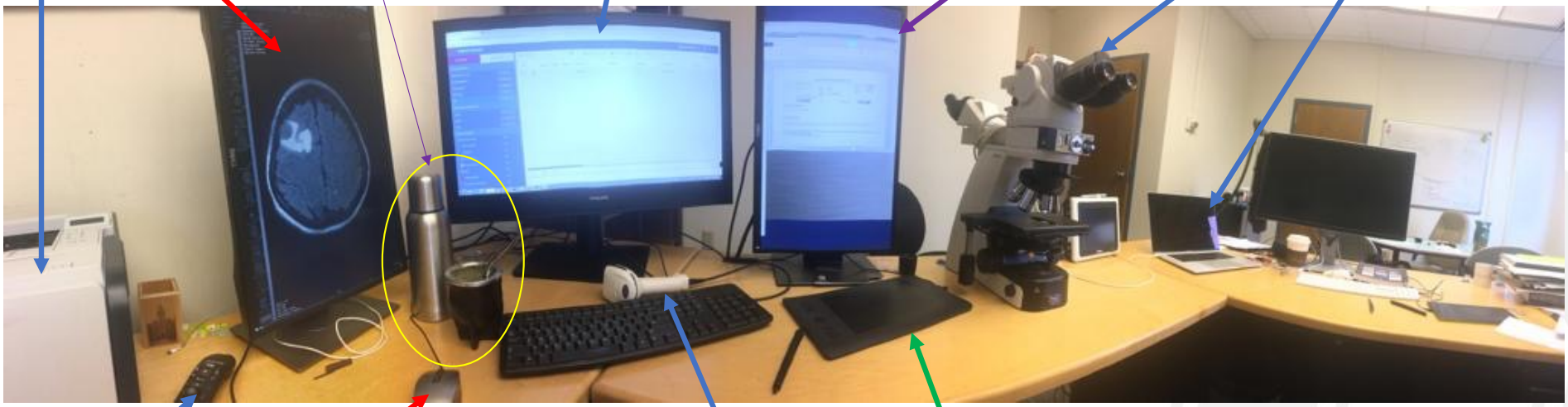
EXCELLENCE IN HISTOLOGY!!!

https://www.captodayonline.com/histology-lab-tips-for-top-tier-whole-slide-images/

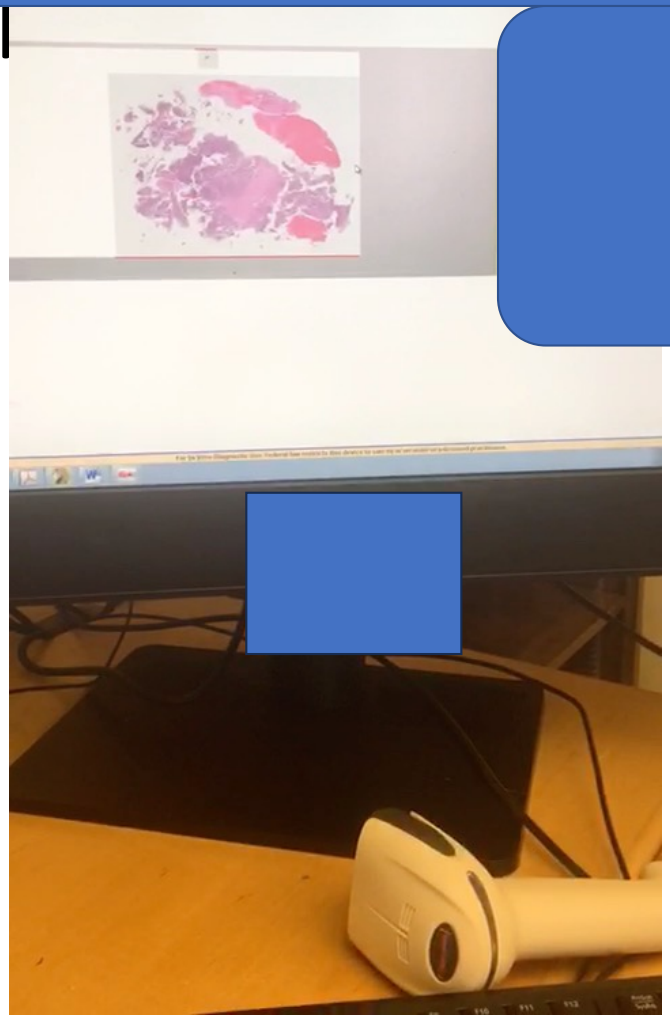
-Problems: folds, stuffing slides resulting in massive scan times, tissue on edge resulting in out of focus, dust on slides, drying of slides...

-***The more things change the more they stay the same.***

Ergonomics for digital



Typical work-speed of digital connection, standard case



Benefits:

- Easily open WSI and Imaging Data

Problems:

- .
- More screen time can make you loose concentration. I recommend closing email applications or other distractors.
- Results in sign out with resident with your “back to them” instead of face to face contact.

Neuropathology Slide Quality

NEUROPATH SLIDE QUALITY

MONTH	Cases for NEURO	Cases Scanned for NEURO	Estimated Slides Scanned for NEURO[1]	Cases "Defer to glass"[2]	Slides "Defer to glass"[2]	Cases w Slide Rescan Requests[3]	Slides w Slide Rescan Requests[3]	% of Cases Containing ANY Slides Unfit for Diagnosis	% of Slides Unfit for Diagnosis
Aug-19	117	103	1,074	4	4	1	1	4.9%	0.5%
Sep-19	112	108	1,126	3	3	2	2	4.6%	0.4%
Oct-19	146	146	1,523	3	3	0	0	2.1%	0.2%
Nov-19	122	122	1,272	9	9	3	4	9.8%	1.0%
Dec-19	117	117	1,220	3	3	4	18	6.0%	1.7%
Jan-20	107	107	1,116	0	0	2	2	1.9%	0.2%
TOTAL	721	703	7332	22	22	12	27	4.8%	0.7%

1= Number of cases times average number of slides scanned per case at DPSC since May 2017 (10.43 slides)

2= Scan Center has detected an unscannable slide and alerted the pathologist

3= Pathologists detects slide unfit for diagnosis, requests rescan

Spike in failed slides in Dec-19 reflects a single case in which 15 slides from the same case were affected by a machine error and had to be rescanned.

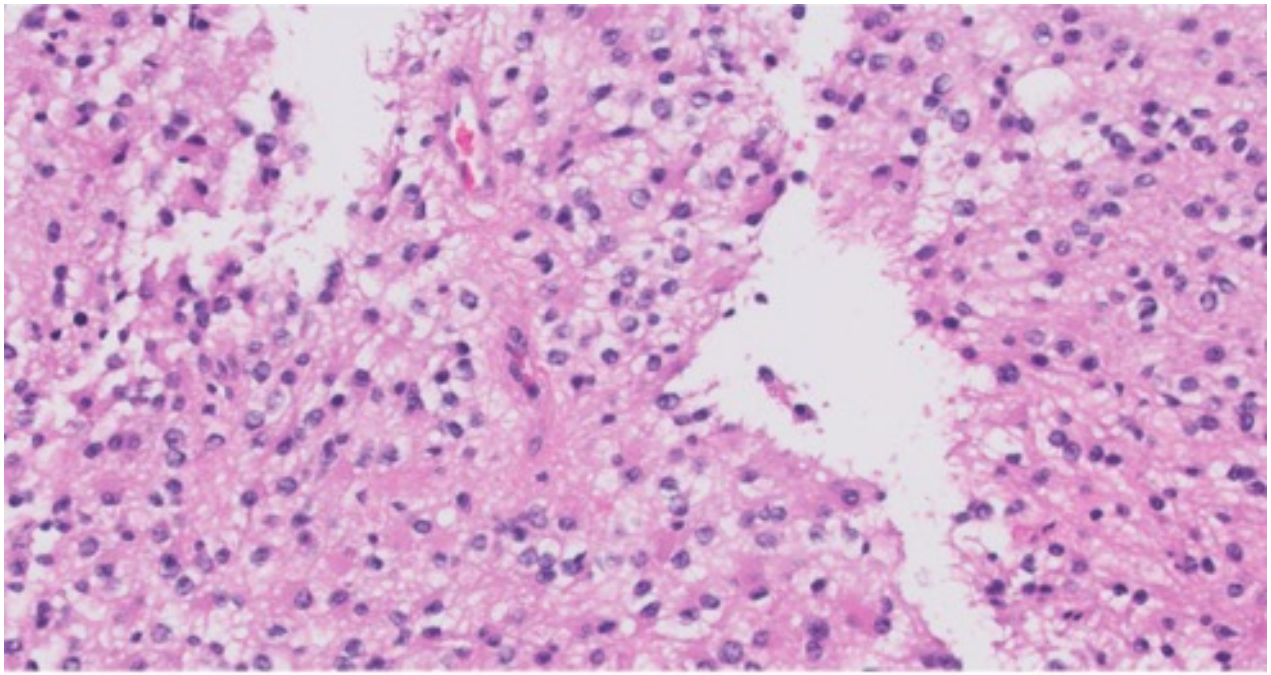
Causes of Slide Scan Failure

CAUSES OF NEUROPATH SLIDE SCAN FAILURE

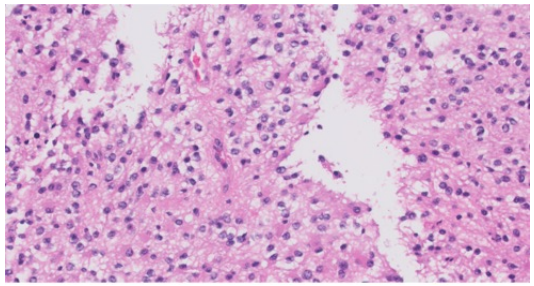
DATE	PATHOL	CASE	SLIDE	STAIN	PROBLEM	CAUSE
8-Aug-19	OTERO	S19-053017	A5-1	HE	ROI	random scanner error
3-Sep-19	OTERO	S19-056245	B2-27	HE	ROI	no detectable tissue
24-Sep-19	OTERO	S19-062785	B1-13	RETIC	OOF	histology/ slide prep
1-Nov-19	THOMAS	S19-071947	A3-3	KI67	OOF	random scanner error
1-Nov-19	THOMAS	S19-071947	A3-7	PDL1	OOF	random scanner error
13-Nov-19	OTERO	S19-073960	B1-5	IDH	OOF	faint
18-Nov-19	OTERO	S19-075170	B2-3	RETIC	OOF	histology/ slide prep
11-Dec-19	OTERO	S19-080136	A3-19	KI67	OOF	histology/ slide prep
17-Dec-19	OTERO	S19-082124	A1-A15	HE	VB error	contaminated scanner optics
19-Dec-19	OTERO	S19-082124	A1-5	NEUROf	OOF	random scanner error
30-Dec-19	OTERO	S19-083937	A1-6	GFAP	ROI	tiny tissue
6-Jan-20	OTERO	S19-083337	A3	KI67	OOF	histology/ slide prep
6-Jan-20	OTERO	S19-084993	A1-6	SYNAPTO	OOF	faint

Tissue type doesn't seem to be much of a factor. Very small biopsies may pose issues but not predictably. Faint IHC stains either negative or weakly positive seem to be most predictive of scan failure but not very.

Traditional View of Pathology (artistic interpretation of images)



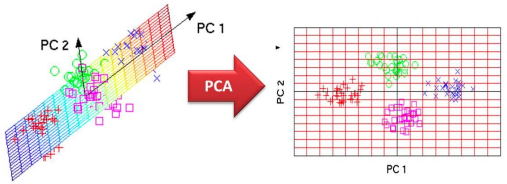
Modern View of Pathology-Big data field

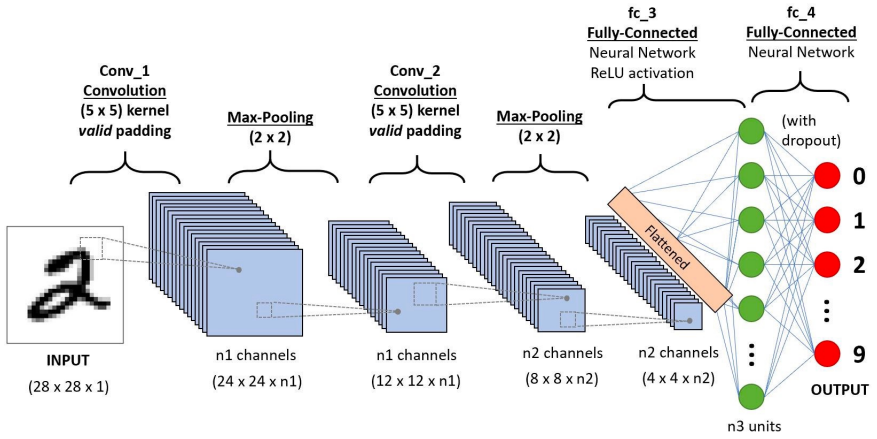


Conditional Probability: Bayes' Theorem

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Dimensionality Reduction
&
Principal Component Analysis





Data types and how to handle them.

Structured Data

- Easiest data to interact with.
- Typically contains observations as rows, and features of each observation as columns, such as an excel file.
- Non-deep learning approaches are often better than deep learning approaches.
  - *Very useful when mechanism is more important over accuracy.*
  - Eg., Telling a patient, "For every kg of weight loss you have, your probability of transitioning from pre-DM to DM reduces by x percent."

Unstructured data

- Data that is not easily condensable into a data table.
- Eg., images
- Typically handled in two manners:
  - Extract features and convert the unstructured data to structured data. (eg., obtain a ki67 proliferation index, or quantify Her2/neu into an ordinal scale of 1+-3+.
  - Utilize deep learning methodologies.
    - Need to be done *if and only if* there is some form of normalization of the data with standardized SOP's, and large quantities of data.
    - *Very useful when accuracy is MORE important than mechanism.*

Data types and how to handle them.

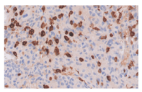
Structured data

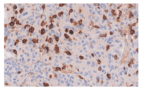
f(x) = some number
(regression)

f(x) = some probability of pertaining to a class

(label prediction, aka discriminative model)

Unstructured data

f( **) = some number**
(regression)

f( **) = some probability of pertaining to a class**

(label prediction, aka discriminative model)

Data types and how to handle them.

Structured Data

- Easiest data to interact with.
- Typically contains observations as rows, and features of each observation as columns, such as an excel file.
- Non-deep learning approaches are often better than deep learning approaches.
  - *Very useful when mechanism is more important over accuracy.*
  - Eg., Telling a patient, "For every kg of weight loss you have, your probability of having DM reduces by x percent."

Unstructured data

- Data that is not easily condensable into a data table.
- Eg., images
- Typically handled in two manners:
  - **Extract features and convert the unstructured data to structured data. (eg., obtain a ki67 proliferation index, or quantify Her2/neu into an ordinal scale of 1+-3+.**
  - Utilize deep learning methodologies.
    - Need to be done *if and only if* there is some form of normalization of the data with standardized SOP's, and large quantities of data.
    - *Very useful when accuracy is MORE important than mechanism.*

Pathology is undergoing a transformation that will change how we practice clinically

- Similar to the genetics transformation of the 1990's and 2000's, machine learning/artificial intelligence is a technology disrupting event.
- Pathologists need to learn the basics of ML/AI as it relates to their clinical practice.
- Pathologists are optimally situated to contribute to this exciting revolution in anatomic pathology by providing the UNIQUE DOMAIN KNOWLEDGE and VISUAL LITERACY necessary to apply these tools.

Differences between Humans and Computers

Human

- Image
- Colors in the visual spectrum
- Magnification (objective size, eg., 10x, 40x, 63x)

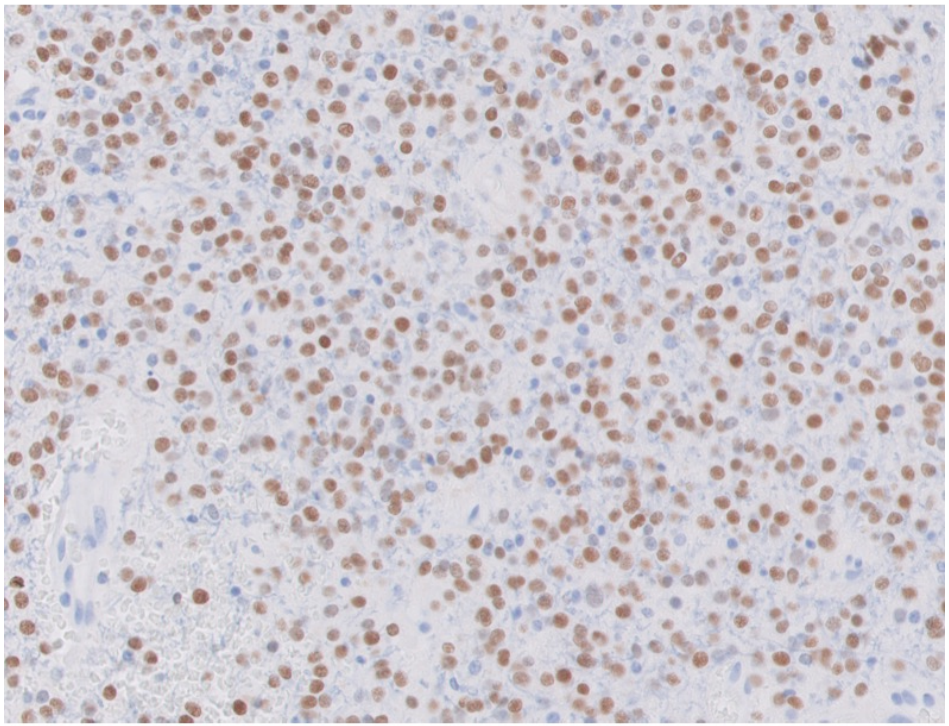
- **Humans can be trained to learn visual information by comparing new information with a set of historical references.**

Computer

- Spreadsheet, matrix, dataframe
- Numbers (0-255 or 0-1)
- Resolution (number of picture elements, or pixels, in a specified volume. Eg., 100 pixels/μm)

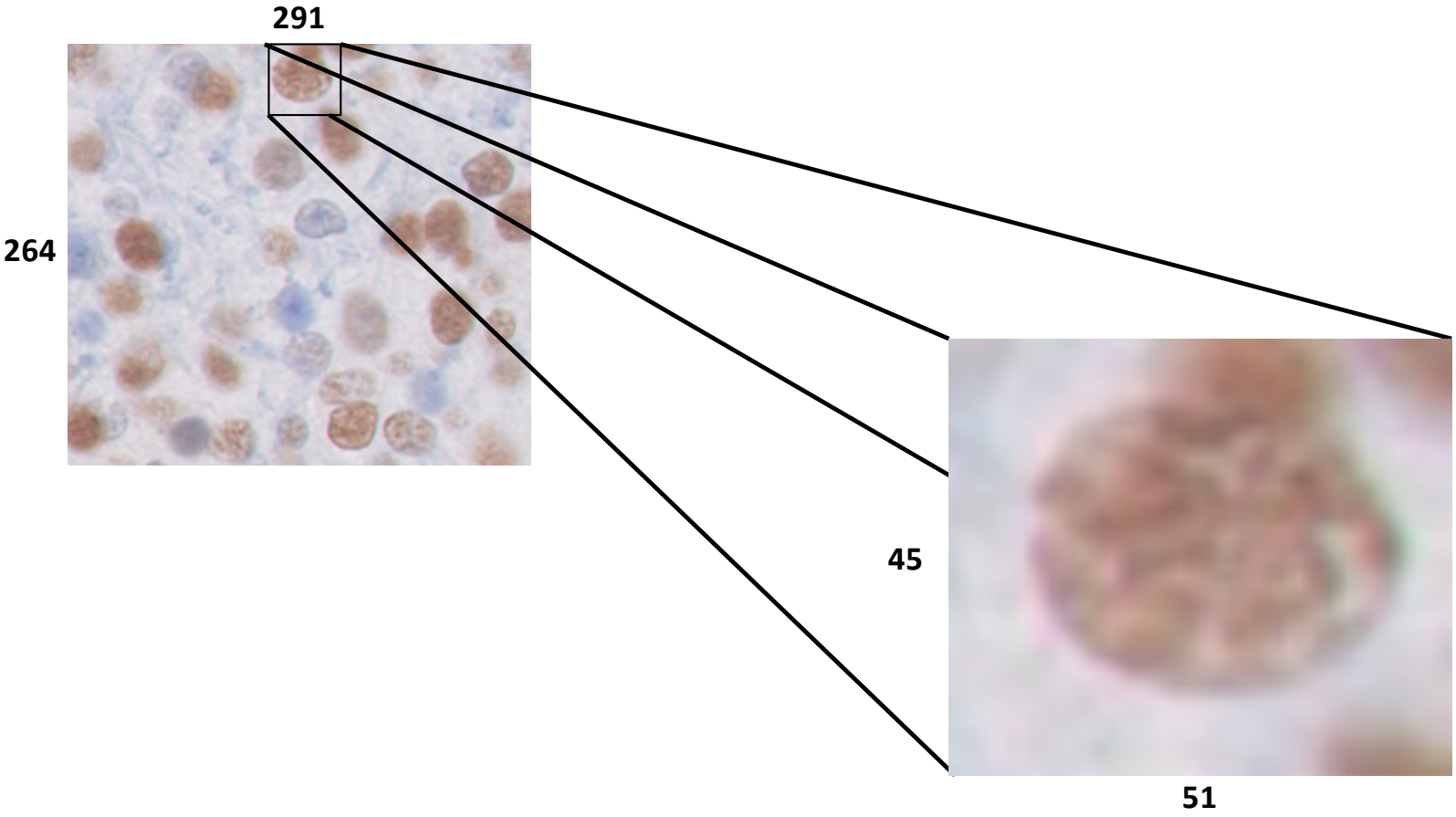
- **Computers can be trained to learn visual information by comparing new information with a set of historical references.**

What is a digital Image?



