





Objective

- Prediction of 5 year cancer outcomes from histology images
- Estimate genomic characteristics of a tumour
 - gene expression data
 - molecular subtype
 - proliferation rate
 - oncogenic pathway activation
 - genomic instability



Buzzwords

- Machine Learning
- Artificial Intelligence
- Deep Learning
- Keras, Tensorflow
- CNN
- Kaggle
- Big Data

- NAÏVE
 - That's me

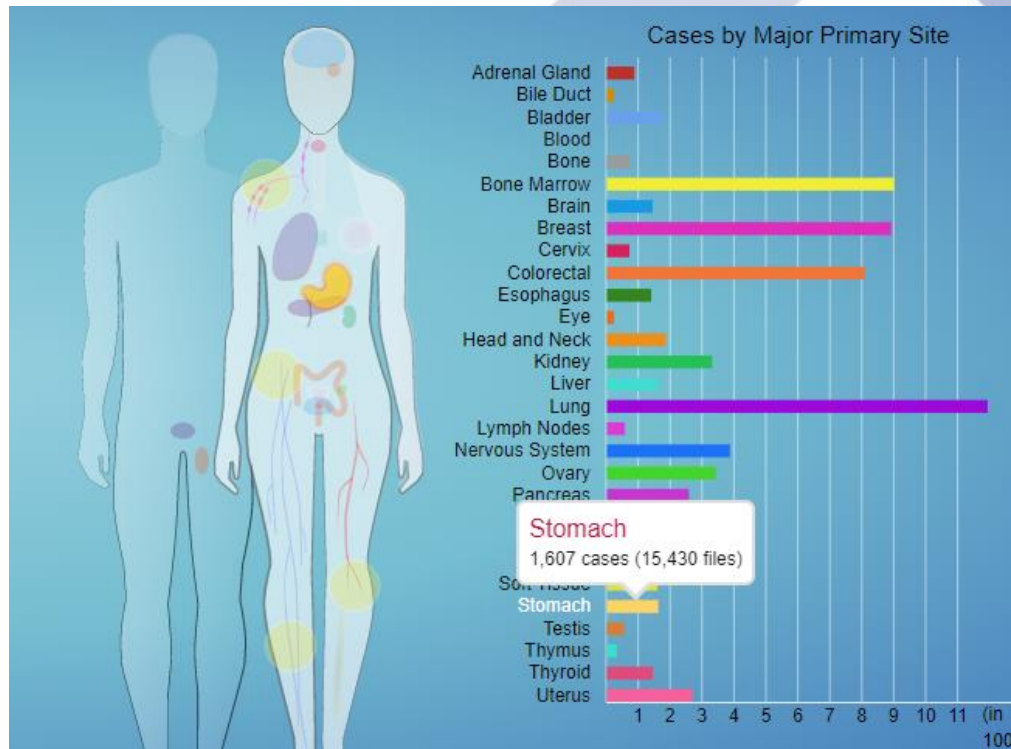




NATIONAL CANCER INSTITUTE GDC Data Portal

<https://portal.gdc.cancer.gov/>

- The Cancer Genome Atlas
- TCGA





NATIONAL CANCER INSTITUTE GDC Data Portal

<https://portal.gdc.cancer.gov/>

Cases and File Counts by Data Category

Data Category	Cases (n=443)	Files (n=12,867)
Sequencing Reads	443	1,845
Transcriptome Profiling	439	2,203
Simple Nucleotide Variation	441	3,536
Copy Number Variation	443	1,813
DNA Methylation	443	470
Clinical	443	463
Biospecimen	443	2,537

