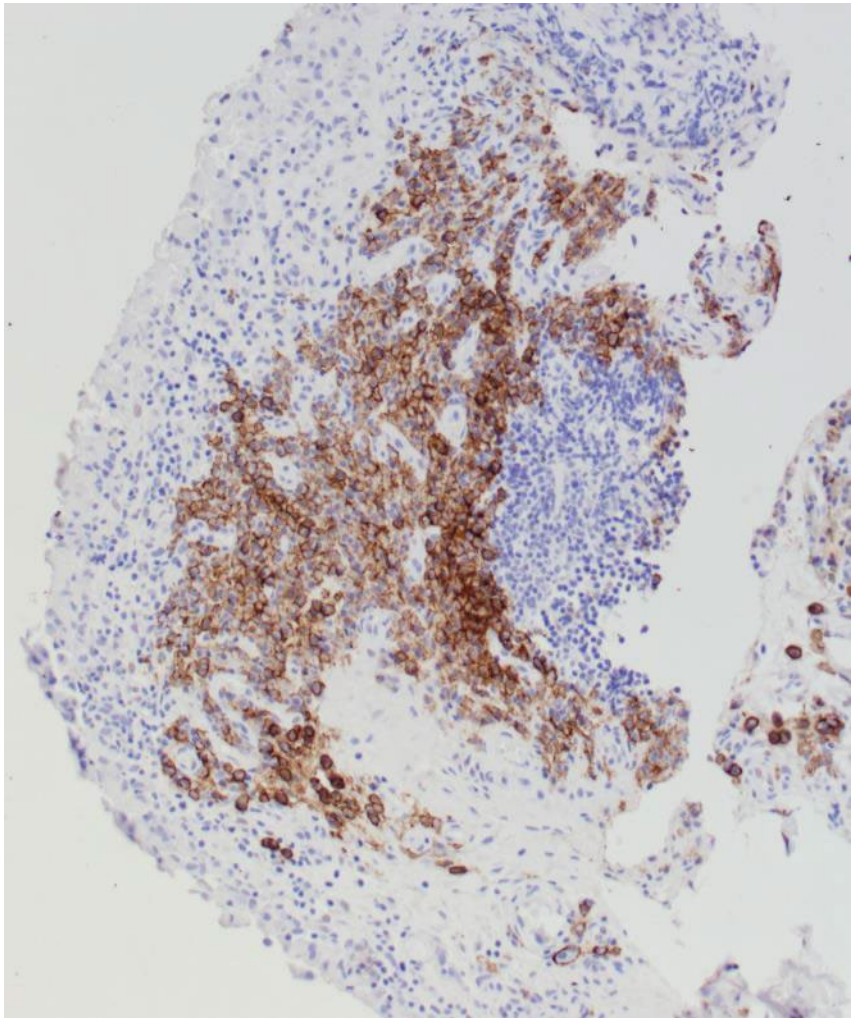
Deep learning applied to synovial histopathology images of Rheumatoid Arthritis patients

Amaya Gallagher-Syed

Research group: Centre for Translational Bioinformatics, William Harvey Research Institute