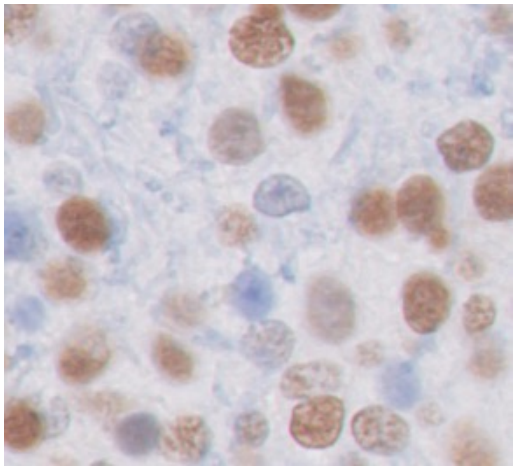
Oligodendroglioma stained with anti-OLIG2 antibody (DAB with Hematoxylin counterstain)

Increasing area covered by each pixel reduces resolution



The Image You See Comes From Three Datasets

RGB IMAGE

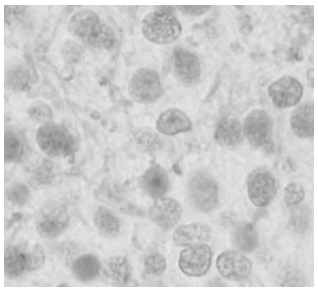


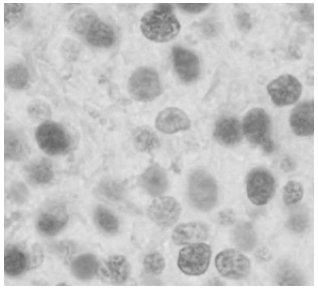
RED

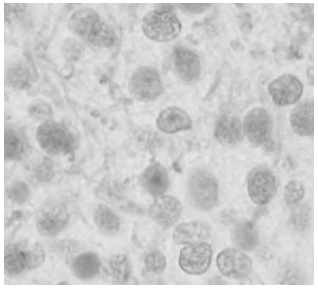
GREEN

BLUE

GRAYSCALE







Spreadsheet

0.6745098	0.68235294	0.69019608	0.69019608	0.6627451	0.65882353
0.6627451	0.66666667	0.6745098	0.69019608	0.6745098	0.65490196
0.67843137	0.69019608	0.70196078	0.71764706	0.71764706	0.72156863
0.70196078	0.69803922	0.70196078	0.70980392	0.72156863	0.73333333
0.74901961	0.7372549	0.70588235	0.68627451	0.68235294	0.6745098
0.76862745	0.76078431	0.70588235	0.68627451	0.69803922	0.6745098
0.79215686	0.77647059	0.7372549	0.72941176	0.73333333	0.71764706
0.78431373	0.77254902	0.76470588	0.74509804	0.72941176	0.7372549
0.76470588	0.78431373	0.76862745	0.72941176	0.68627451	0.69411765
0.74901961	0.79215686	0.77647059	0.71372549	0.65490196	0.65882353
0.70196078	0.74509804	0.76078431	0.7254902	0.69019608	0.66666667
0.68627451	0.71372549	0.74117647	0.7372549	0.73333333	0.69411765
0.70196078	0.70588235	0.69019608	0.68627451	0.72156863	0.7372549
0.70980392	0.71372549	0.67058824	0.68627451	0.72941176	0.75686275
0.76078431	0.77647059	0.74509804	0.74117647	0.75294118	0.7254902

0.62745098	0.63529412	0.64313725	0.65490196	0.64313725	0.62745098
0.63529412	0.63921569	0.65098039	0.66666667	0.65490196	0.65490196
0.65490196	0.64705882	0.63921569	0.65490196	0.65490196	0.64313725
0.68235294	0.63921569	0.60392157	0.60784314	0.61568627	0.58823529
0.70196078	0.68235294	0.64313725	0.61960784	0.61176471	0.58039216
0.69019608	0.70588235	0.67843137	0.65098039	0.65882353	0.63137255
0.67843137	0.69411765	0.68627451	0.65882353	0.64313725	0.63529412
0.63921569	0.6627451	0.69411765	0.63921569	0.58823529	0.60784314
0.60392157	0.65490196	0.6745098	0.62745098	0.57647059	0.58039216
0.57254902	0.64313725	0.65882353	0.61568627	0.57647059	0.55686275
0.54901961	0.6	0.62352941	0.61176471	0.6	0.57254902
0.56078431	0.57254902	0.58823529	0.60784314	0.62745098	0.60784314
0.61568627	0.61568627	0.59215686	0.59607843	0.63921569	0.63529412
0.66666667	0.67058824	0.63137255	0.63529412	0.66666667	0.64313725
0.67843137	0.69411765	0.66666667	0.65490196	0.66666667	0.61960784

0.64313725	0.6745098	0.70196078	0.68627451	0.63921569	0.61960784
0.64313725	0.67843137	0.71372549	0.71764706	0.68627451	0.67843137
0.63137255	0.6627451	0.69019608	0.70588235	0.70196078	0.68627451
0.62745098	0.62745098	0.63529412	0.65490196	0.67843137	0.63921569
0.6627451	0.65882353	0.63529412	0.62352941	0.62745098	0.58039216
0.66666667	0.67058824	0.62745098	0.61176471	0.62745098	0.57647059
0.67843137	0.67843137	0.65490196	0.64705882	0.64705882	0.61568627
0.65490196	0.66666667	0.67843137	0.65098039	0.62745098	0.62745098
0.63137255	0.6627451	0.65882353	0.61568627	0.56470588	0.57254902
0.61176471	0.65490196	0.63921569	0.57254902	0.50980392	0.51764706
0.53333333	0.58431373	0.60784314	[illegible]	[illegible]	[illegible]
0.48235294	0.52941176	0.57254902	0.56078431	0.54901961	0.52156863
0.54901961	0.56470588	0.56078431	0.5568275	0.5882353	0.58431373
0.60784314	0.61568627	0.57647059	0.6	0.64705882	0.61960784

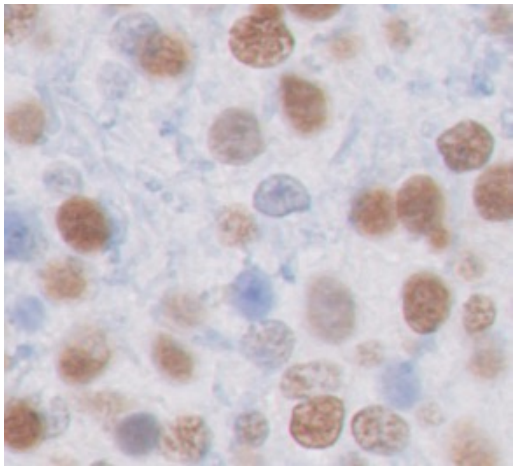
Common Pixel Scales

8 bit camera image (2^8 = 256):0 = Black, 255 = White

Many programs normalize this by dividing by 255 to make values 0 = Black, 1 = White

The Image You See Comes From Three Datasets

RGB IMAGE

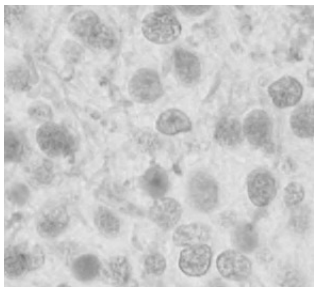


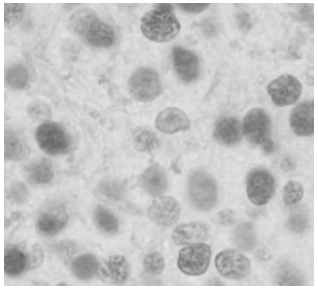
RED

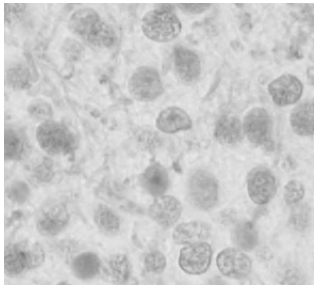
GREEN

BLUE

GRAYSCALE







Spreadsheet

0.6745098	0.68235294	0.69019608	0.69019608	0.6627451	0.65882353
0.6627451	0.66666667	0.6745098	0.69019608	0.6745098	0.65490196
0.67843137	0.69019608	0.70196078	0.71764706	0.71764706	0.72156863
0.70196078	0.69803922	0.70196078	0.70980392	0.72156863	0.73333333
0.74901961	0.7372549	0.70588235	0.68627451	0.68235294	0.6745098
0.76862745	0.76078431	0.70588235	0.68627451	0.69803922	0.6745098
0.79215686	0.77647059	0.7372549	0.72941176	0.73333333	0.71764706
0.78431373	0.77254902	0.76470588	0.74509804	0.72941176	0.7372549
0.76470588	0.78431373	0.76862745	0.72941176	0.68627451	0.69411765
0.74901961	0.79215686	0.77647059	0.71372549	0.65490196	0.65882353
0.70196078	0.74509804	0.76078431	0.7254902	0.69019608	0.66666667
0.68627451	0.71372549	0.74117647	0.7372549	0.73333333	0.69411765
0.70196078	0.70588235	0.69019608	0.68627451	0.72156863	0.7372549
0.70980392	0.71372549	0.67058824	0.68627451	0.72941176	0.75686275
0.76078431	0.77647059	0.74509804	0.74117647	0.75294118	0.7254902

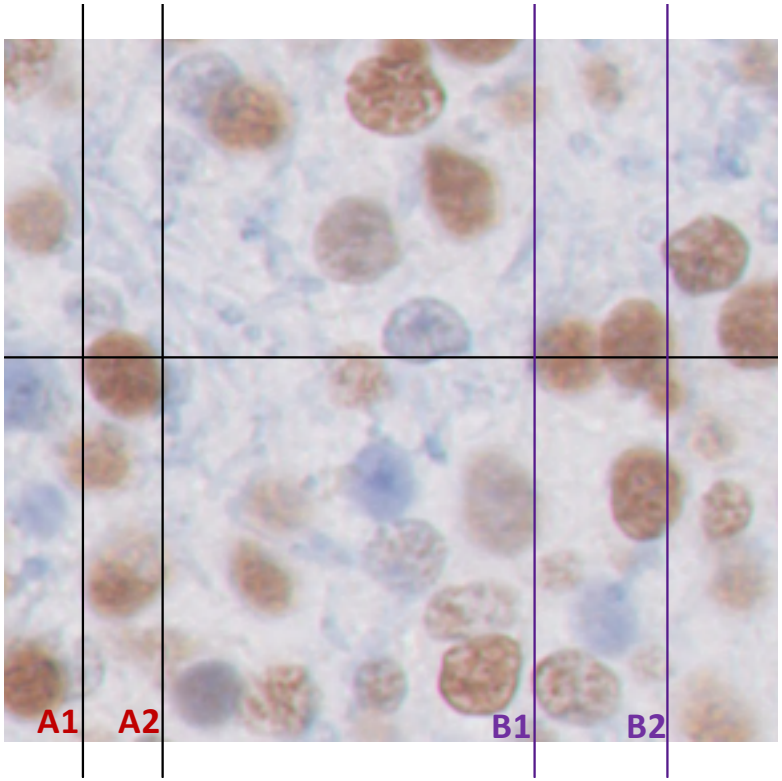
0.62745098	0.63529412	0.64313725	0.65490196	0.64313725	0.62745098
0.63529412	0.63921569	0.65098039	0.66666667	0.65490196	0.65490196
0.65490196	0.64705882	0.63921569	0.65490196	0.65490196	0.64313725
0.68235294	0.63921569	0.60392157	0.60784314	0.61568627	0.58823529
0.70196078	0.68235294	0.64313725	0.61960784	0.61176471	0.58039216
0.69019608	0.70588235	0.67843137	0.65098039	0.65882353	0.63137255
0.67843137	0.69411765	0.68627451	0.65882353	0.64313725	0.63529412
0.63921569	0.6627451	0.69411765	0.63921569	0.58823529	0.60784314
0.60392157	0.65490196	0.6745098	0.62745098	0.57647059	0.58039216
0.57254902	0.64313725	0.65882353	0.61568627	0.57647059	0.55686275
0.54901961	0.6	0.62352941	0.61176471	0.6	0.57254902
0.56078431	0.57254902	0.58823529	0.60784314	0.62745098	0.60784314
0.61568627	0.61568627	0.59215686	0.59607843	0.63921569	0.63529412
0.66666667	0.67058824	0.63137255	0.63529412	0.66666667	0.64313725
0.67843137	0.69411765	0.66666667	0.65490196	0.66666667	0.61960784

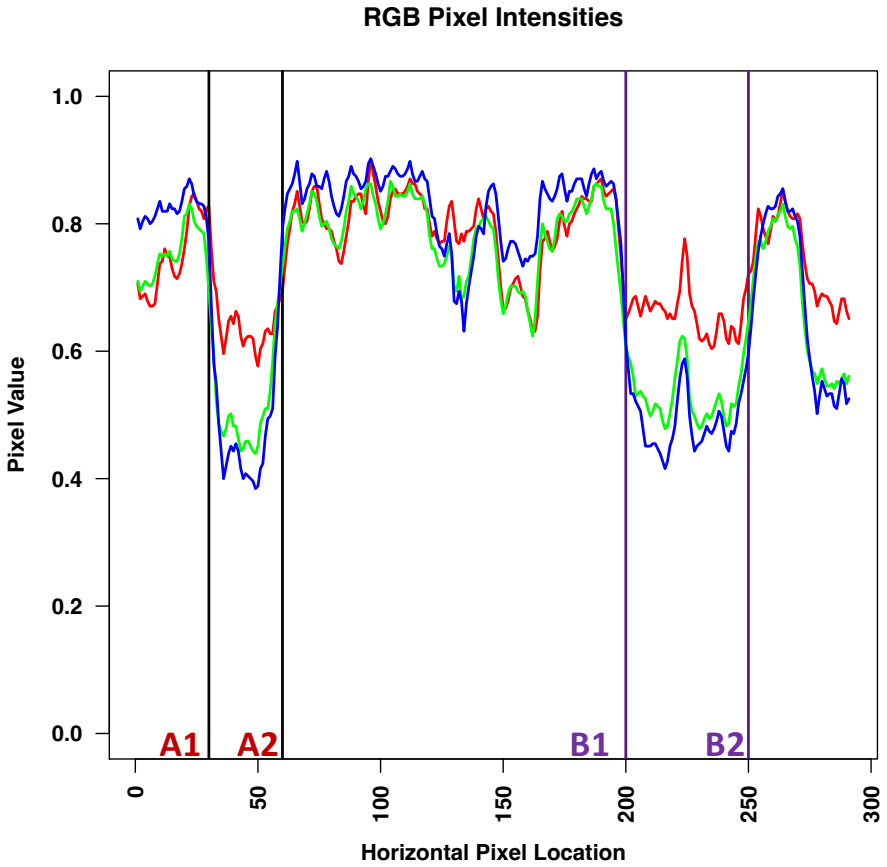
0.64313725	0.6745098	0.70196078	0.68627451	0.63921569	0.61960784
0.64313725	0.67843137	0.71372549	0.71764706	0.68627451	0.67843137
0.63137255	0.6627451	0.69019608	0.70588235	0.70196078	0.68627451
0.62745098	0.62745098	0.63529412	0.65490196	0.67843137	0.63921569
0.6627451	0.65882353	0.63529412	0.62352941	0.62745098	0.58039216
0.66666667	0.67058824	0.62745098	0.61176471	0.62745098	0.57647059
0.67843137	0.67843137	0.65490196	0.64705882	0.64705882	0.61568627
0.65490196	0.66666667	0.67843137	0.65098039	0.62745098	0.62745098
0.63137255	0.6627451	0.65882353	0.61568627	0.56470588	0.57254902
0.61176471	0.65490196	0.63921569	0.57254902	0.50980392	0.51764706
0.53333333	0.58431373	0.60784314	0.54509804	0.47843137	0.50196078
0.48235294	0.52941176	0.57254902	0.54901961	0.54901961	0.52156863
0.54901961	0.56470588	0.56078431	0.55686275	0.58823529	0.58431373
0.60784314	0.61568627	0.57647059	0.6	0.64705882	0.61960784

This spreadsheet, also known as a **MATRIX**, has a number in each cell which gives a ***value regarding the light intensity*** in that pixel.

Example of an Image Trace

```
> dim(im)
[1] 291 264   3
```





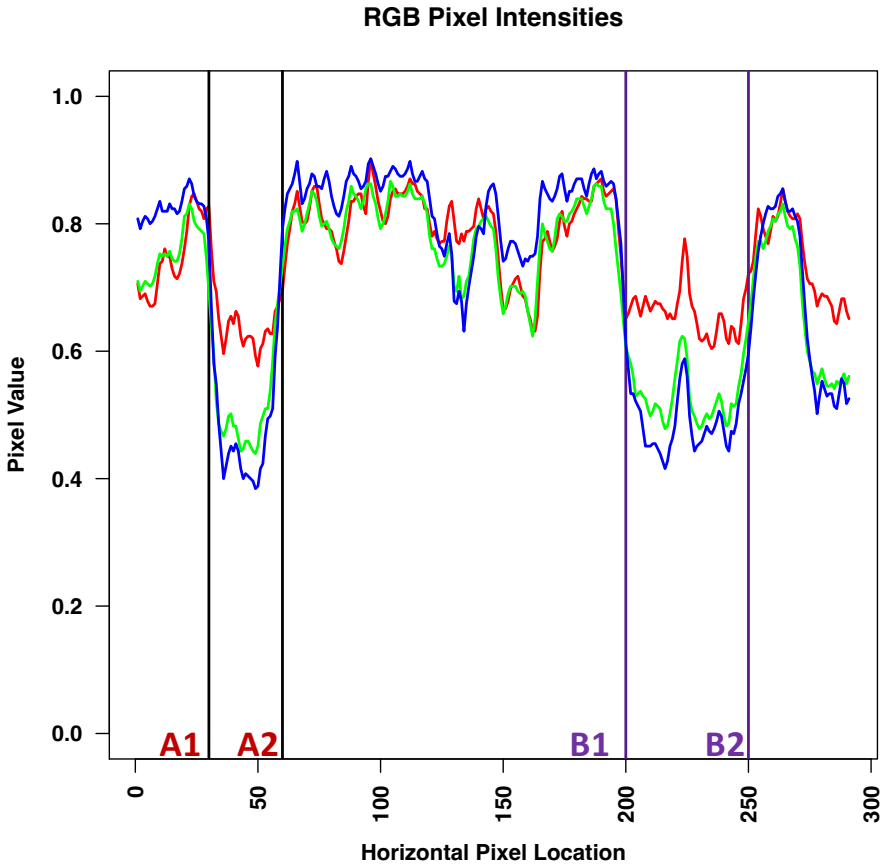
What does image analysis do?

Matrix of light intensities

0.6745098	0.68235294	0.69019608	0.69019608	0.6627451	0.65882353
0.6627451	0.66666667	0.6745098	0.69019608	0.6745098	0.65490196
0.67843137	0.69019608	0.70196078	0.71764706	0.71764706	0.72156863
0.70196078	0.69803922	0.70196078	0.70980392	0.72156863	0.73333333
0.74901961	0.7372549	0.70588235	0.68627451	0.68235294	0.6745098
0.76862745	0.76078431	0.70588235	0.68627451	0.69803922	0.6745098
0.79215686	0.77647059	0.7372549	0.72941176	0.73333333	0.71764706
0.78431373	0.77254902	0.76470588	0.74509804	0.72941176	0.7372549
0.76470588	0.78431373	0.76862745	0.72941176	0.68627451	0.69411765
0.74901961	0.79215686	0.77647059	0.71372549	0.65490196	0.65882353
0.70196078	0.74509804	0.76078431	0.7254902	0.69019608	0.66666667
0.68627451	0.71372549	0.74117647	0.7372549	0.73333333	0.69411765
0.70196078	0.70588235	0.69019608	0.68627451	0.72156863	0.7372549
0.70980392	0.71372549	0.67058824	0.68627451	0.72941176	0.75686275
0.76078431	0.77647059	0.74509804	0.74117647	0.75294118	0.7254902

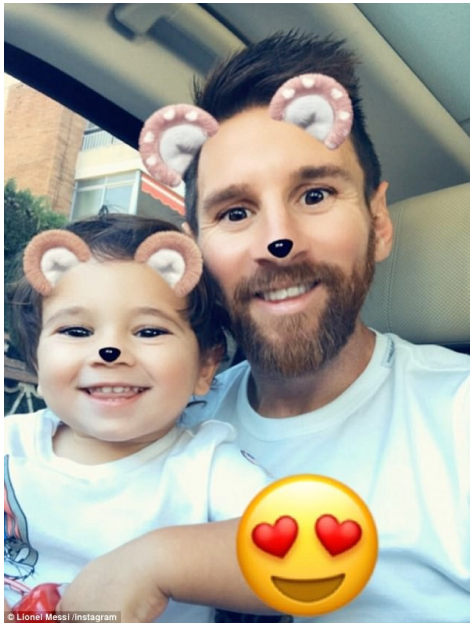
Most image analysis performs mathematical calculations, or functions, that are intended to provide you with information that you would like to extract from the image.

The pathologist plays a key role in this revolution since we already know what information is needed from an image.