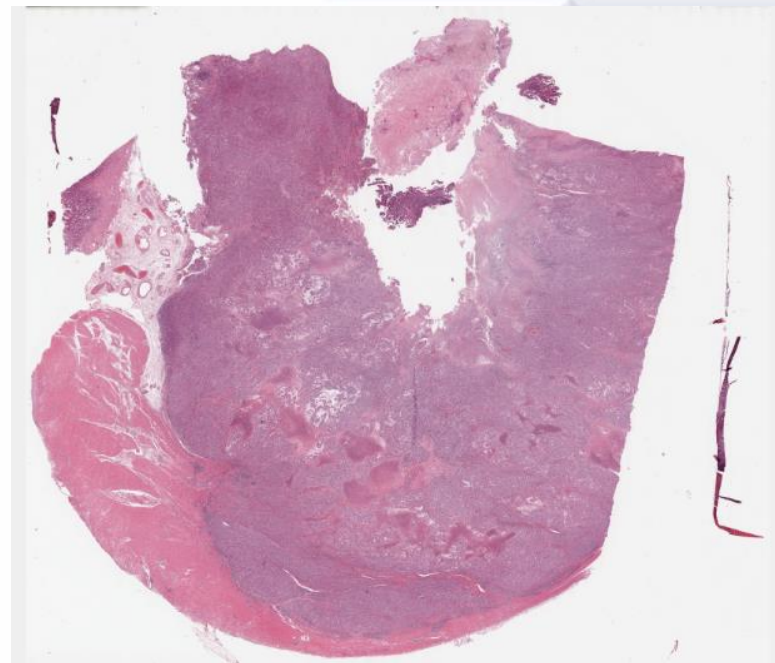
Cases and File Counts by Experimental Strategy

Experimental Strategy	Cases (n=443)	Files (n=12,867)
Diagnostic Slide	416	442
Tissue Slide	432	755
WXS	443	4,462
RNA-Seq	380	1,628
miRNA-Seq	436	1,473
ATAC-Seq	21	21
Genotyping Array	443	1,813
Methylation Array	443	470