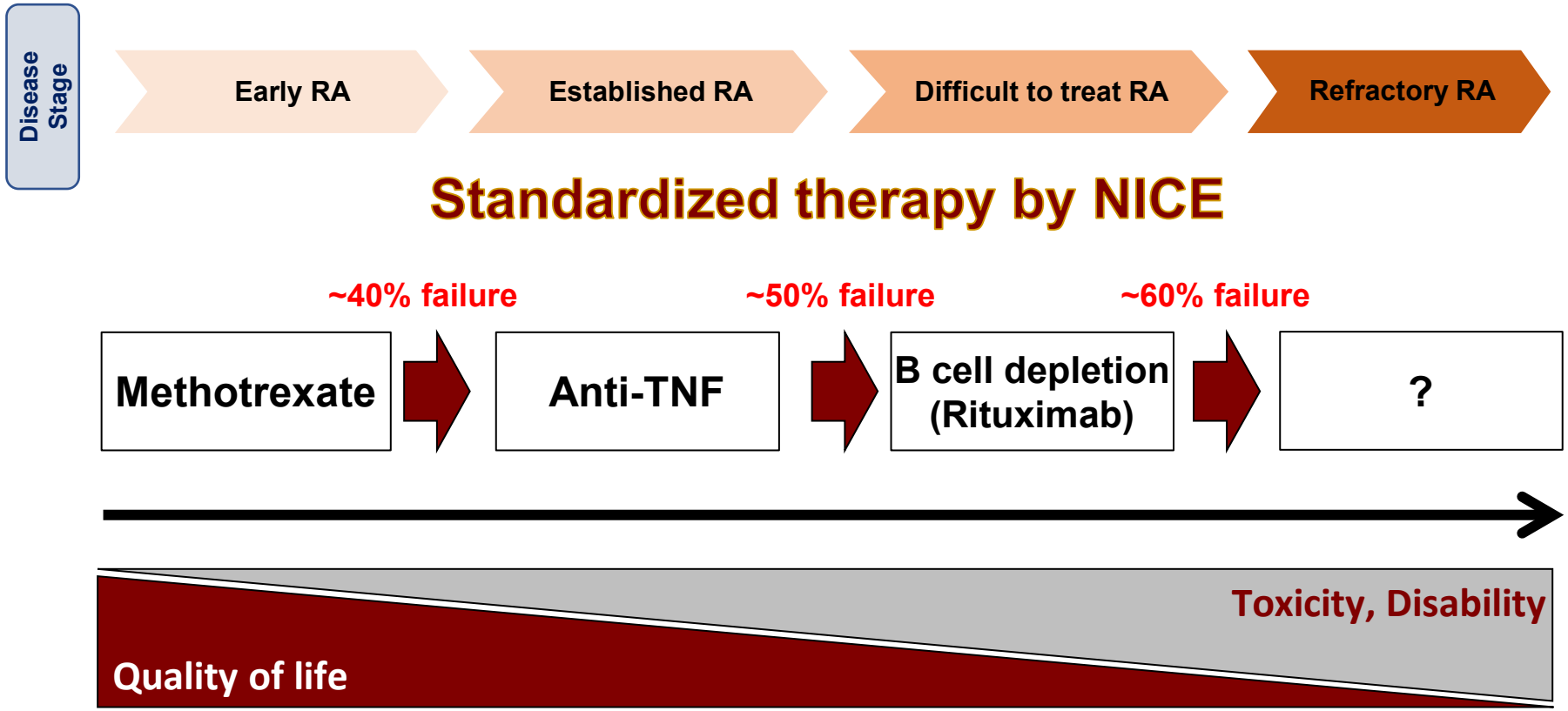
Supervisors: Prof. Mike Barnes – Dr. Myles Lewis