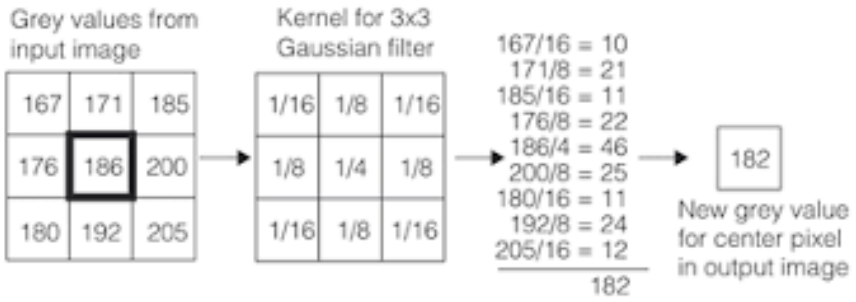


Filters, what are they and why do they matter

Social Media Filters (Messi, Daily Mail)

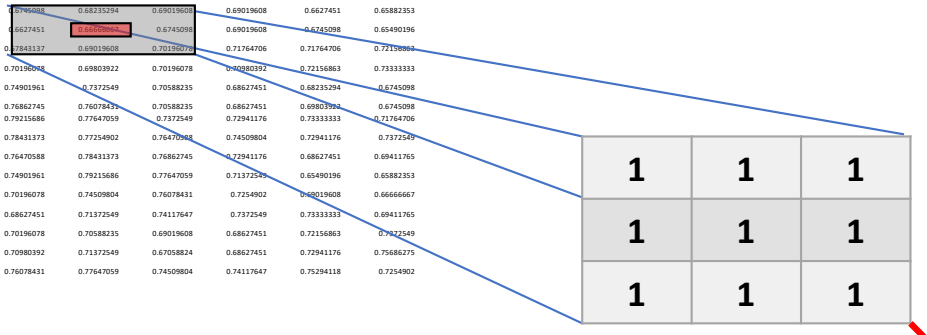


An Image Processing Filter (aka, Kernel)



These filters can be used to modify an image.
The process by which they do so is termed convolution

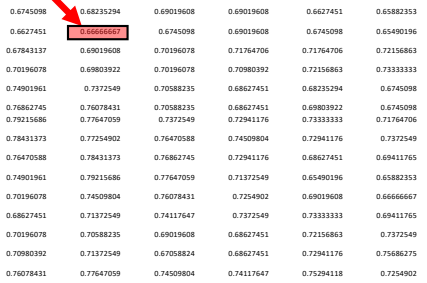
How do filters work? (eg., 3x3 filter)



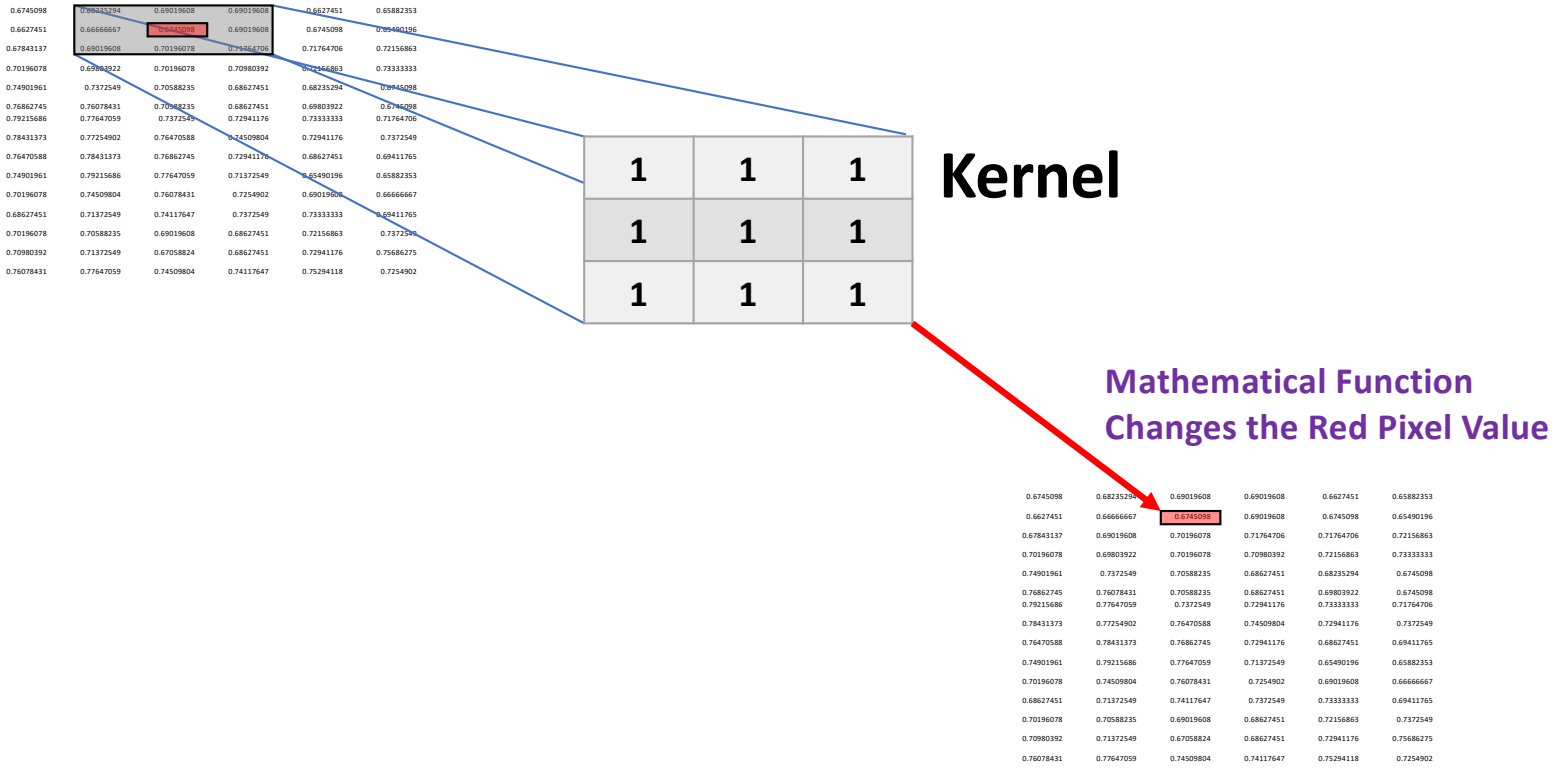
1	1	1
1	1	1
1	1	1

Kernel

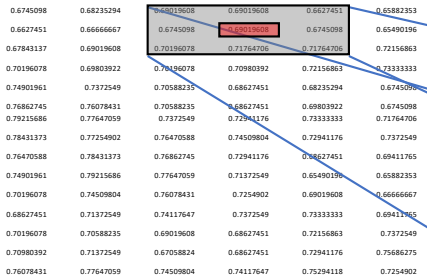
Mathematical Function Changes the Red Pixel Value



How do filters work? (eg., 3x3 filter)



How do filters work? (eg., 3x3 filter)

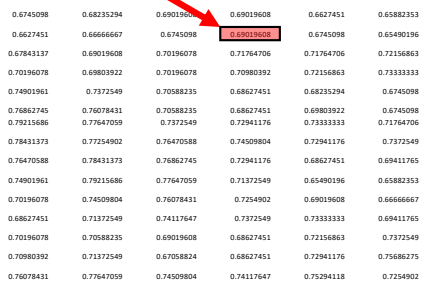


1	1	1
1	1	1
1	1	1

Kernel

Mathematical Function Changes the Red Pixel Value

...



This process is called *CONVOLUTION*

How is this working?

Kernel

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

Image

1x 0.77	1x 0.77	1x 0.78	1x 0.78	1x 0.78
1x 0.78	1x 0.8	1x 0.82	1x 0.82	1x 0.82
1x 0.79	1x 0.82	Result	1x 0.85	1x 0.86
1x 0.83	1x 0.84	1x 0.85	1x 0.85	1x 0.86
1x 0.85	1x 0.85	1x 0.85	1x 0.85	1x 0.85

Multiple each corresponding cell by the value, and then take the median to replace the central pixel value.

Gaussian Smoothing

The kernel has a similar function in this. The Gaussian outputs a `weighted average' of each pixel's neighborhood, with the average weighted more towards the value of the central pixels.

Difference between techniques is based on the difference of the kernel function.

0	1	2	1	0
1	6	10	6	1
2	10	*16*	10	2
1	6	10	6	1
0	1	2	1	0

Take each cell, and multiply times the value, and then take the average. Note that it is weighted for the central pixel in the kernel.

Less Weight **More Weight**

Gaussian Smoothing

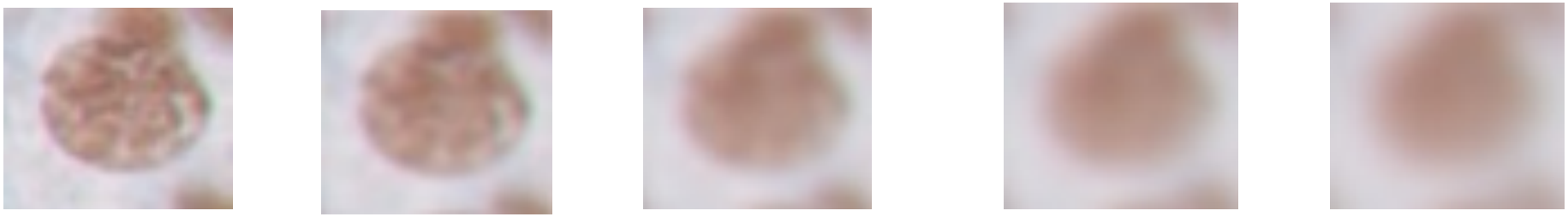
The kernel has a similar function in this. The Gaussian outputs a `weighted average' of each pixel's neighborhood, with the average weighted more towards the value of the central pixel.

Difference between techniques is based on the difference of the kernel function.

MEDIAN FILTER

Original | 3x3 | 5x5 | 7x7 | 9x9

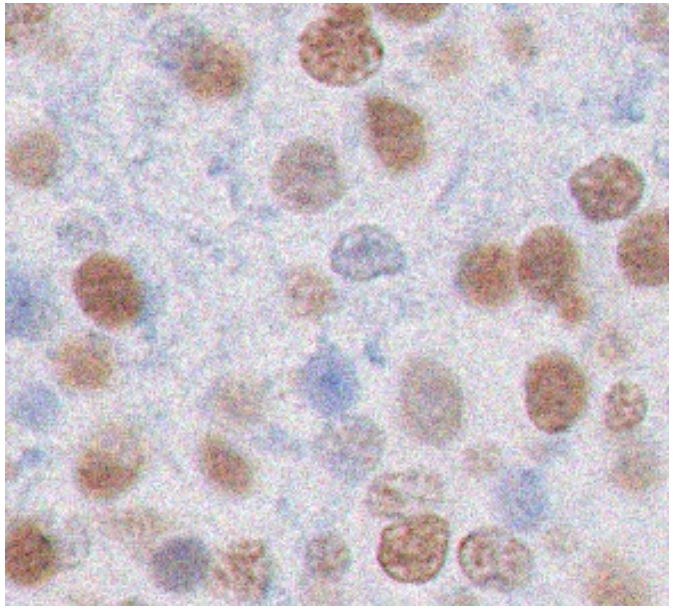
GAUSSIAN FILTER



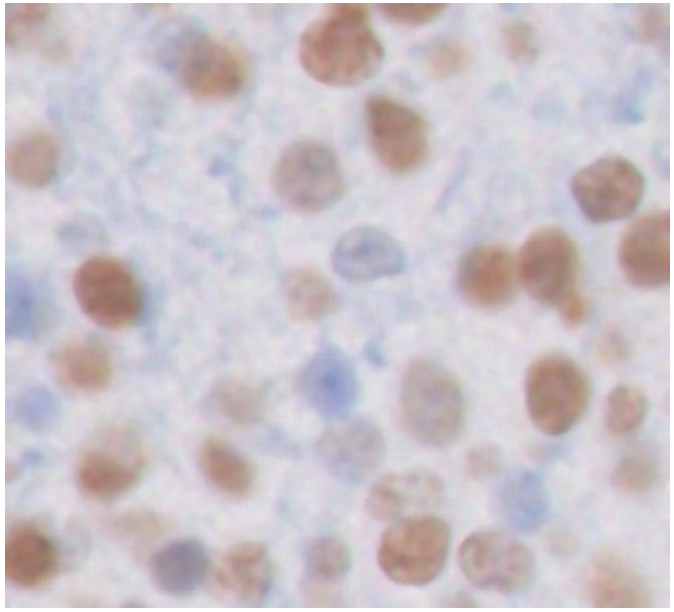
Original | $\sigma = 1$ | $\sigma = 2$ | $\sigma = 3$ | $\sigma = 4$

Example of application: Noise Reduction. Note that Gaussian Filter Preserves Edges

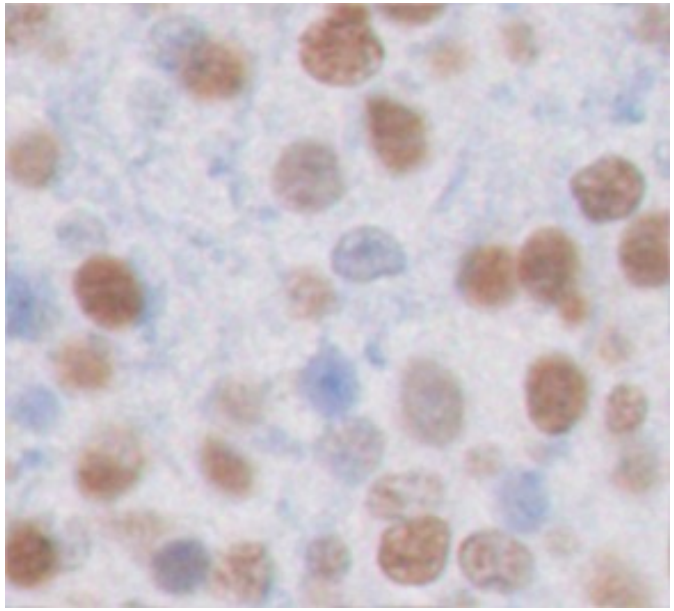
Image with Noise Added



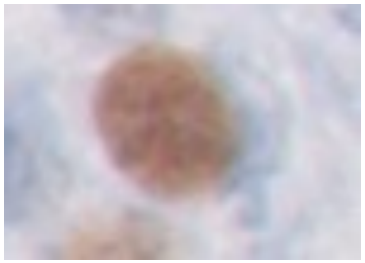
Median Filter, 5x5 kernal



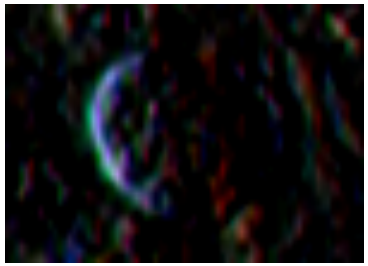
Gaussian Filter, sigma = 1



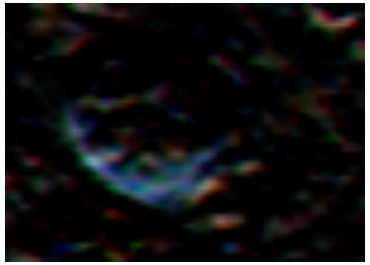
Example of application: Edge Detection. Note that Gaussian Filter Preserves Edges



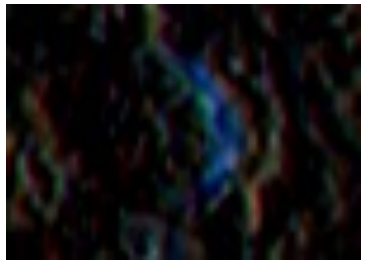
1	0	-1
2	0	-2
1	0	-1



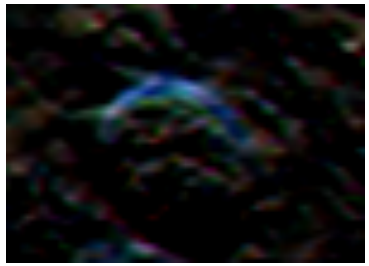
1	2	1
0	0	0
-1	-2	-1



-1	0	1
-2	0	2
-1	0	1



-1	-2	-1
0	0	0
1	2	1



Example application: identification of membrane staining such as Her2Neu

Image Segmentation

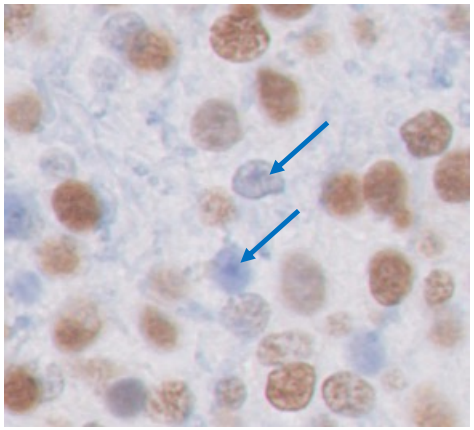
- Partition an image into various segments, with each segment holding a specific label.
- Potential Applications:
  - I want to identify only the nuclei of an image.
  - I want to identify the vessels of an image
  - I want to identify necrosis versus total tumor area.

  - **These techniques provide an unbiased methodology to obtain percentages.**

Classical Color Deconvolution

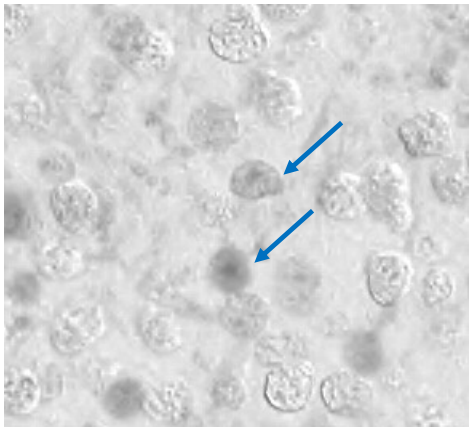
Color deconvolution is a method used in diagnostic brightfield microscopy to transform **color images** of multiple stained biological samples into **images** representing the stain concentrations. It is applied by decomposing the absorbance values of stain mixtures into absorbance values of single stains.

Original Image



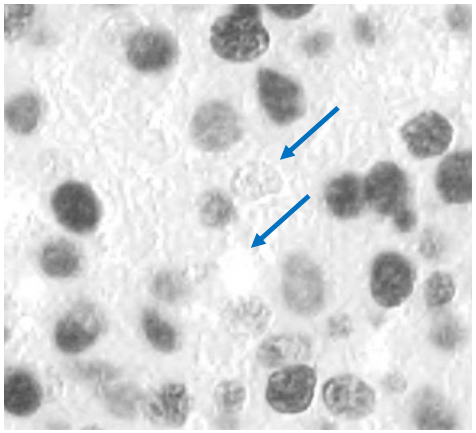
Colormode = RGB

Hematoxylin



Colormode = Grayscale

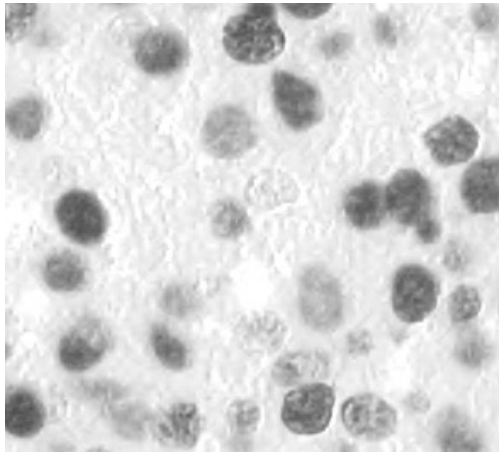
DAB (Olig2)



Colormode = Grayscale

Inversion (subtracting the actual pixel value from the maximal possible pixel value)

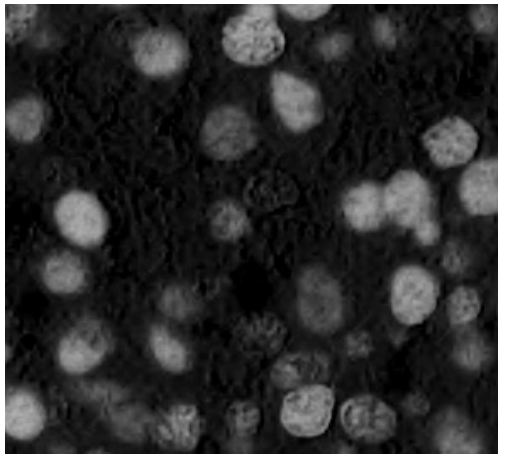
DAB (Olig2)



White = Negative
Black = Positive

1-image

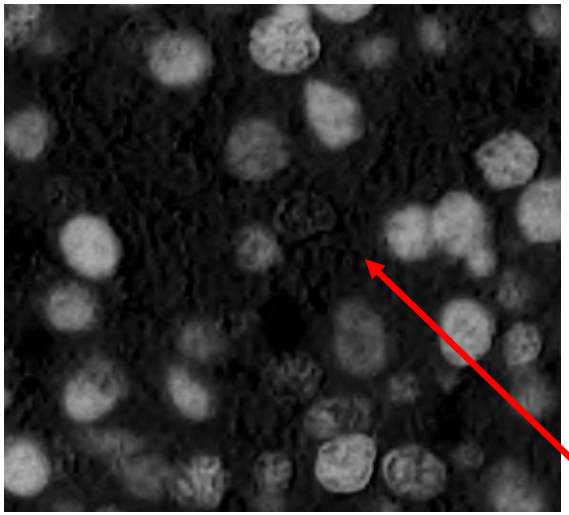
Inverted Image



White = Positive
Black = Negative

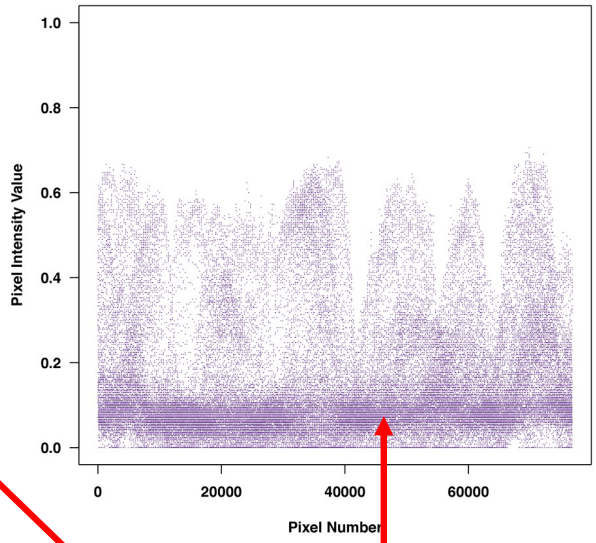
Distribution of Pixel Intensities- continuous numbers from 0 (black) to 1(white)