NATIONAL CANCER INSTITUTE GDC Data Portal

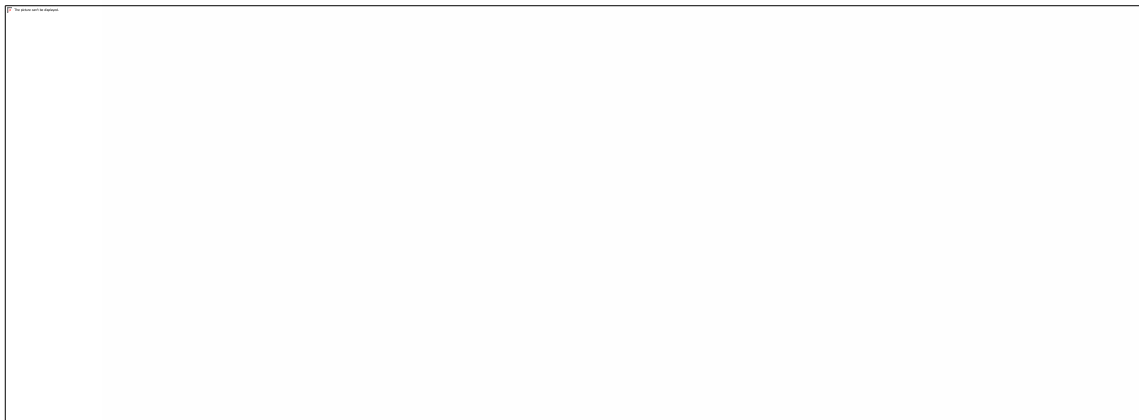
<https://portal.gdc.cancer.gov/>

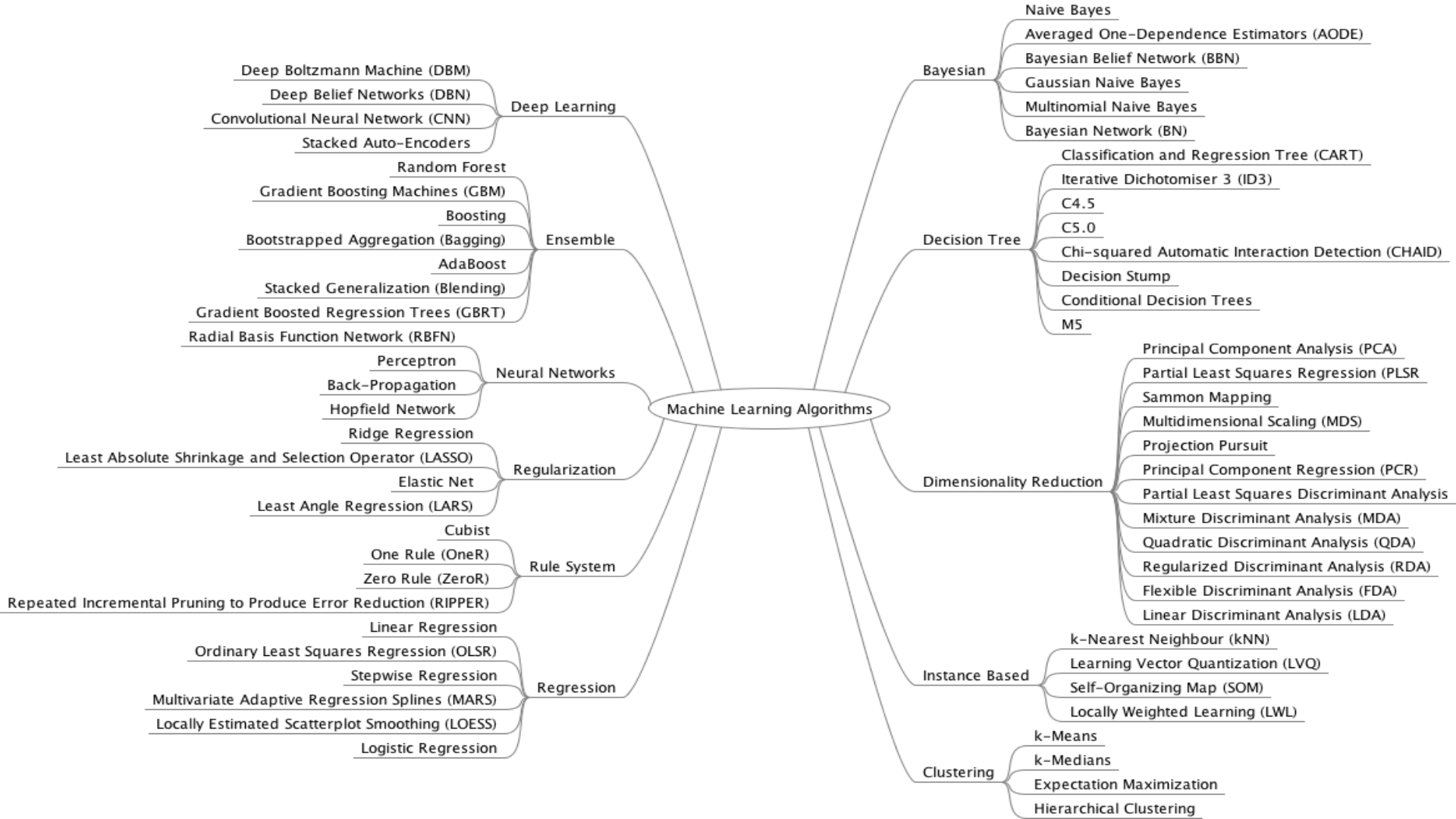




NATIONAL CANCER INSTITUTE GDC Data Portal

- Fischer, W., Moudgalya, S., Cohn, J. et al.
- Sparse coding of pathology slides compared to transfer learning with deep neural networks.
- BMC Bioinformatics 19, 489 (2018). <https://doi.org/10.1186/s12859-018-2504-8>





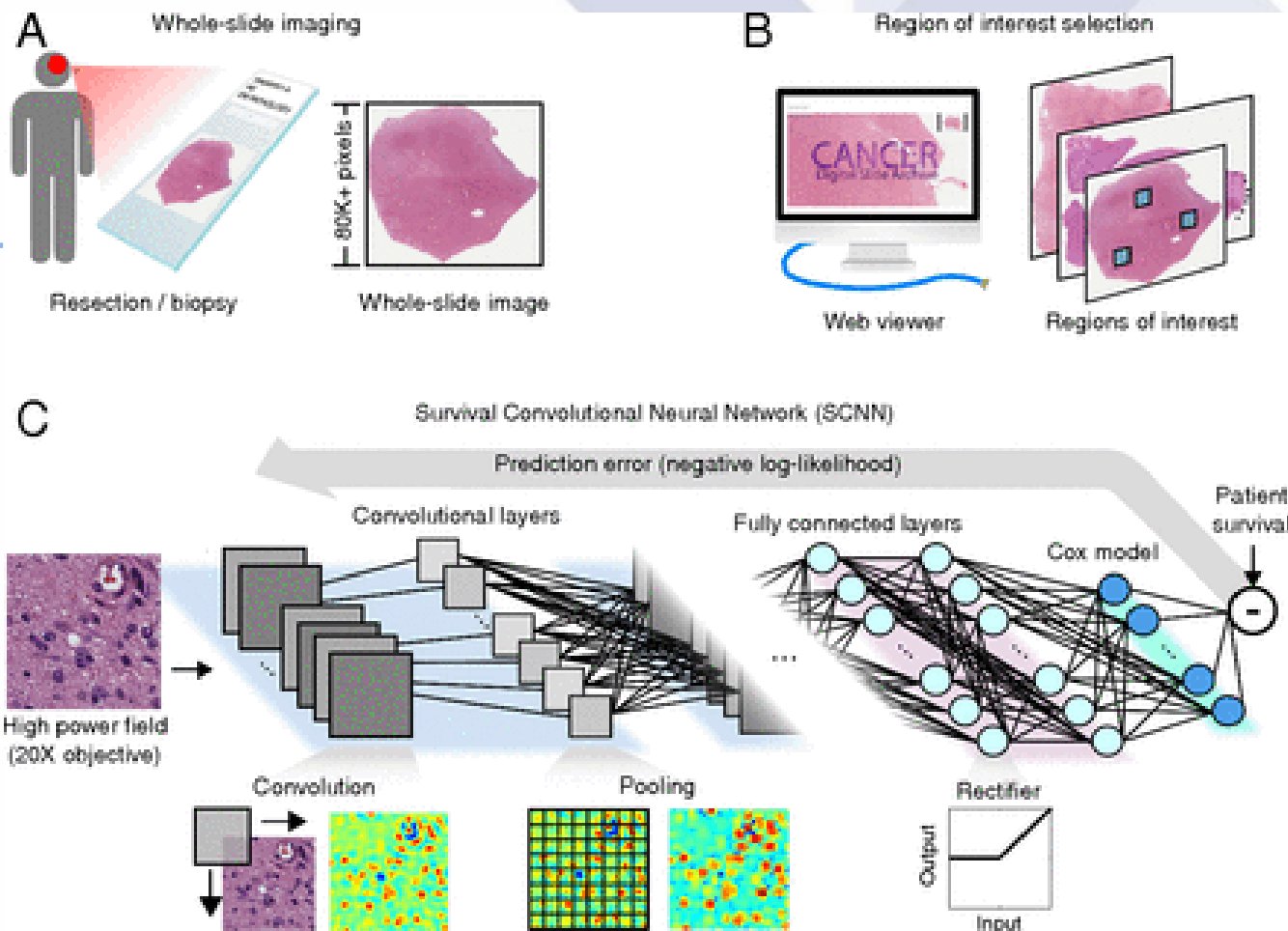
SCNN

<https://github.com/CancerDataScience/SCNN>

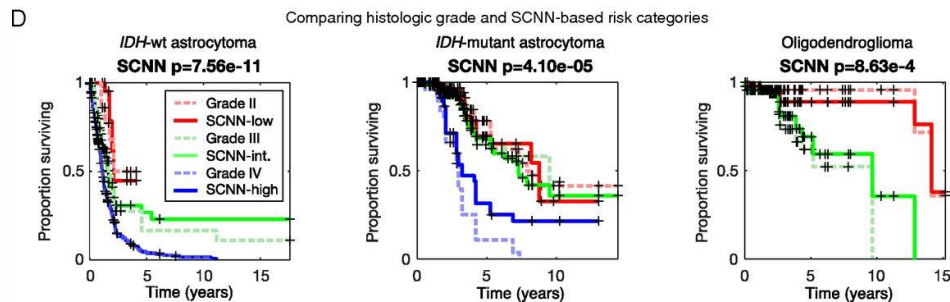
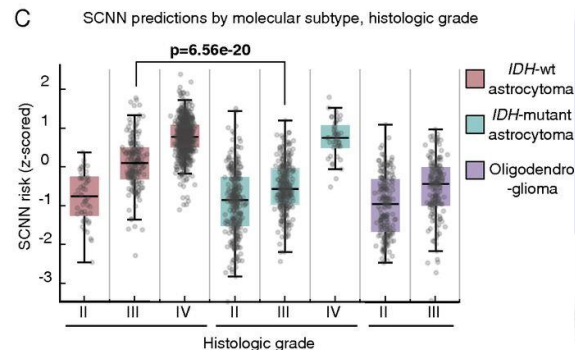
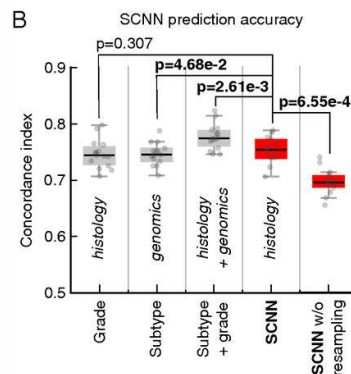
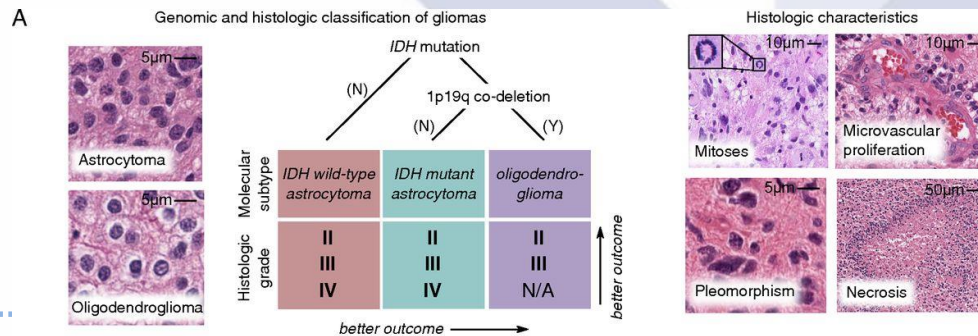
- Survival Convolutional Neural Network
- Docker Container
- User friendliness???
- [Mobadersany, et al. "Predicting cancer outcomes from histology and genomics using convolutional networks" *PNAS* published online March 12, 2018 ahead of print.](#)