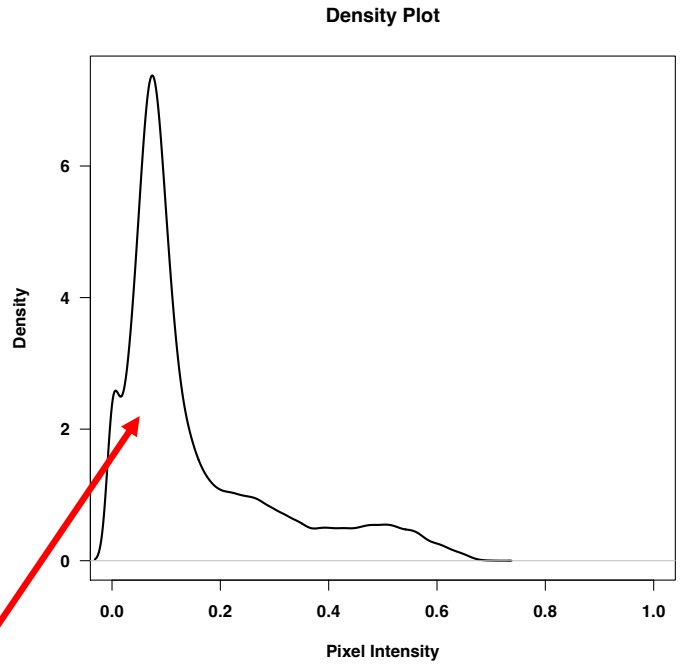
Inverted DAB (Olig2) Image



Dim = 291 x 264, total of 76824

Scatter Plot of Pixels

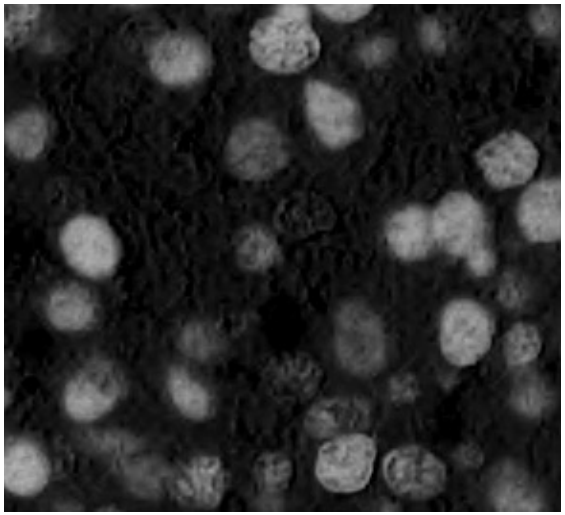




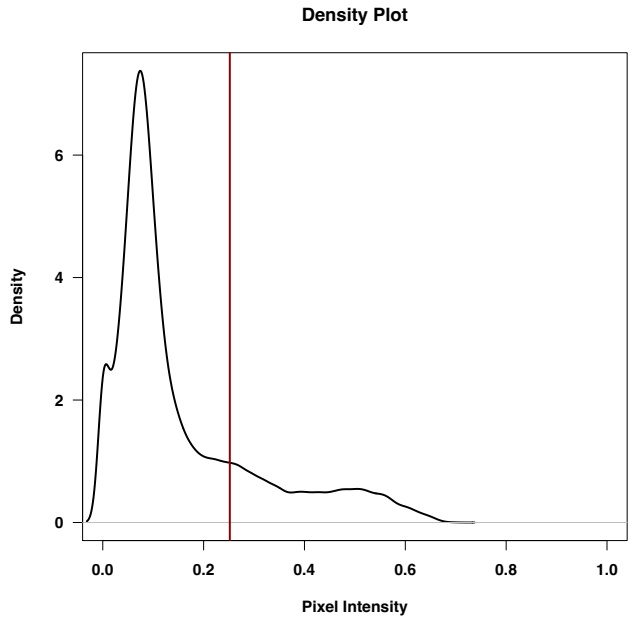
Majority of the pixels of dark (close to 0)

Global Thresholding: Choosing a value in the pixel histogram of the entire image and then assigning each pixel to either 0 or 1 (i.e., a binary output)

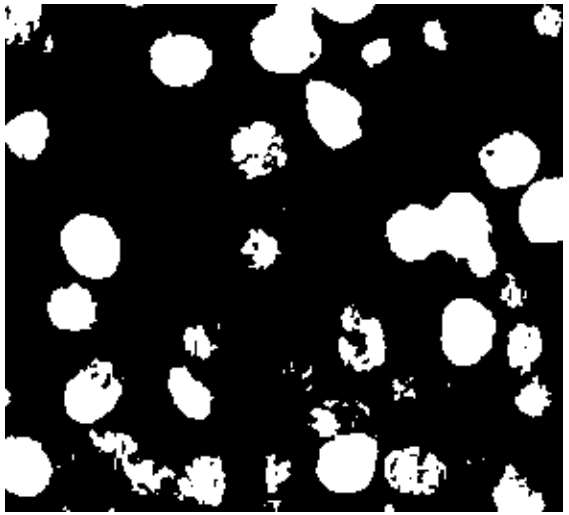
Inverted DAB (Olig2) Image



Dim = 291 x 264, total of 76824

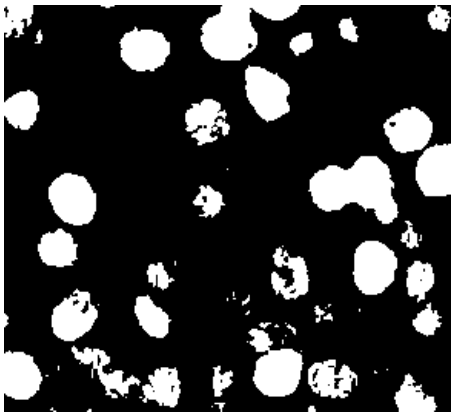


Binary DAB (Olig2) Image



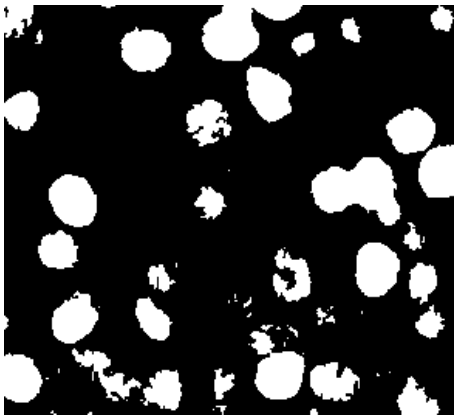
Segmentation of each DAB nucleus

Binary Thresholded Image



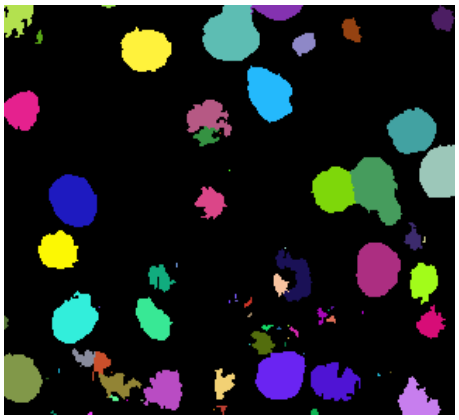
Threshold Operation

Fill Hull and
Watershed Operation



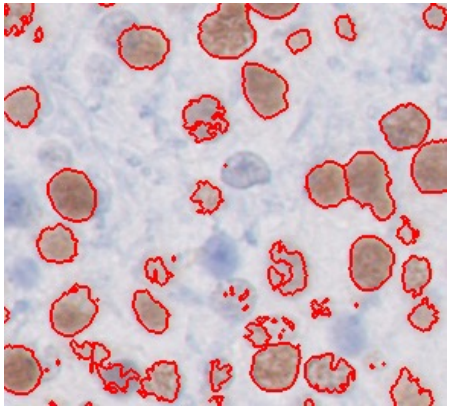
Morphological Operation

Generation of
Mask



Segmentation Operation

Mask Overlay
On Original Image

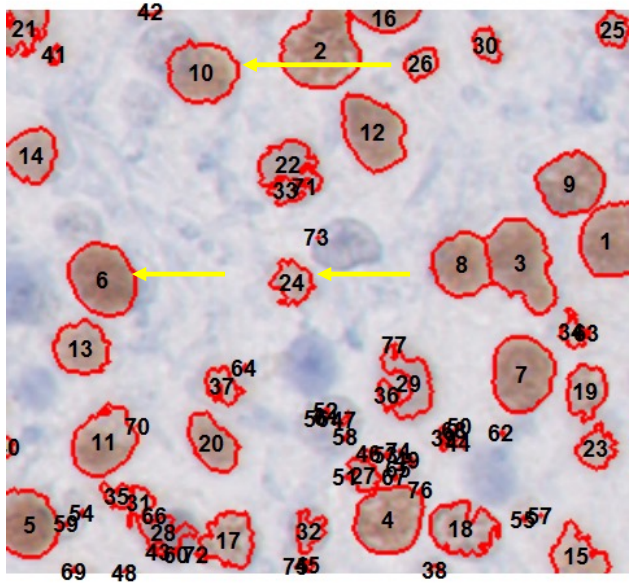


Overlay Operation

Having placed a mask on top of our image, we can now extract multiple imaging features from each segmented object.

What imaging features can we extract?

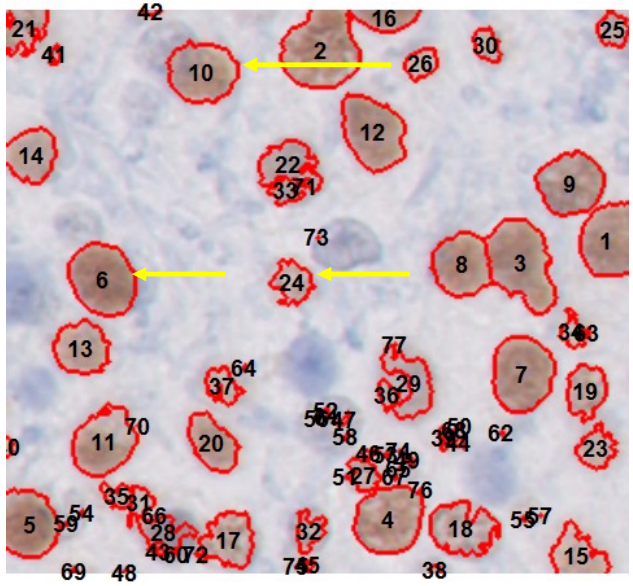
- **Location features within the image. Usually not useful in pathology unless relative to a landmark or an internal fiducial mark.**



Location Example:
Object 10 is located at coordinates (91, 30)
Object 6 is located at coordinates (45, 126)
Object 24 is located at coordinates (134, 127)

What imaging features can we extract?

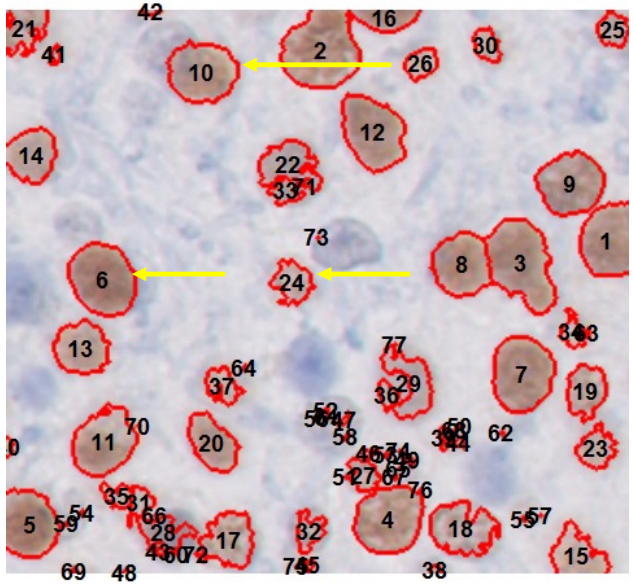
- **Shape features (roundness, eccentricity, area, perimeter).**



Location Example:
Object 10 has an area equal to 707 pixels^2
Object 6 has an area equal to 808 pixels^2
Object 24 has an area equal to 156 pixels^2

What imaging features can we extract?

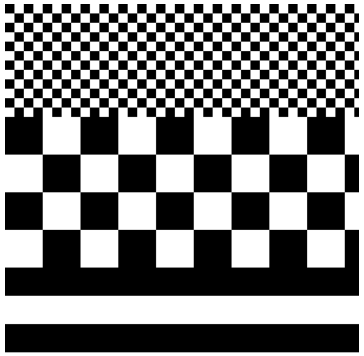
- **Pixel intensity feature (mean pixel intensities, standard deviations, variance, quartiles).**



Location Example:
Object 10 has mean intensity of (**0.67**, **0.61**, **0.58**).
Object 6 has mean intensity of (**0.66**, **0.52**, **0.49**).
Object 24 has mean intensity of (**0.74**, **0.67**, **0.68**).

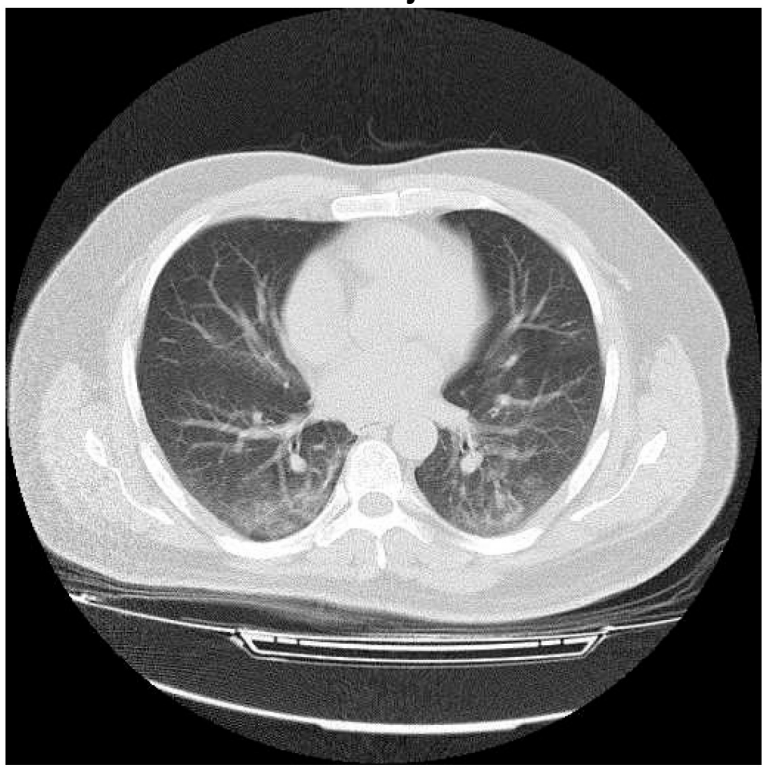
Texture Features

- Gives us information about the spatial arrangement of color or intensities in an image or selected region of an image.
- Many people use the Haralik features
  - Set of 14 texture features described in a manuscript published by Haralick, et al., IEEE Transactions on Systems, Man, and Cybernetics. 1973, page 610-621.



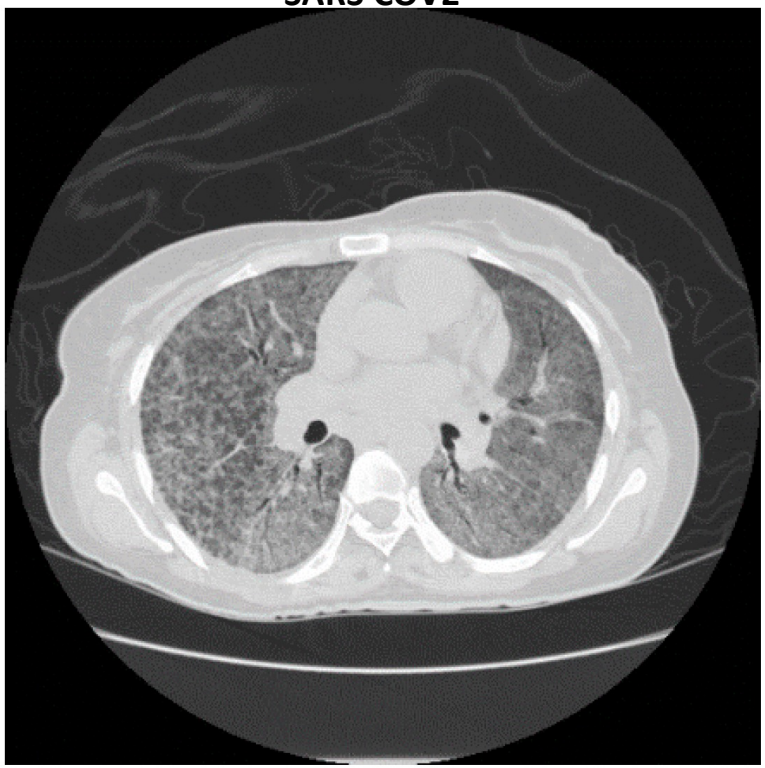
Texture Features

Control Subject



(a)

SARS COV2

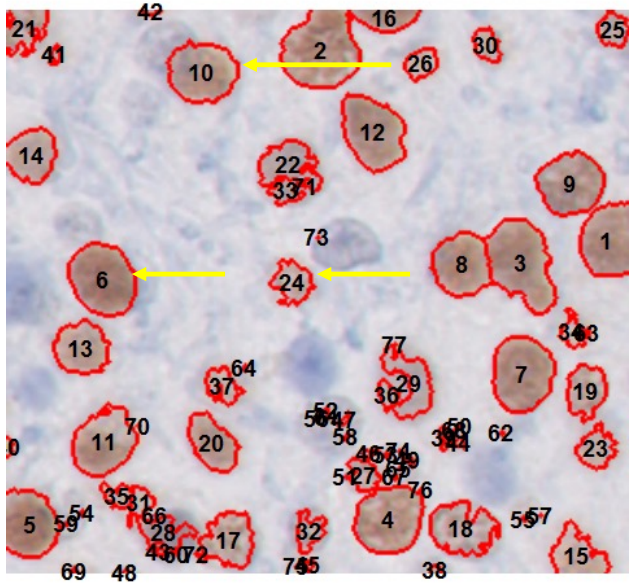


(b)

Hasan, et al., *Entropy* **2020**, *22*(5), 517

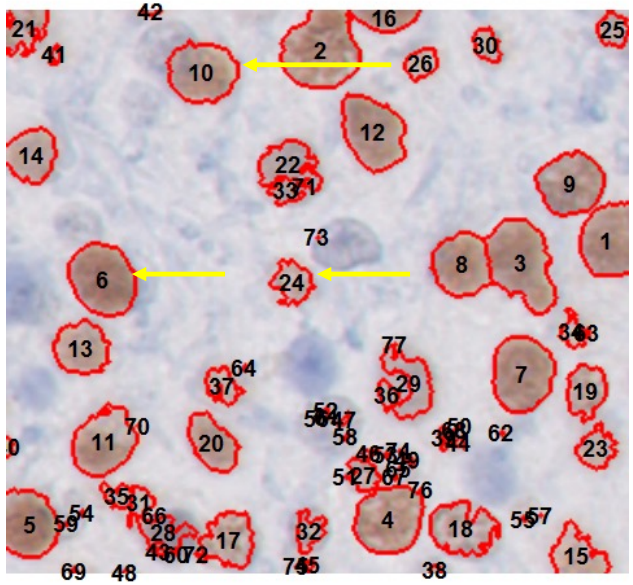
What imaging features can we extract?

- **Texture Features**



Location Example:
Object 10 has entropy of (**1.4**, **1.4**, **1.4**).
Object 6 has entropy of (**1.6**, **1.5**, **1.4**).
Object 24 has entropy of (**1.0**, **1.0**, **1.2**).

Each object (nucleus) is now segmented and features identified from the image mask



Imaging Data for Each Feature includes

- 89 features in the red
- 89 features in the green
- 89 features in the blue
- **Total of 267 features for each nucleus!!!**

This is a lot of data, and we need computational support to take advantage of this.

Machine Learning Models helps us transform these massive amounts of data into knowledge.

- **Incorporation of ML techniques is a necessary component of Modern Pathology Practice**

What is a Model?

What is a Model?

What is a Model?

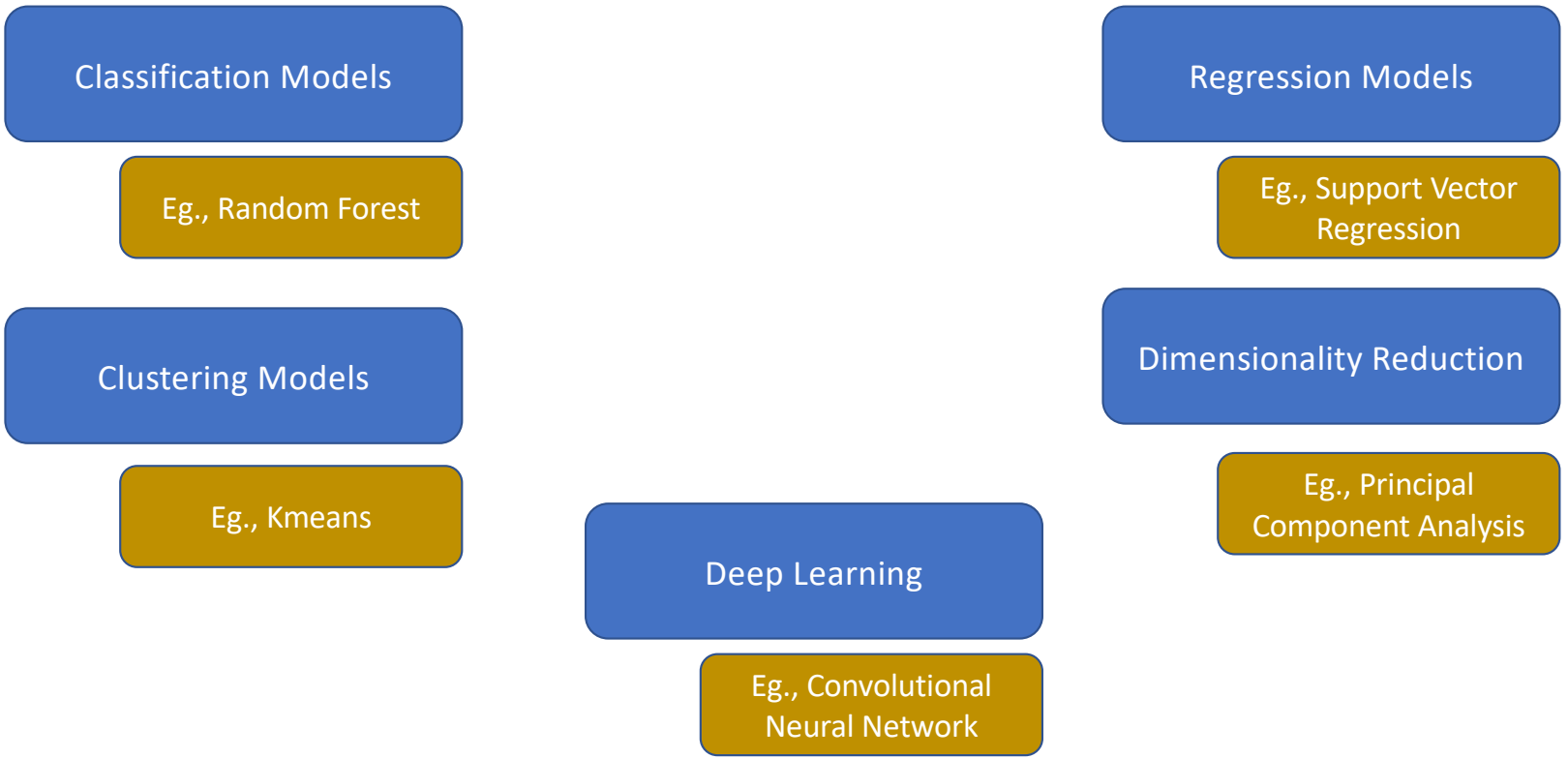
What is a model?

Now we use model to:

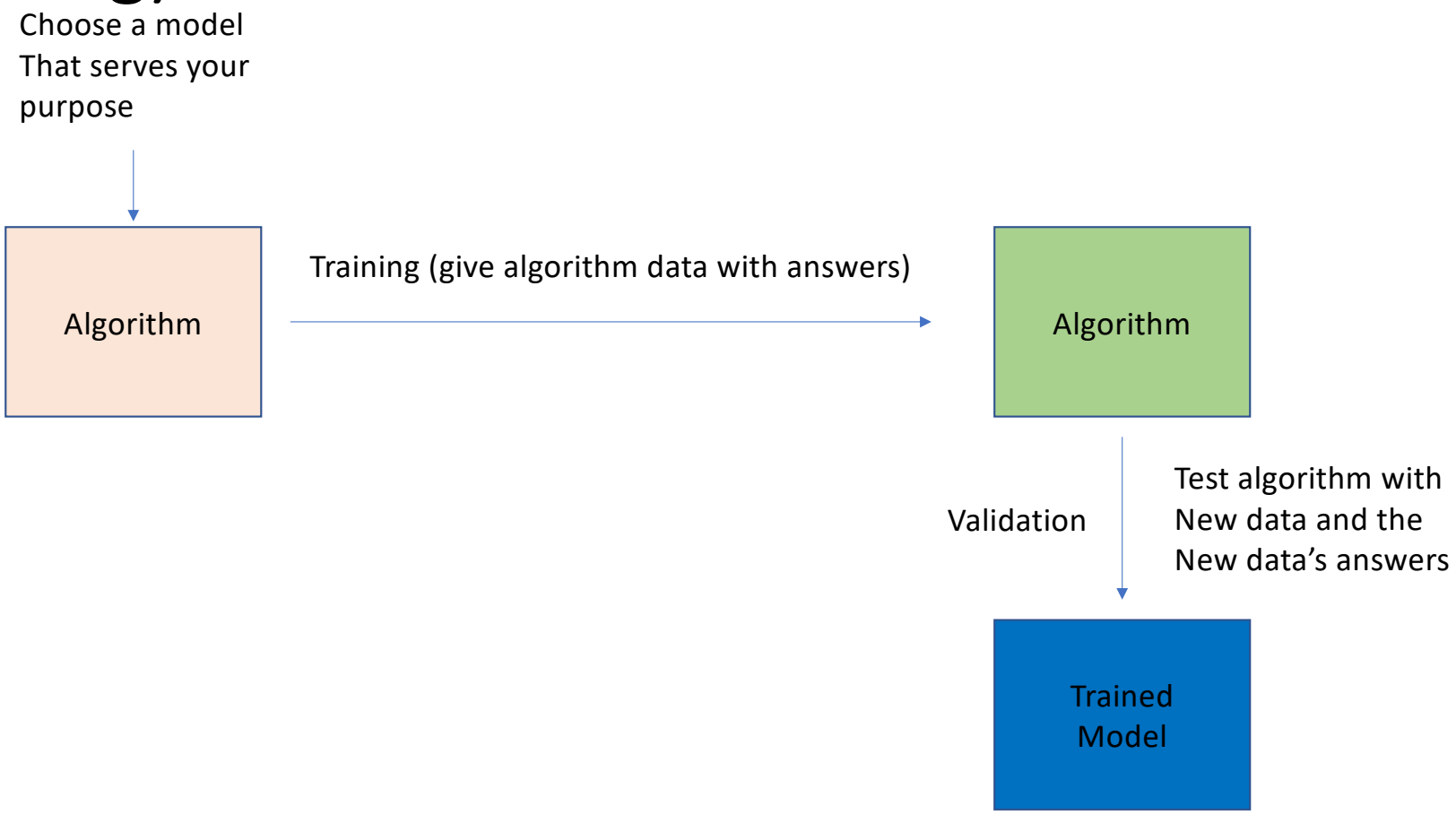
- **Describe a mathematical formula.**
- **Explore relationships between variables.**
- **To simulate possibilities.**

$$\frac{dG}{dt} = f_1(G, I) + J(t),$$
$$\frac{dI}{dt} = f_2(G, I),$$

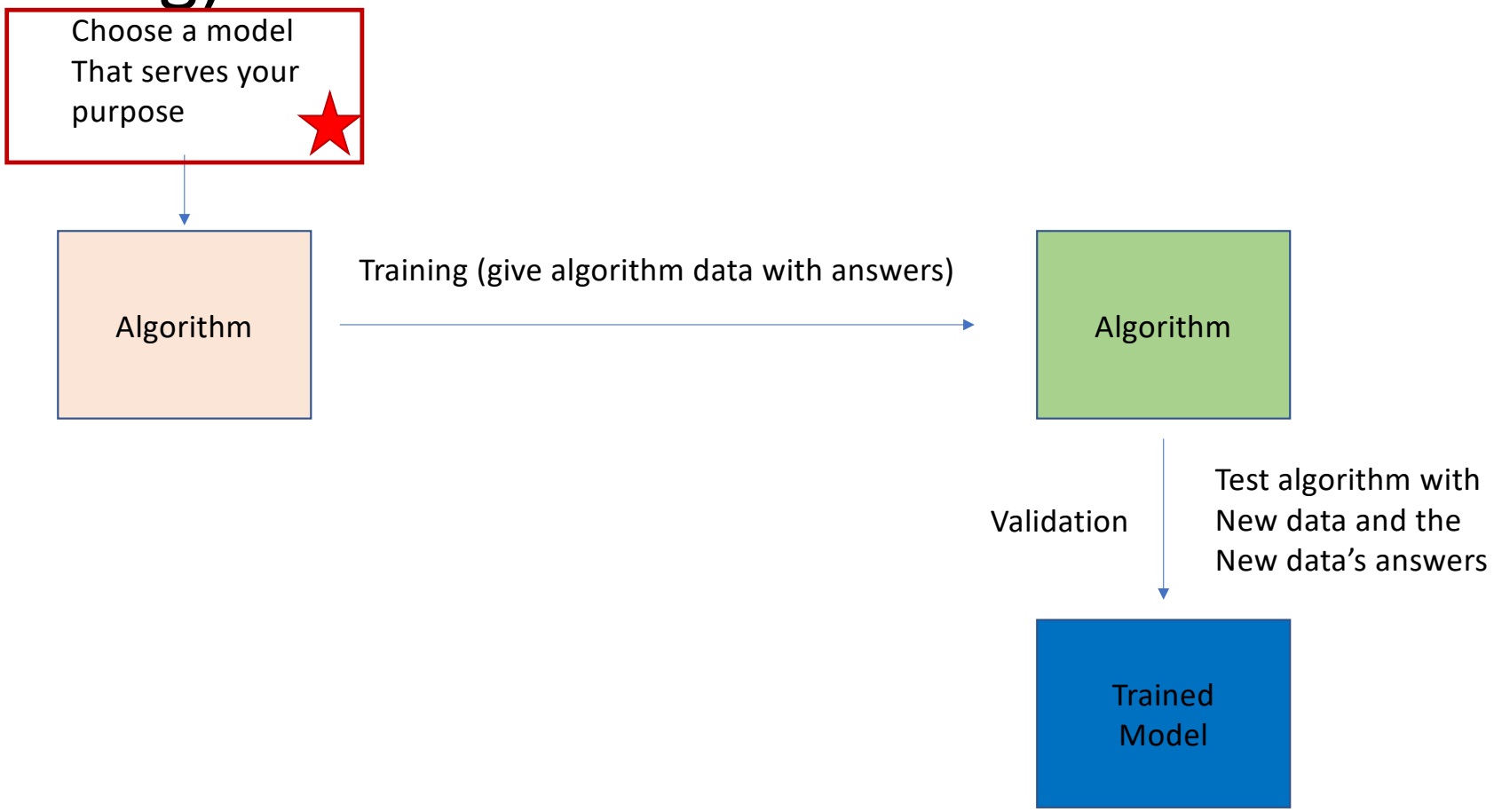
Many pre-existing models to utilize.



How does ML work? (example of supervised learning)



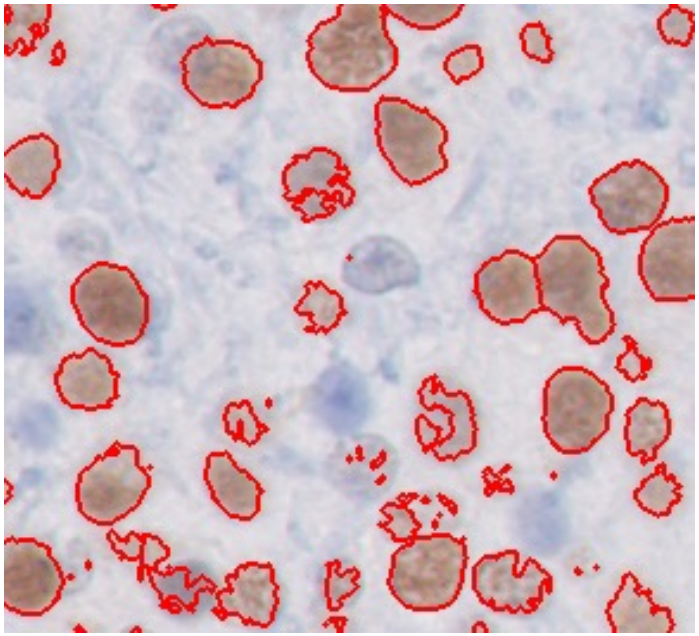
How does ML work? (example of supervised learning)



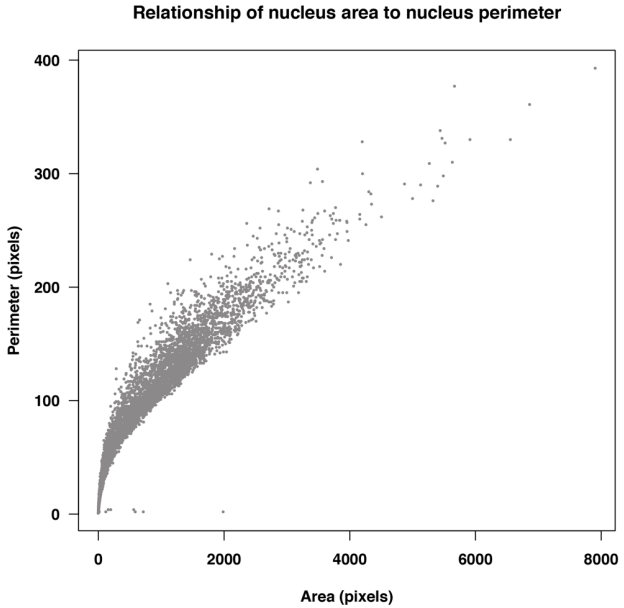
Regression Techniques Provide Predictions of Numbers

Used to predict a number or predict a probability from 0 (unlikely) to 1 (highly likely)

Segmented Cell Data



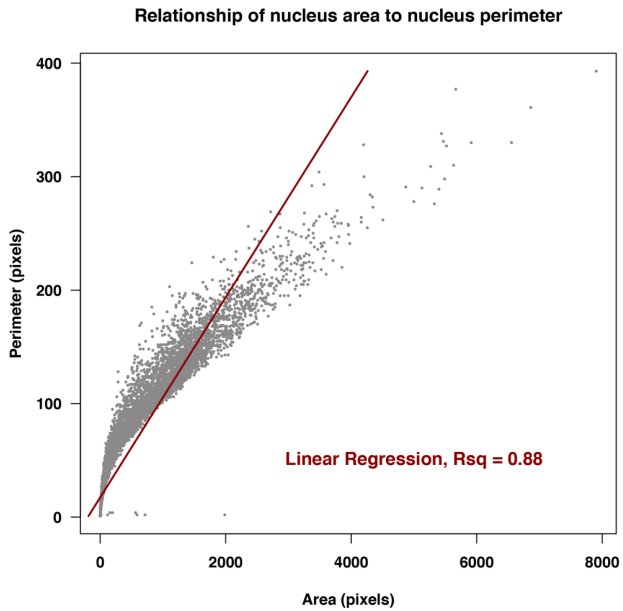
Scatter Plot



Best Fit Curves are Simple Examples of Machine Learning

Linear Regression

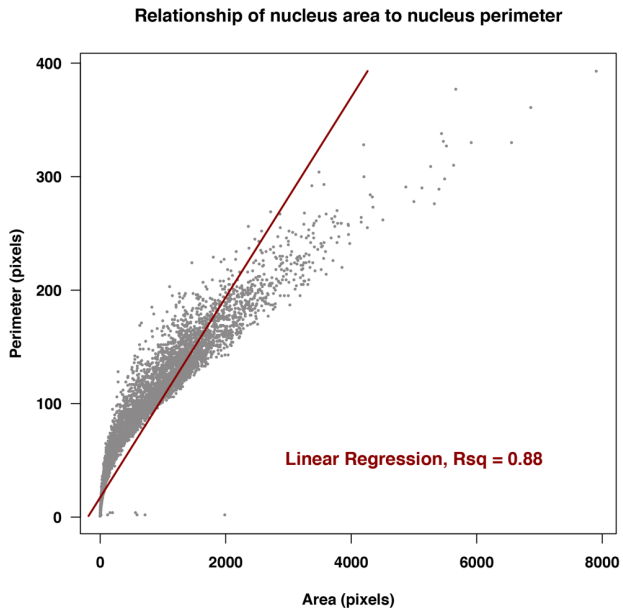
- **Per = β_0 A^0 + β_1 A^1**



Best Fit Curves are Simple Examples of Machine Learning

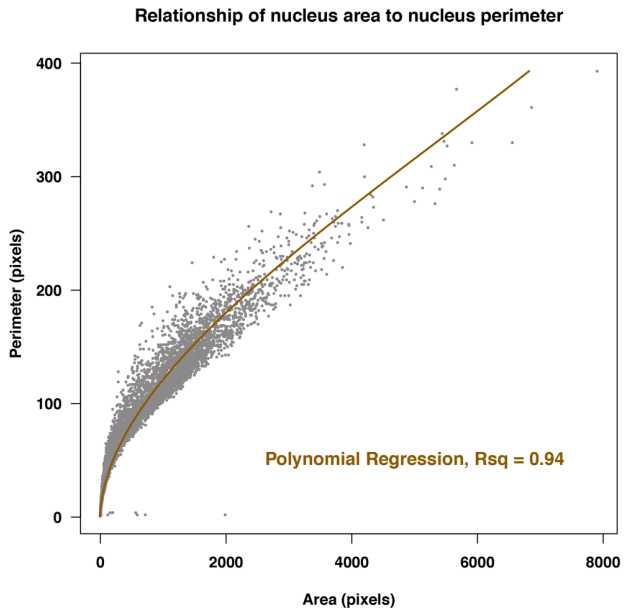
Linear Regression

- **Per = $\beta_0 A^0 + \beta_1 A^1$**



Polynomial Regression

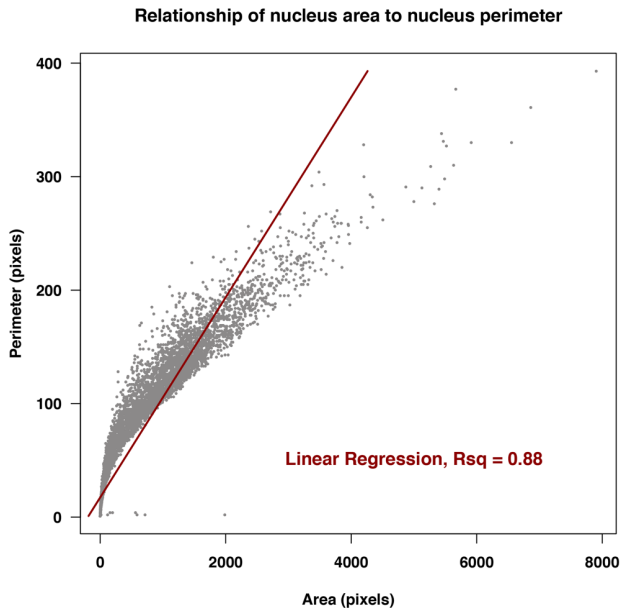
- **Per = $\beta_0 A^0 + \beta_1 A^1 + \beta_2 A^2 + \beta_3 A^3$**



If you have ever had a computer make a best-fit curve for you, you have already done machine learning!!!

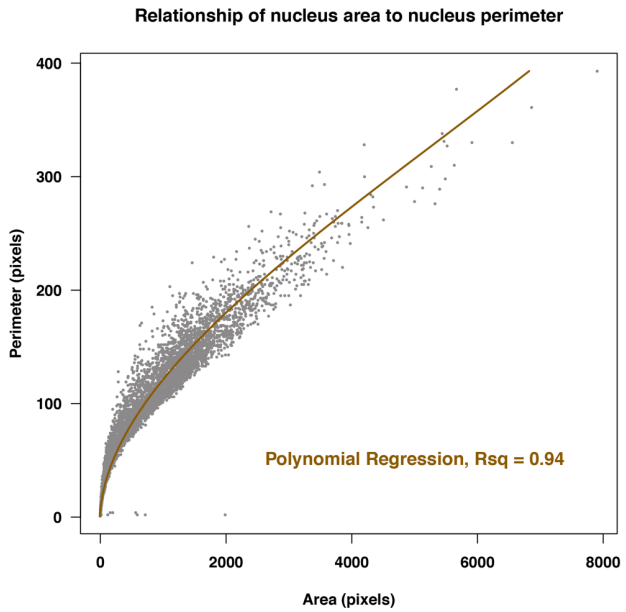
Linear Regression

- **Per = $\beta_0 + \beta_1 A$**



Polynomial Regression

- **Per = $\beta_0 A^0 + \beta_1 A^1 + \beta_2 A^2 + \beta_3 A^3$**



Data types and how to handle them.