SCNN



SCNN



Keras - R

<https://github.com/keras-team/keras>

- User friendliness
- Modularity
- Easy extensibility
- Work with Python

model definition

```
model <- keras_model_sequential()
```

```
model %>% layer_dense(units = 256, activation = 'relu',  
  input_shape = c(784)) %>%  
  layer_dropout(rate = 0.4) %>%  
  layer_dense(units = 128, activation = 'relu') %>%  
  layer_dropout(rate = 0.3) %>%  
  layer_dense(units = 10, activation = 'softmax')
```

```
model %>% compile(loss = 'categorical_crossentropy',  
  optimizer = optimizer_rmsprop(), metrics = c('accuracy'))
```

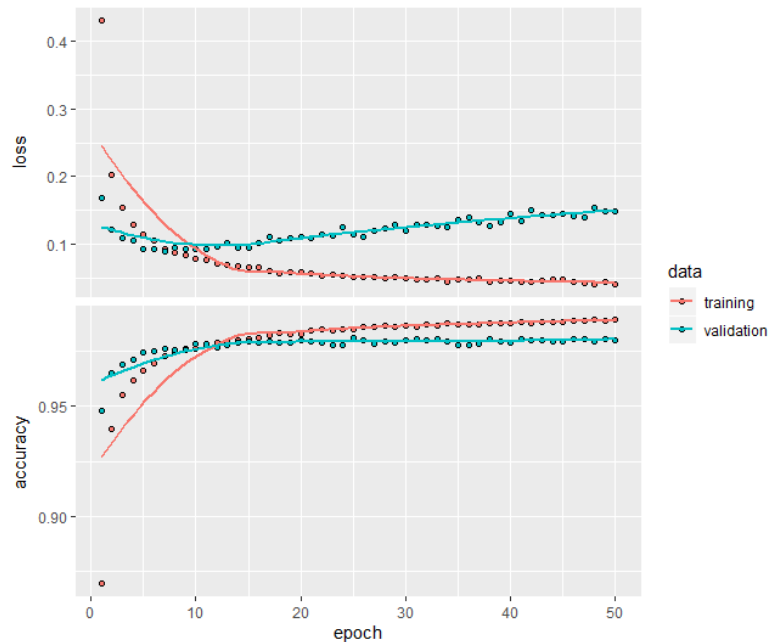
training and validation

```
history <- model %>% fit(x_train, y_train,  
  epochs = 50, batch_size = 128, validation_split = 0.2)
```

Keras - R

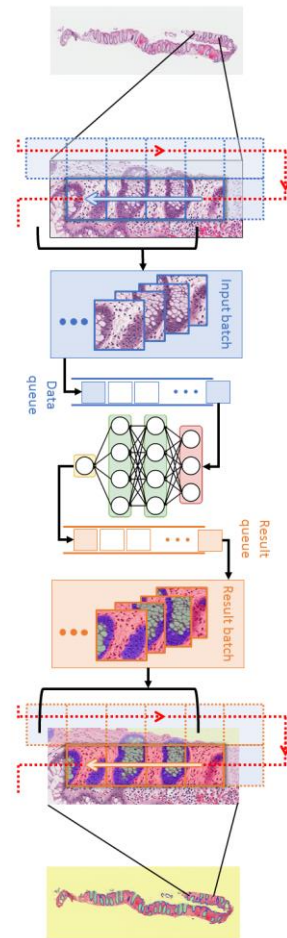
<https://github.com/keras-team/keras>

- User friendliness
 - Modularity
 - Easy extensibility
 - Work with Python
-
- Test Models and Example Data 😊
 - Real Data ☹️