Structured Data

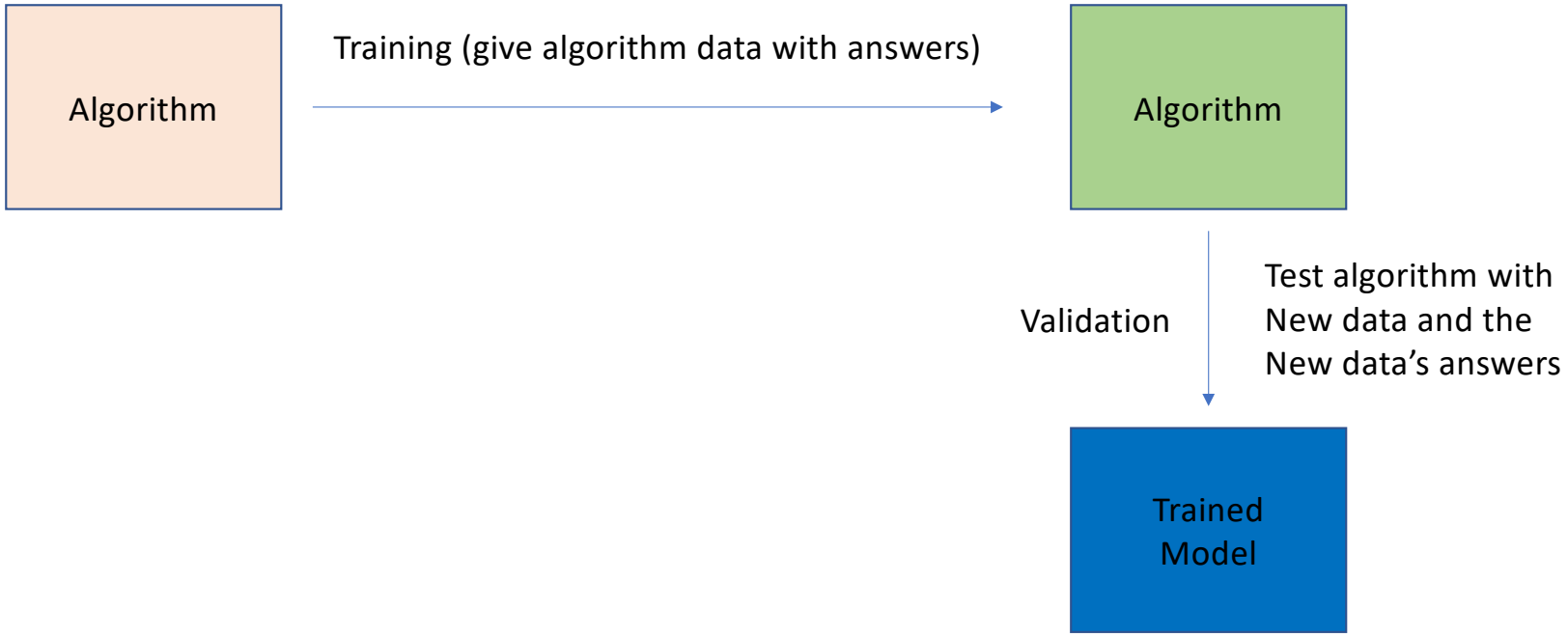
- Easiest data to interact with.
- Typically contains observations as rows, and features of each observation as columns, such as an excel file.
- Example: the data table we had from our collaborators.
- Non-deep learning approaches are often better than deep learning approaches.
  - *Very useful when mechanism is more important over accuracy.*
  - Eg., Telling a patient, “For every kg of weight loss you have, your probability of having DM reduces by x percent.”

Unstructured data

- Data that is not easily condensable into a data table.
- Eg., images, videos, and audio.
- Typically handled in two manners:
  - Extract features and convert the unstructured data to structured data. (eg., obtain a ki67 proliferation index, or quantify Her2/neu into an ordinal scale of 1+-3+.
  - Utilize deep learning methodologies.
    - Need to be done *if and only if* there is some form of normalization of the data with standardized SOP’s, and large quantities of data.
    - *Very useful when accuracy is MORE important than mechanism.*

Regression Techniques Provide Predictions of Numbers

Used to predict a number or predict a probability from 0 (unlikely) to 1 (highly likely)

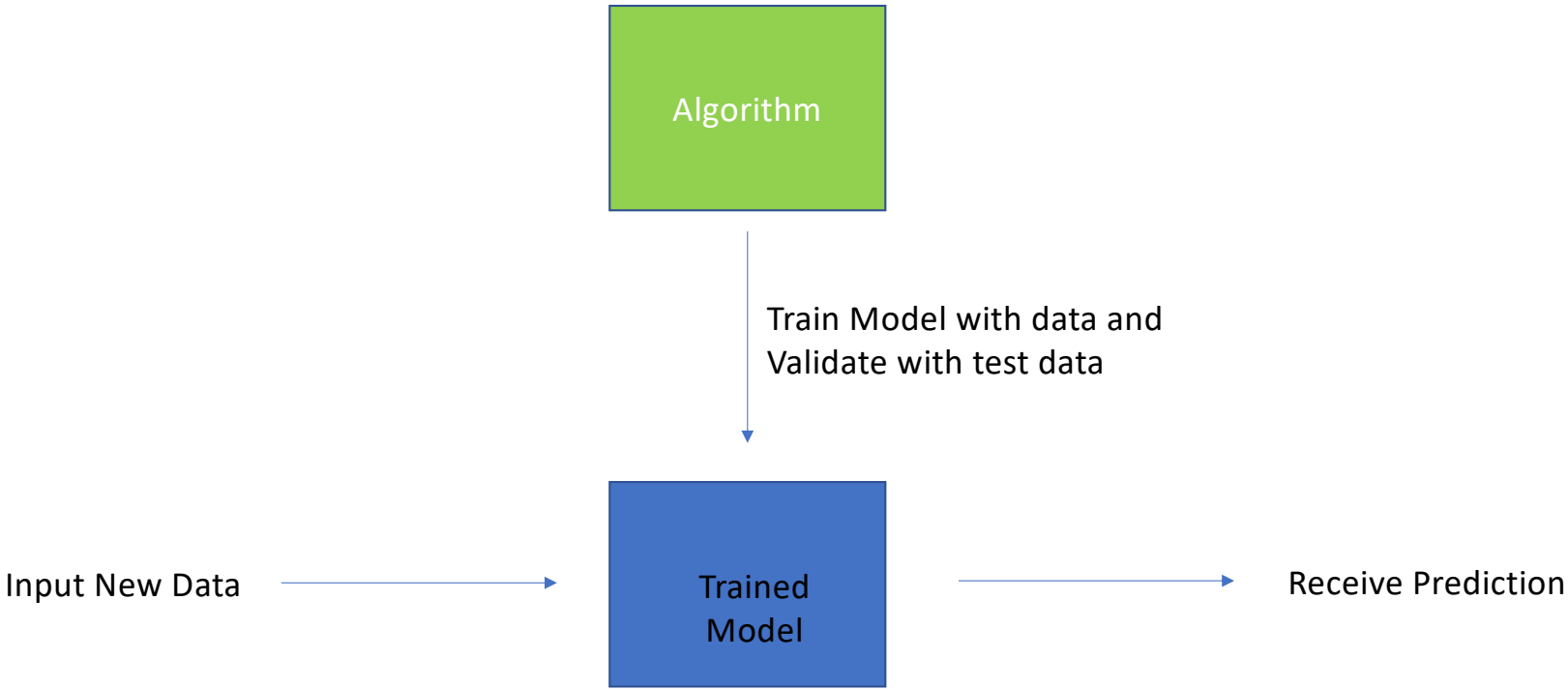


How can regression modelling help pathologists as a clinical diagnostic aid?

- Case Study*:
  - Prostate cancer research study of 97 men recorded many parameters including:
    - lcavol: log of cancer volume.
    - lweight: log of the prostate gland weight
    - age: patient age
    - lbph: log of the amount of BPH
    - **svi: seminal vesical invasion**
    - **lcp :log of capsular penetration**
    - gleason: gleason score
    - pgg45: percentage of tumor with Gleason's score of 4 or 5
    - **lpsa: log of the PSA**

*Stamey TA, Kabalin JN, McNeal JE, Johnstone IM, Freiha F, Redwine EA, Yang N. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. II. Radical prostatectomy treated patients. J Urol. 1989 May;141(5):1076-83. doi: 10.1016/s0022-5347(17)41175-x. PMID: 2468795.

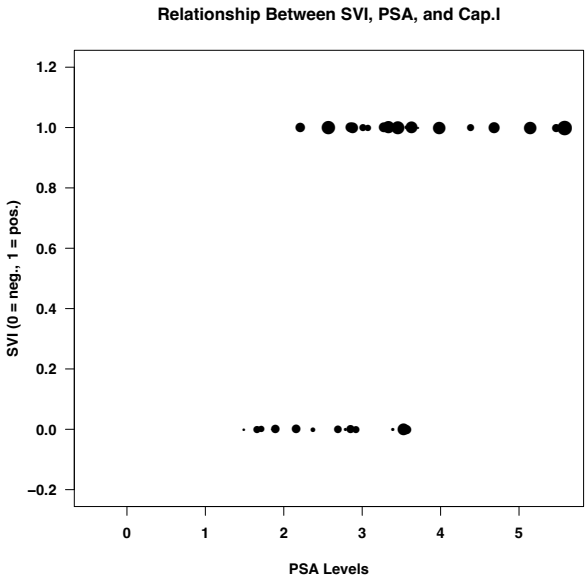
How is the model utilized?



Regression can be used to calculate probability: Logistic regression (provides a probability)

Actual Data

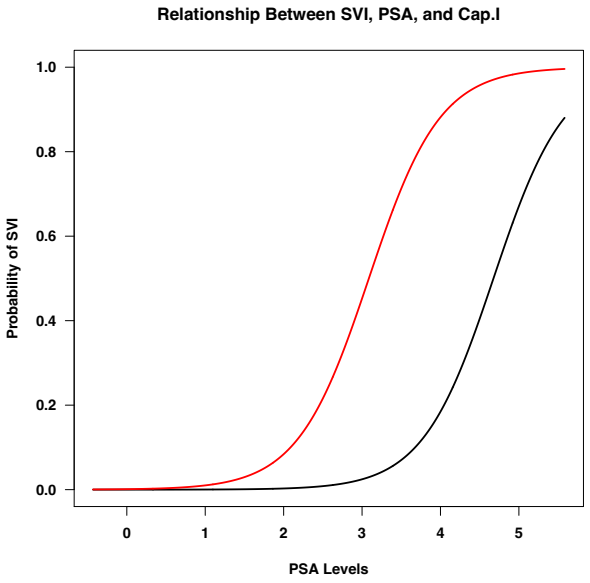
Point Size = extent of Caps. Inv



Model
SVI ~ PSA + caps.

Logistic Regression

Black is low Caps. Inv., Red is high Caps.Inv



Now lets take our models and input new examples-Logistic Regression Model

Patient 1

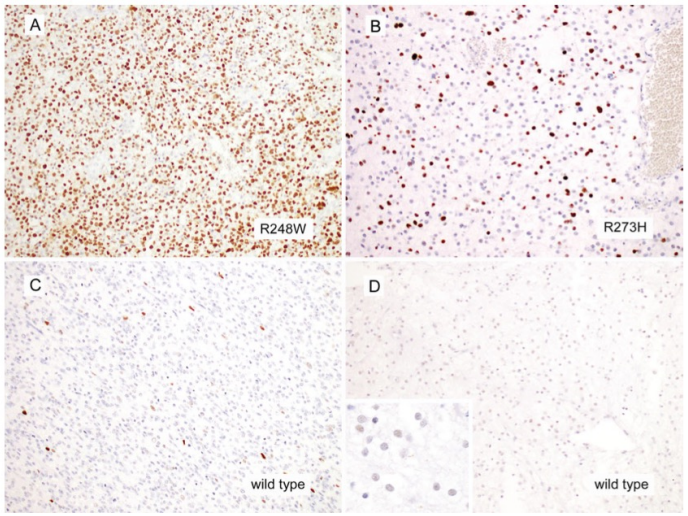
- PSA = 3.2
- No capsular penetration
- **Probability of SVI = prostate is 0.03**

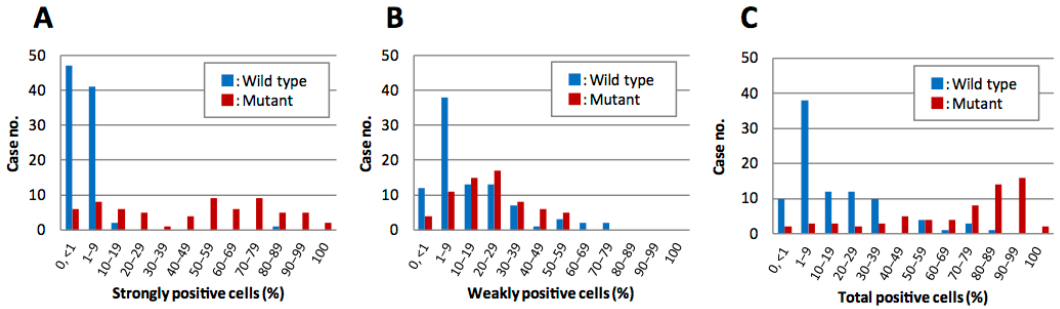
Patient 2

- PSA = 3.2
- Capsular penetration
- **Probability of SVI = prostate is 0.55**

Example of how to combine imaging tools in Pathology for Classification

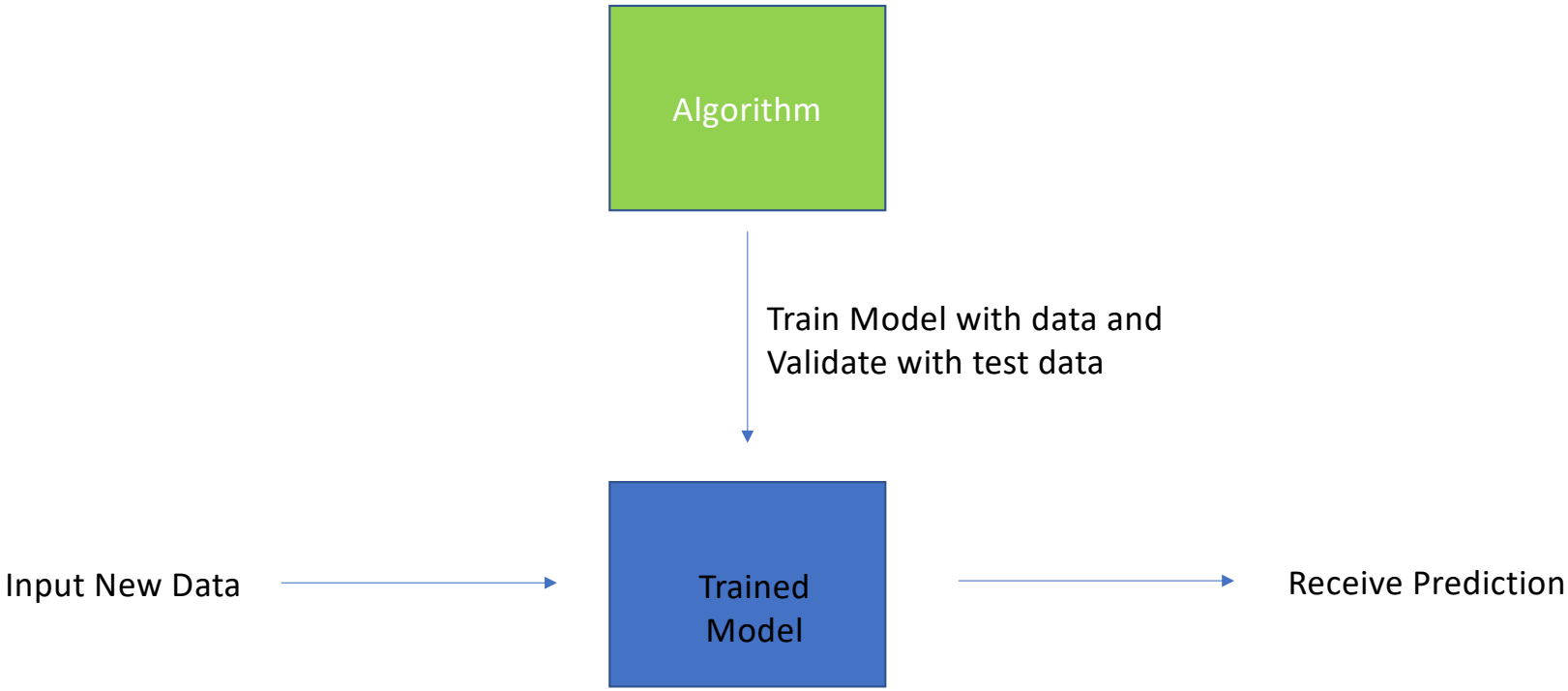
- **Case Study in Classification Machine Learning:**
  - P53 immunohistochemistry is used as a proxy of *TP53* mutation, with intense immunoreactivity in many tumor cells being highly associated with the *TP53* mutations. ***IHC surrogates for molecular findings save time and money.***





Takami, *et al*., *Brain Pathology*, 25 (2015) , 256-266.

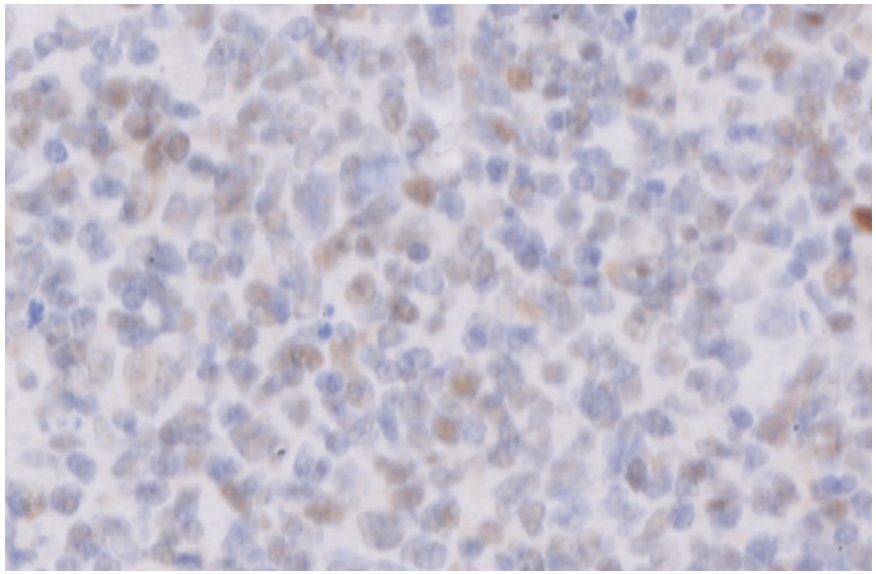
How is the model utilized?



Step 1: Evaluation of Ground Truth

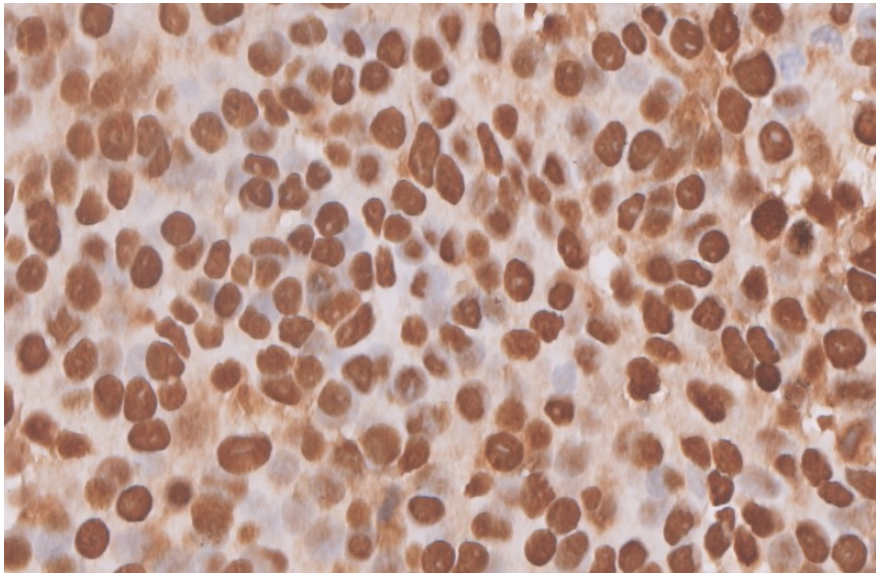
Benign Lymph Node

(known negative TP53 mut)



Ovarian Serous Carcinoma

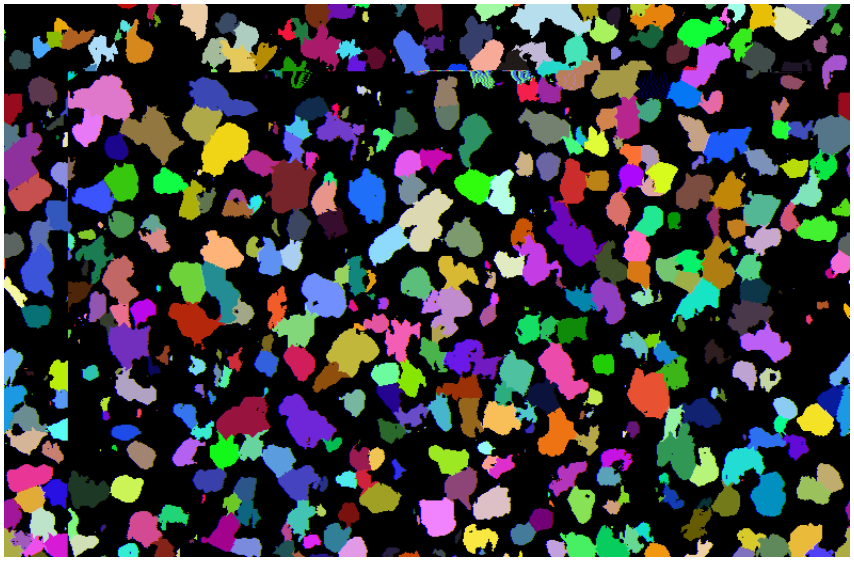
(known positive TP53 mut.)



Step 2: Image Segmentation

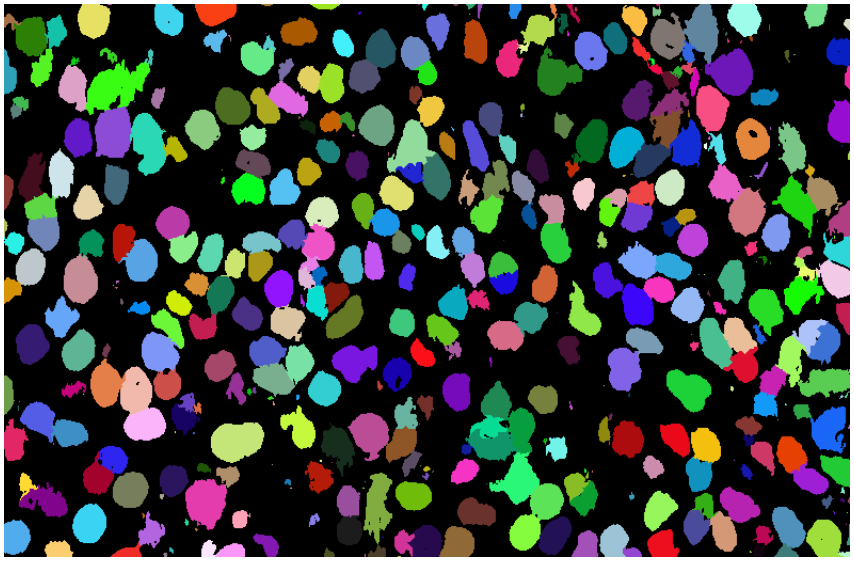
Benign Lymph Node

(known negative TP53 mut)



Ovarian Serous Carcinoma

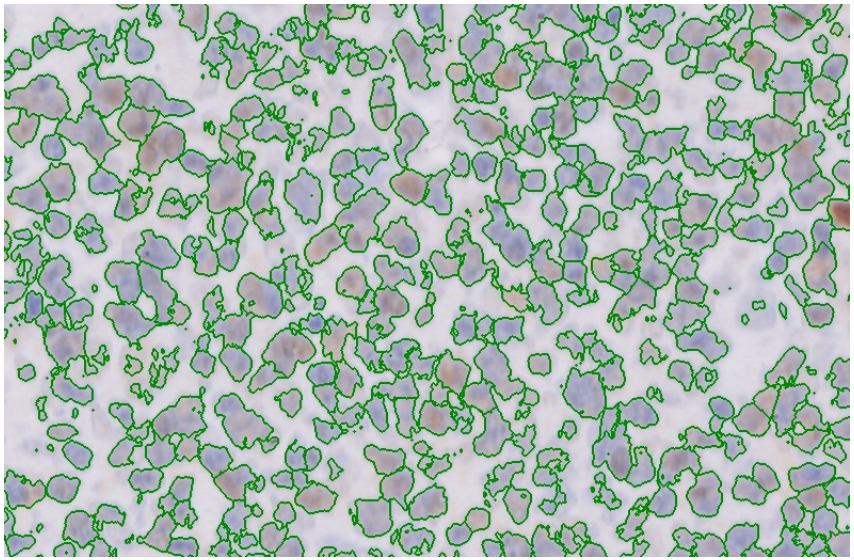
(known positive TP53 mut.)



Step 3: Overlay Mask and Extract Imaging Features

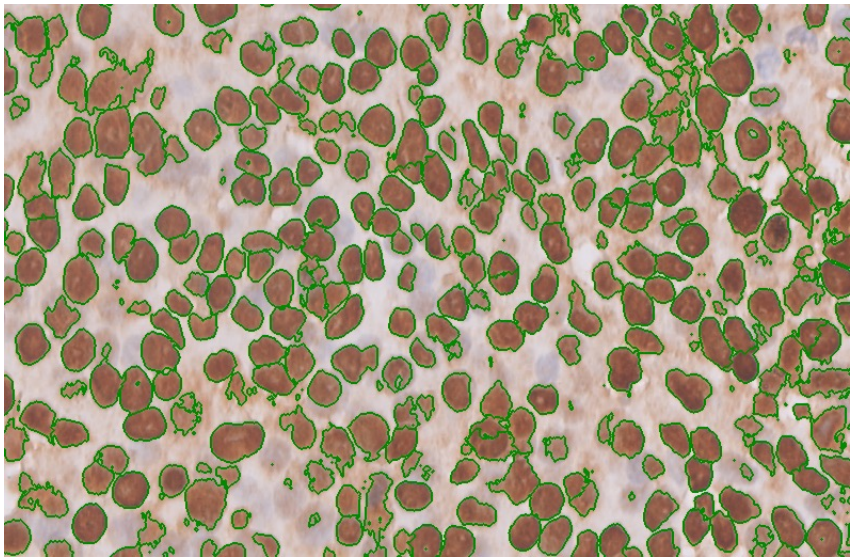
Ovarian Serous Carcinoma

(known positive TP53 mut.)



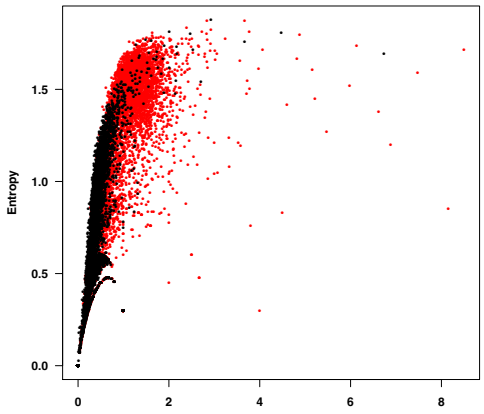
Ovarian Serous Carcinoma

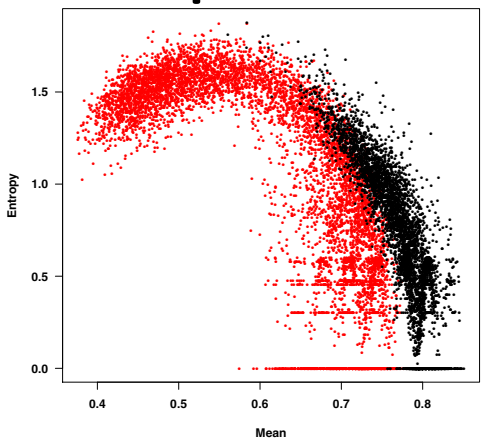
(known positive TP53 mut.)

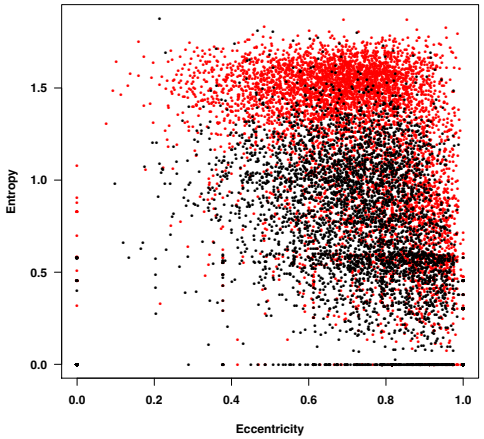


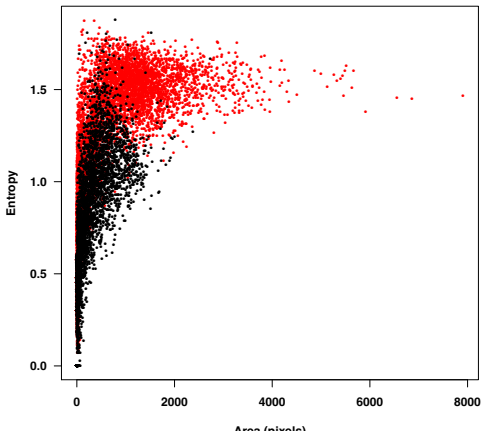
Step 4: Reduce Dimensions (we started with 267 dimensions/features per nucleus)

Red = P53 Positive
Black = P53 Negative





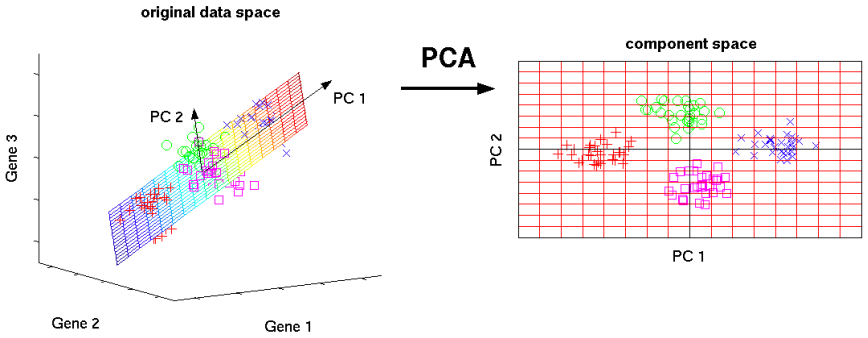




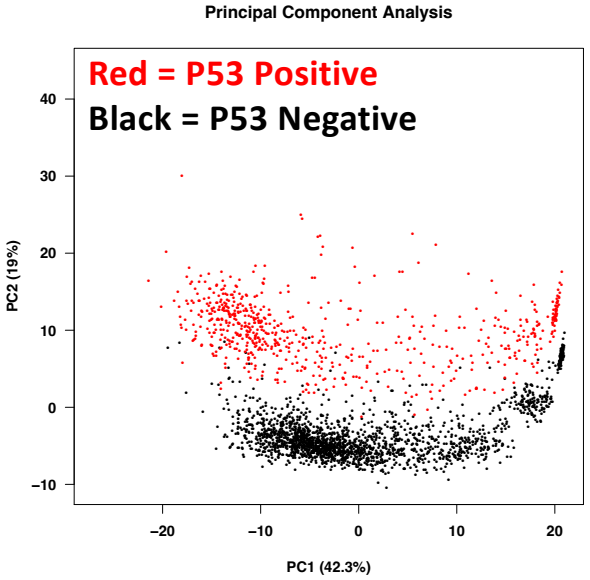
Visualizing so many combinations
Of graphs would make you go crazy as
There are way too many 2x2 combinations!!!!

Step 4: Reduce Dimensions (we started with 267 dimensions/features per nucleus)

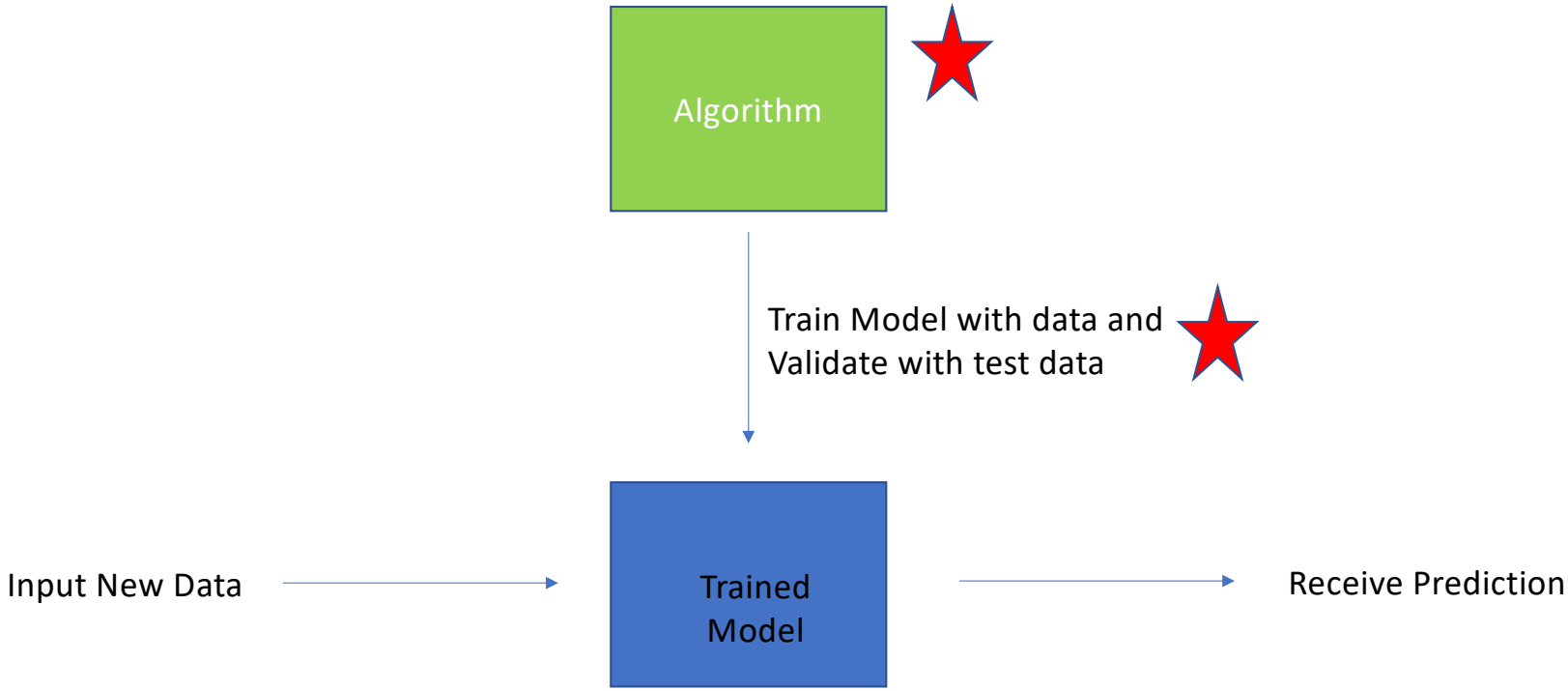
Principal Component Analysis



Easy Separation when using only ~62% of the variance.



How is the model utilized?



Step 5: Train Data and Test Data

- Test algorithm on 30% of remaining data (aka Test data)

Reference

Accuracy : 0.9975

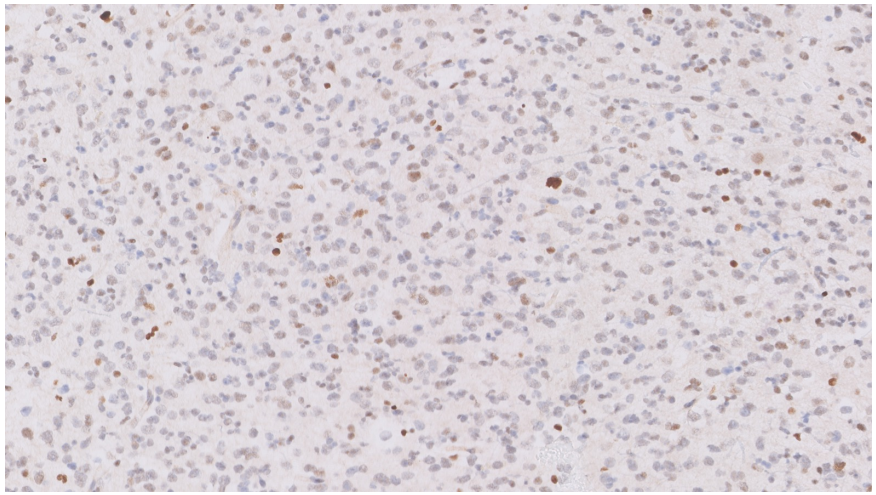
95% CI : (0.991, 0.9997)

No Information Rate : 0.7704

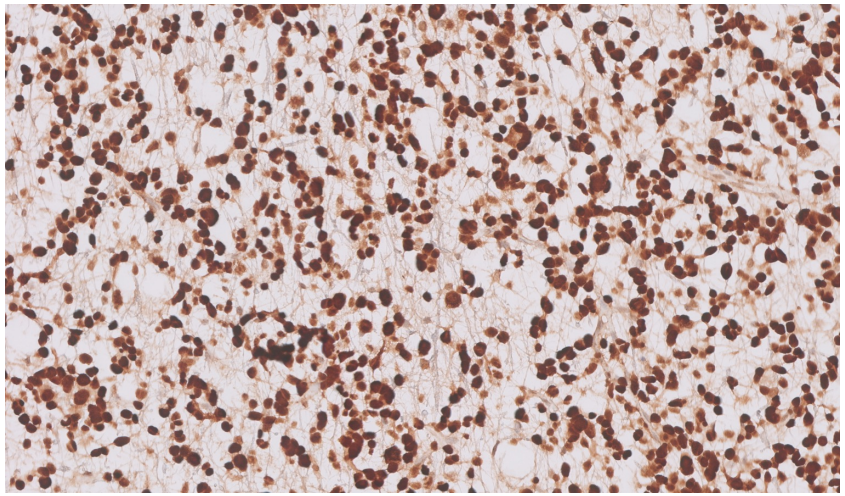
P-Value [Acc > NIR] : <2e-16

Step 6: Deploy on real data (glioma analysis)