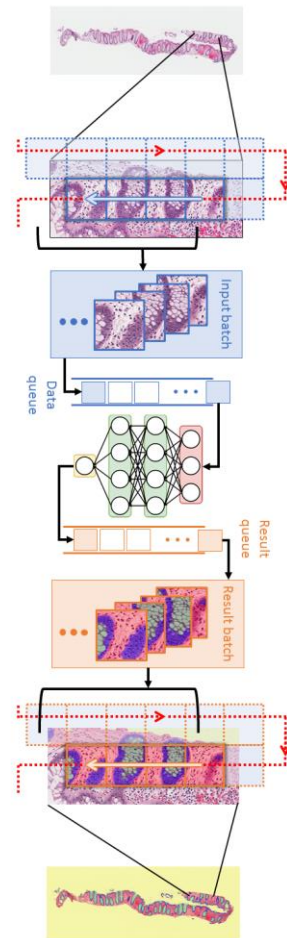
Python

- <https://openslide.org/api/python/>
- <https://github.com/quolc/keras-OpenSlideGenerator>
- <https://github.com/zhaoxuanma/Deeplearning-digital-pathology>
 - Segmentation
- <https://ysbecca.github.io/programming/2018/05/22/py-wsi.html>
 - Patch Sampling for Deep Learning
- AND
- <https://github.com/CancerDataScience/SCNN>



Python

- <https://openslide.org/api/python/>
- <https://github.com/quolc/keras-OpenSlideGenerator>
- <https://github.com/zhaoxuanma/Deeplearning-digital-pathology>
 - Segmentation
- <https://ysbecca.github.io/programming/2018/05/22/py-wsi.html>
 - Patch Sampling for Deep Learning
- AND
- <https://github.com/CancerDataScience/SCNN>



Python/BASH

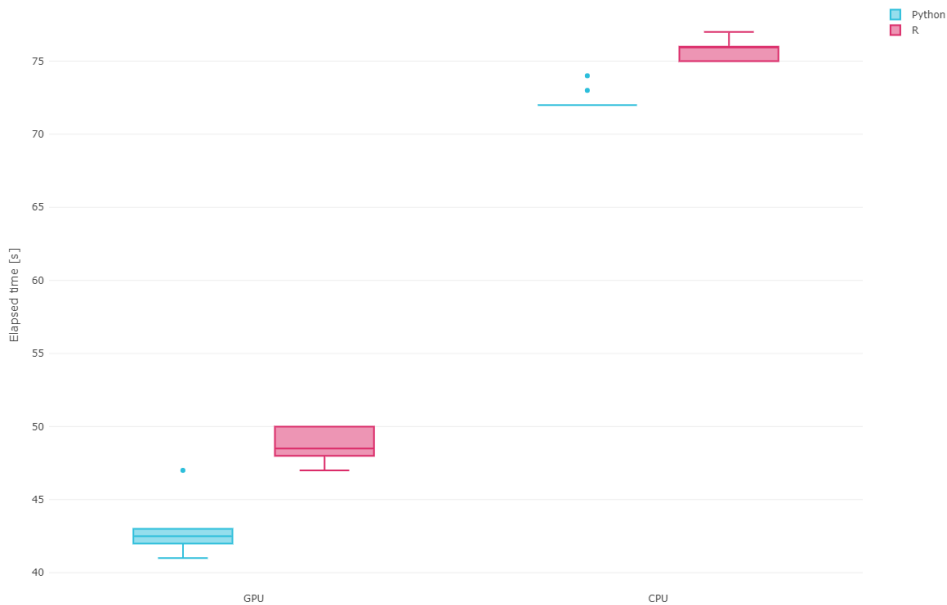
```
cut -d$'\t' -f 2 gdc_manifest.txt | grep -E '\.*-DX[^\-]\w*.'
```

```
python generate_rois.py
```

Inside the Docker container and I can't read it ☹

Python v R

Elapsed Time for 10 Epochs of CNN Training



Final Accuracy after 10 Epochs of CNN Training