Case 1: Low grade glioma p53 stain

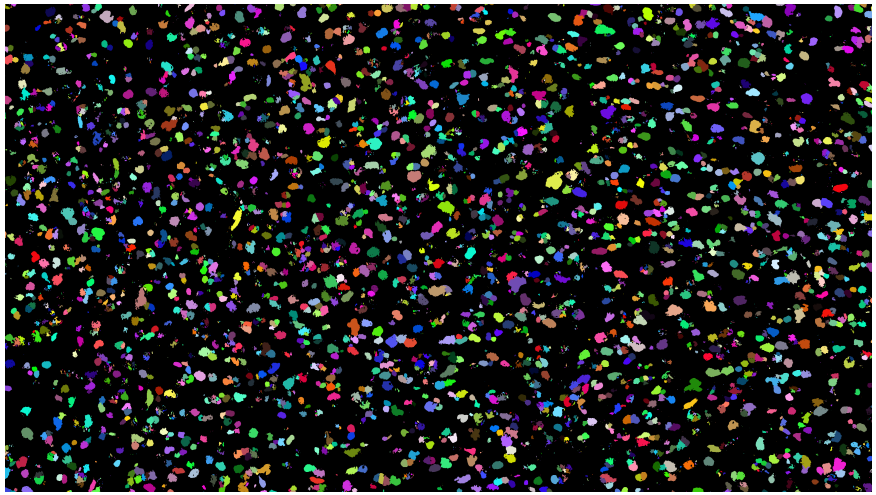


Case 2: Low grade glioma p53 stain

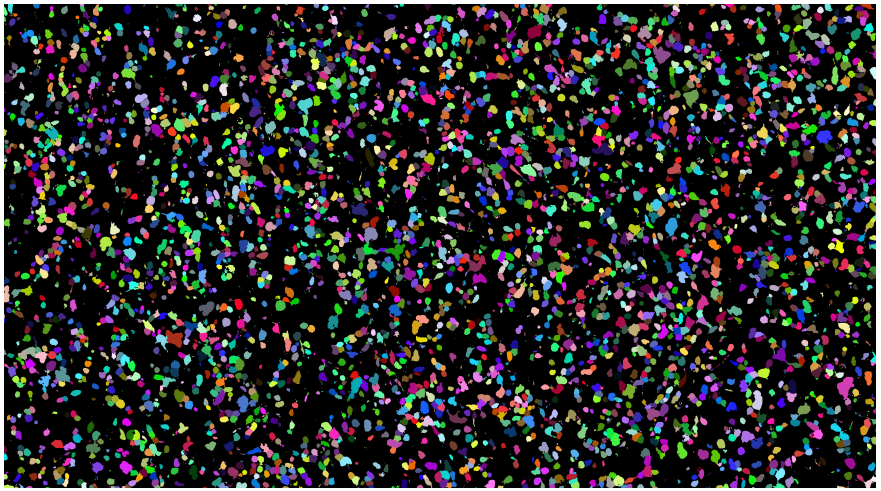


Step 6a:Segment and Extract Imaging Features

Case 1: Low grade glioma p53 stain

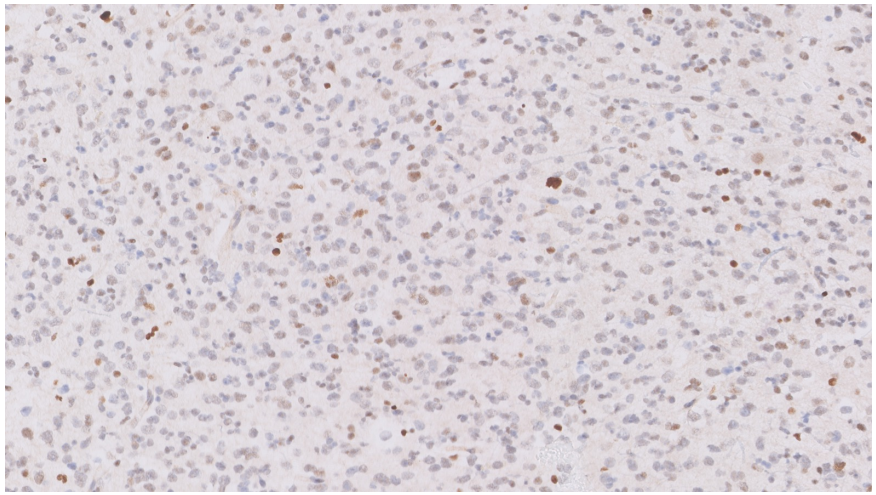


Case 2: Low grade glioma p53 stain

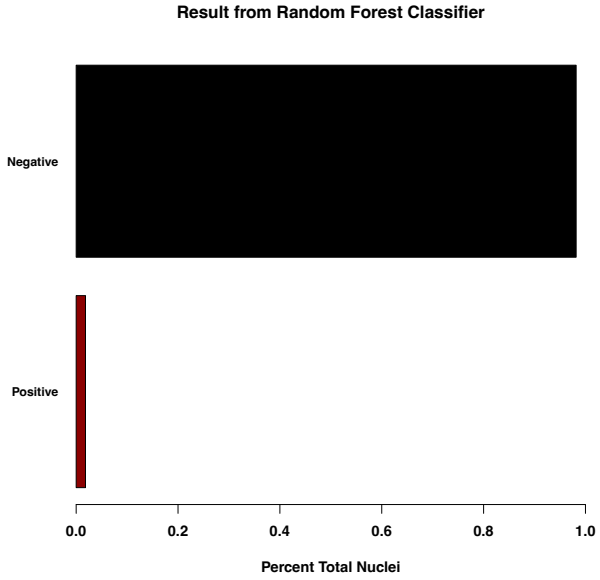


Case 1 has low probability of TP53 mutation

Original Image

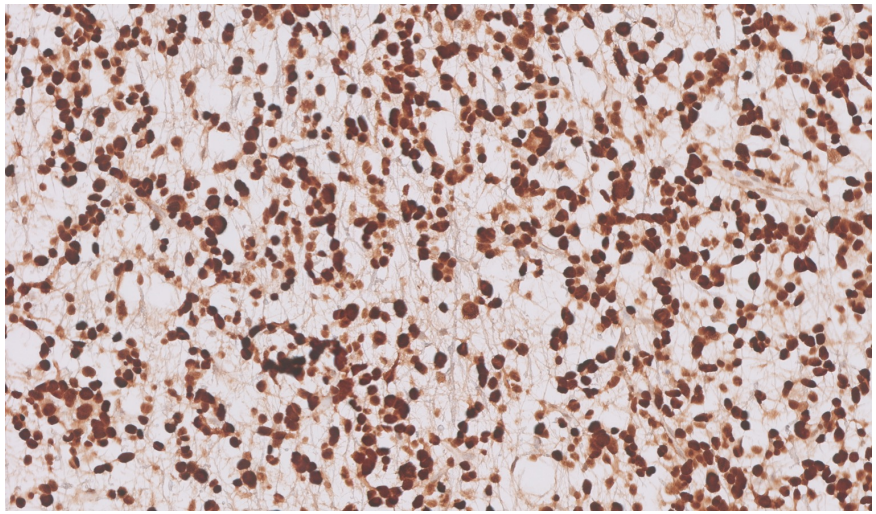


Quantification by Random Forest

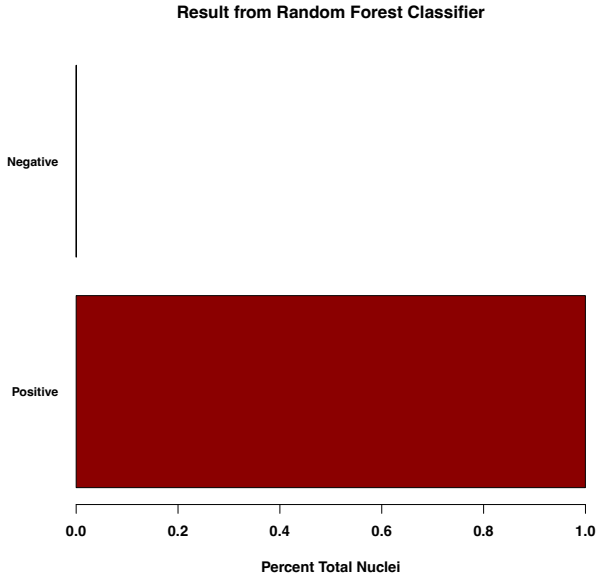


Case 2 has high probability of TP53 mutation

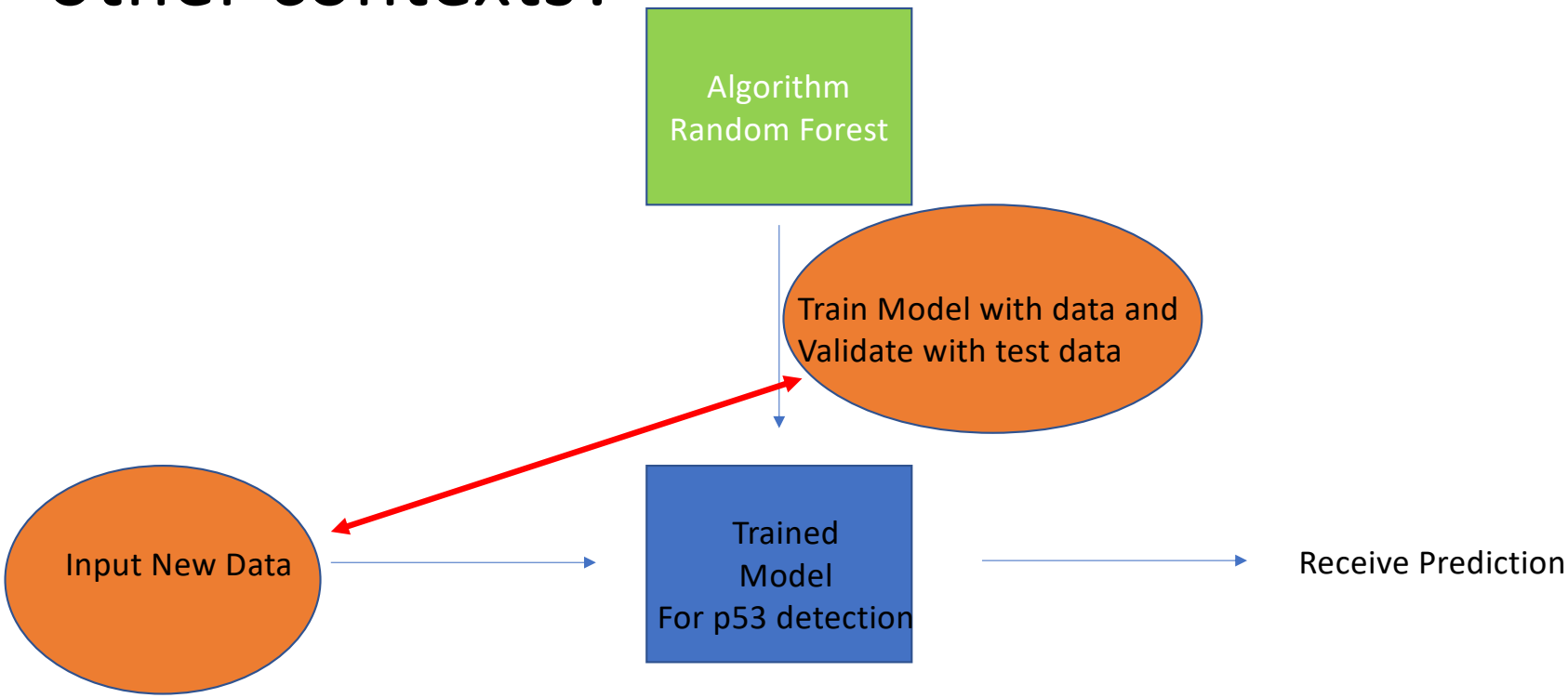
Original Image



Quantification by Random Forest

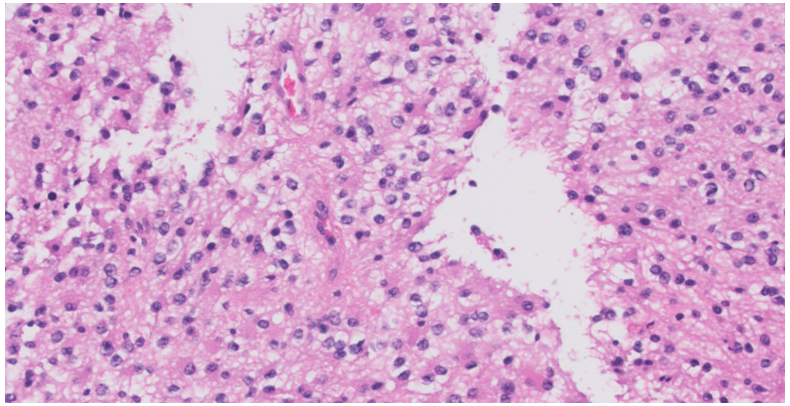


Is it appropriate to re-use this model in other contexts?



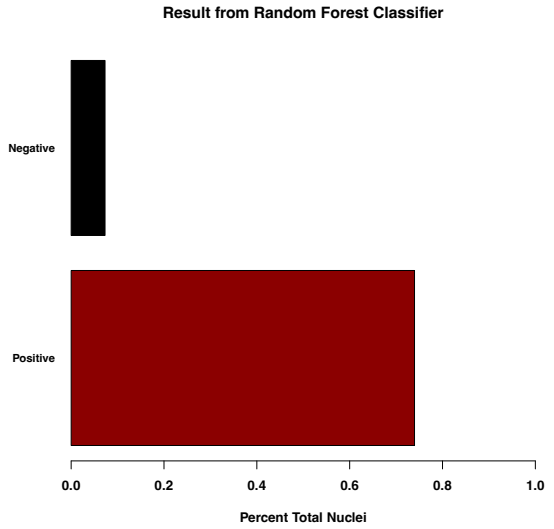
How can this go wrong?

- What if we use the model to predict things for which it was not intended for?
- Eg., how many p53 positive cells are in this image?



Insert data to
The model

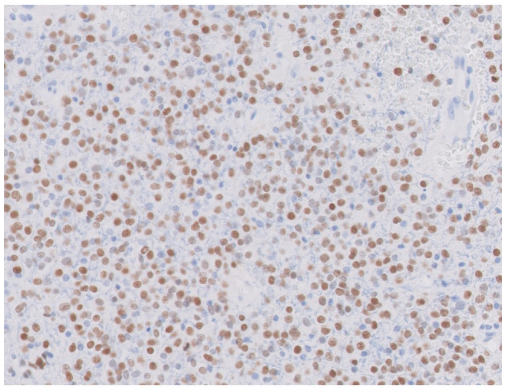
Inserting non-sensical data into
An algorithm causes non-sensical results!!!



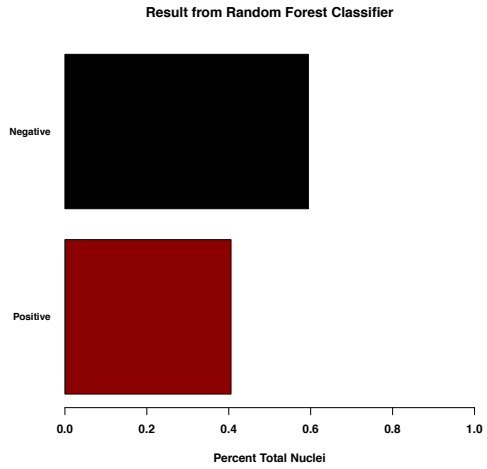
Can we use our classifier to label DAB+ cells?

- As an example, can we determine the percentage of OLIG2+ cells in thisimage?

Inserting non-sensical data into
An algorithm causes non-sensical results!!!



Insert data to
The model



Using Trained ML Models on Classification Problems for which They Were Not Intended for is Very Problematic

“Negative” Nuclei

“Positive” Nuclei

Conclusion: When using and designing Machine Learning tools for pathology, you must understand the purpose of the classifier.

What is a neural network?

- Pathologists will have to become joined with computer scientists similarly to how radiologists are joined with physicists.
  - Similar to our merger with molecular biology that started ~ 20 years ago.
- Nodes:
  - Input Nodes/Neurons
  - Hidden Nodes/Neurons
  - Output Nodes/Neurons
- Transition from numerical data into the input node is modified by
  - Weight (called a coefficient in other machine learning systems)
  - Bias (a constant)
  - Activation function (a non-linear transformation of the data)
  - Although weight and bias is linear, placing the transformation into an activation function turns this to a non-linear relationship.

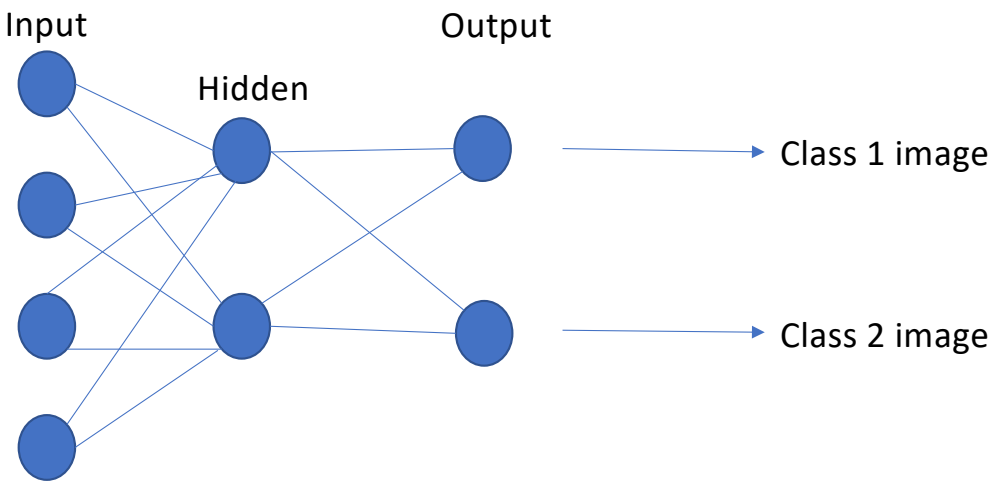
How would a neural network classify?

Class 1 image

1	**0.5**
0.5	**1**

Class 2 image

1	**0**
0.5	0.25



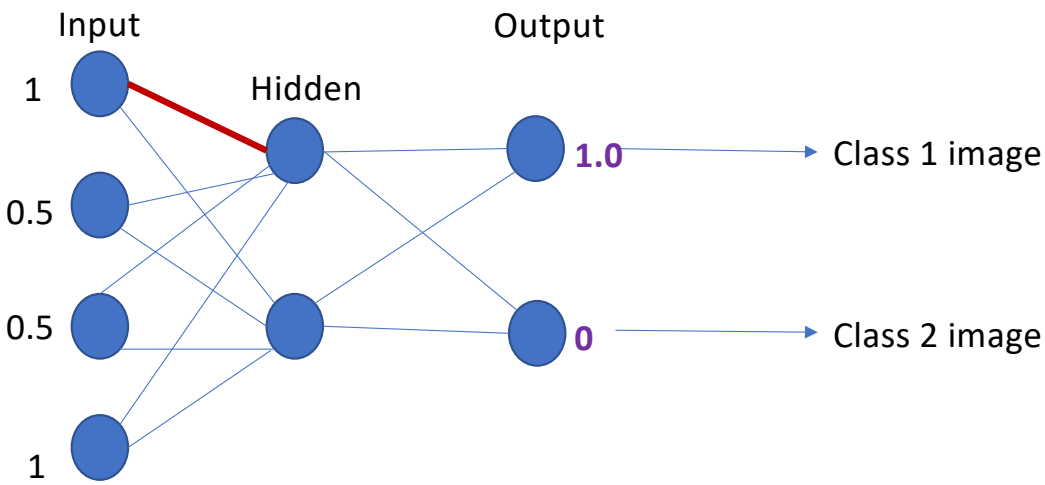
How would a neural network classify?

Class 1 image

1	**0.5**
0.5	**1**

Class 2 image

1	**0**
0.5	0.25



The value of the top node will become:
x = (Input*weight + bias)
Value for next layer will be the value of the activation function.
Eg., softplus function is defined as $f(x) = \ln(1+ e^x)$

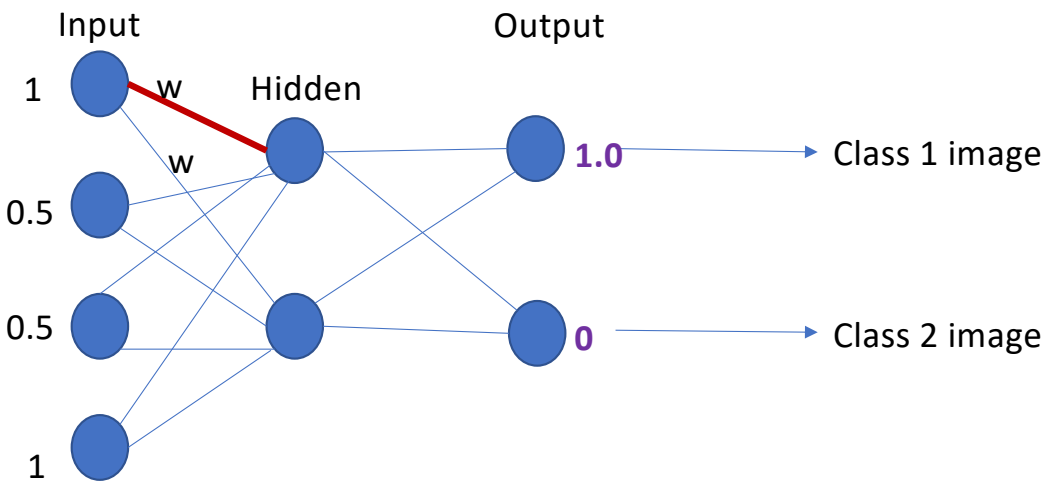
How would a neural network classify?

Class 1 image

1	**0.5**
0.5	**1**

Class 2 image

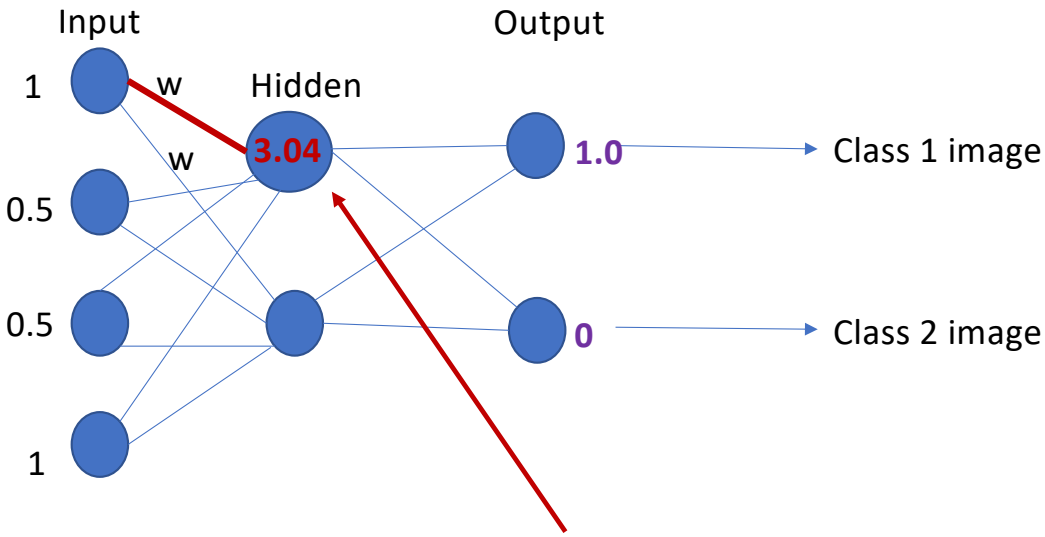
1	**0**
0.5	0.25



The value of the top node will become:
x = (Input*weight + bias)
Value for next layer will be the value of the activation function.
Eg., softplus function is defined as $f(x) = \ln(1+ e^x)$

Eg., weight = 1, bias = 2
x = 1*1 + 2 = 3
$f(3) = \ln(1+e^3) = 3.04$

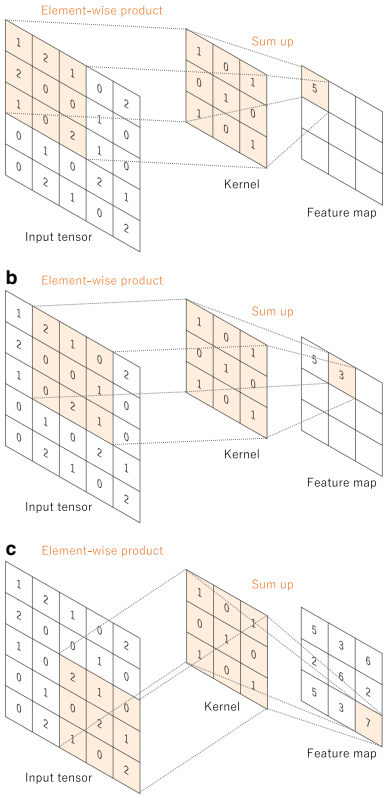
How would a neural network classify?

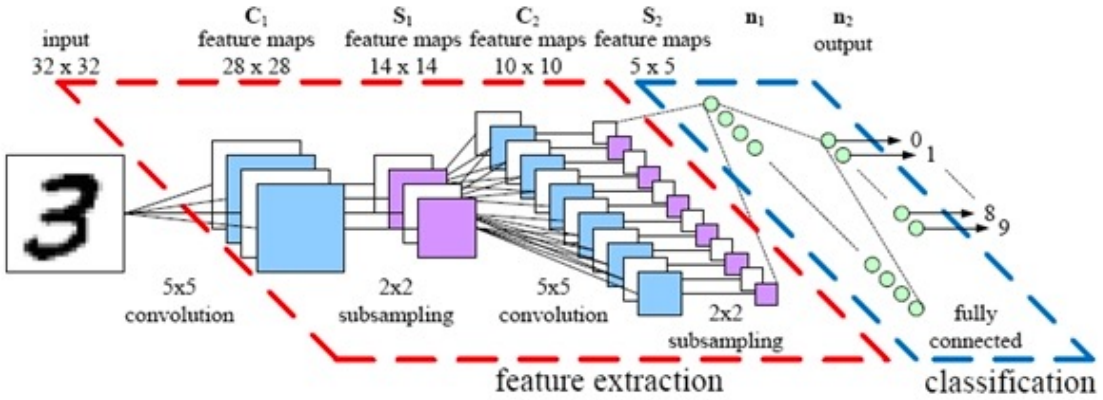


The value of the top node will become:
x = (Input*weight + bias)
Value for next layer will be the value of the activation function.
Eg., softplus function is defined as $f(x) = \ln(1+ e^x)$

Eg., weight = 1, bias = 2
x = 1*1 + 2 = 3
$f(3) = \ln(1+e^3) = 3.04$

Overview of Convolutional Neural Networks





Neural network models work well on large datasets and are fast.

Yamashita, et al., Insights into imaging 2018:9:611-629

Closing Remarks: Such an Exciting Time to Be A Pathologist!!!!

- Digital pathology is converting our histpathology slides into files with massive amounts of numerical data.
- These data are highly amenable to utilizing machine learning tools.
- Pathologists need to learn now how and when to use machine learning and artificial intelligence tools to improve patient care.