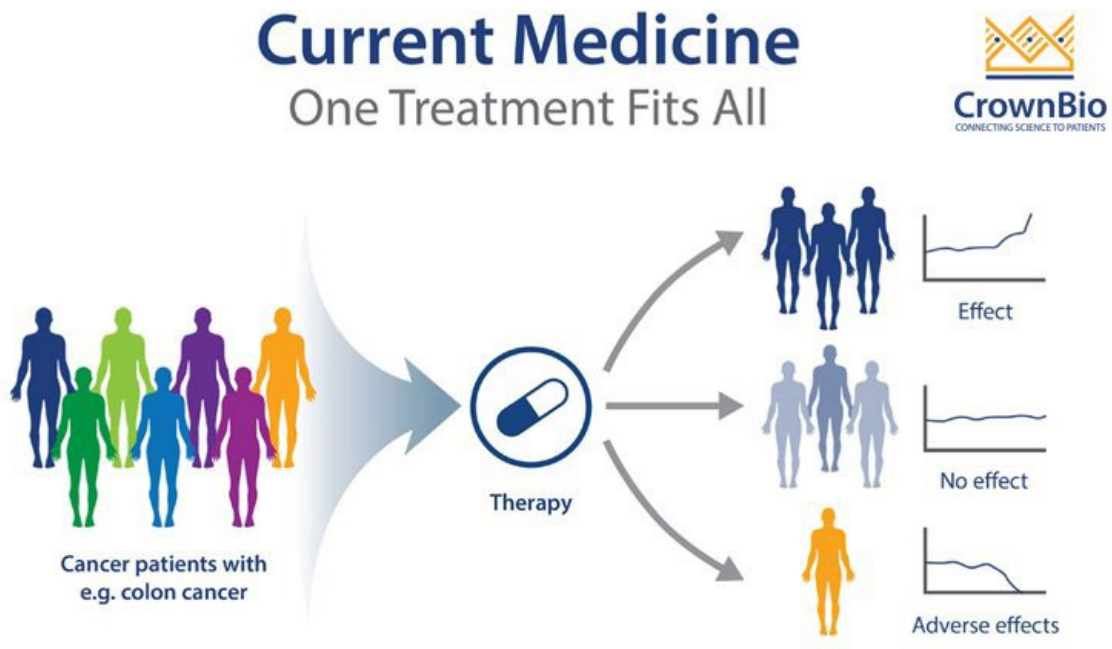
Rheumatoid arthritis

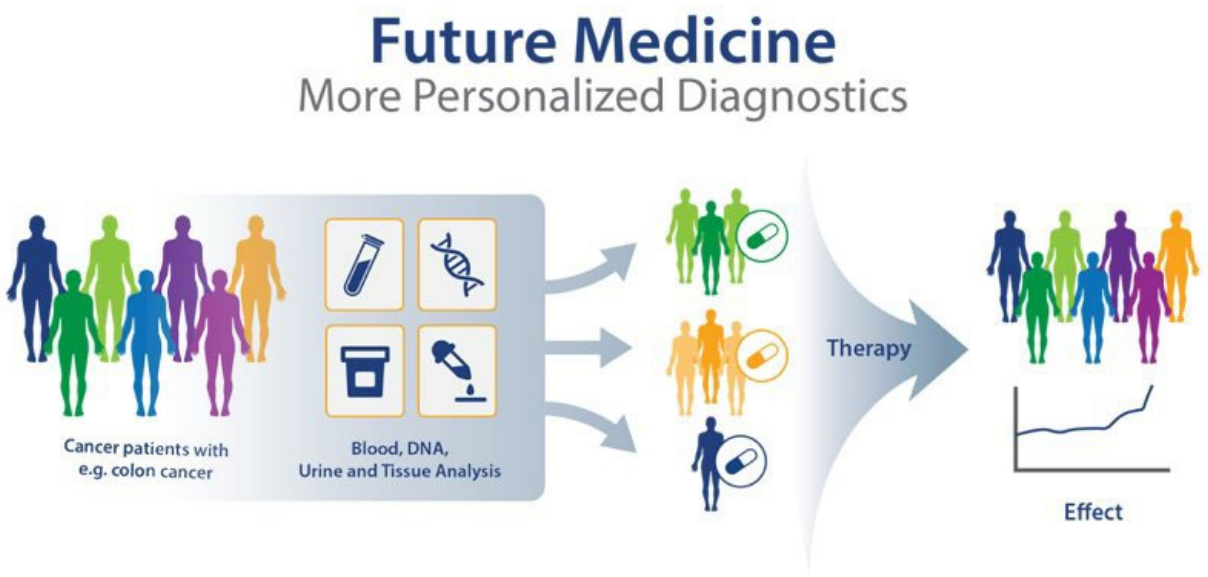
- Immune mediated disease affecting approx. 1 % of world population
- Affects the synovial membrane of joints.
- As it progressed it becomes a systemic disease affecting pulmonary, cardiac, neurologic, muscular, renal, vascular, hepatic and ophthalmologic systems.
- It causes chronic pain, disability, increased mortality
- No known cure

Treatment options for RA: trial and error



Personalised medicine





Pathobiology for Early Arthritis Cohort (PEAC)

- Aimed at determining biomarkers related to treatment response in treatment naïve early rheumatoid arthritis patients.
- Samples were taken of synovial tissue, blood, urine, and X-ray, doppler and histopathology imaging was used. Approx. 250 patients recruited from 2008 to 2014.
- Importantly disease tissue (synovial tissue from the joint) heterogeneity was postulated to play a major role in incomplete drug response.
- During this trial three RA pathotypes were described and linked to diverging transcriptomic signatures and response to therapy.

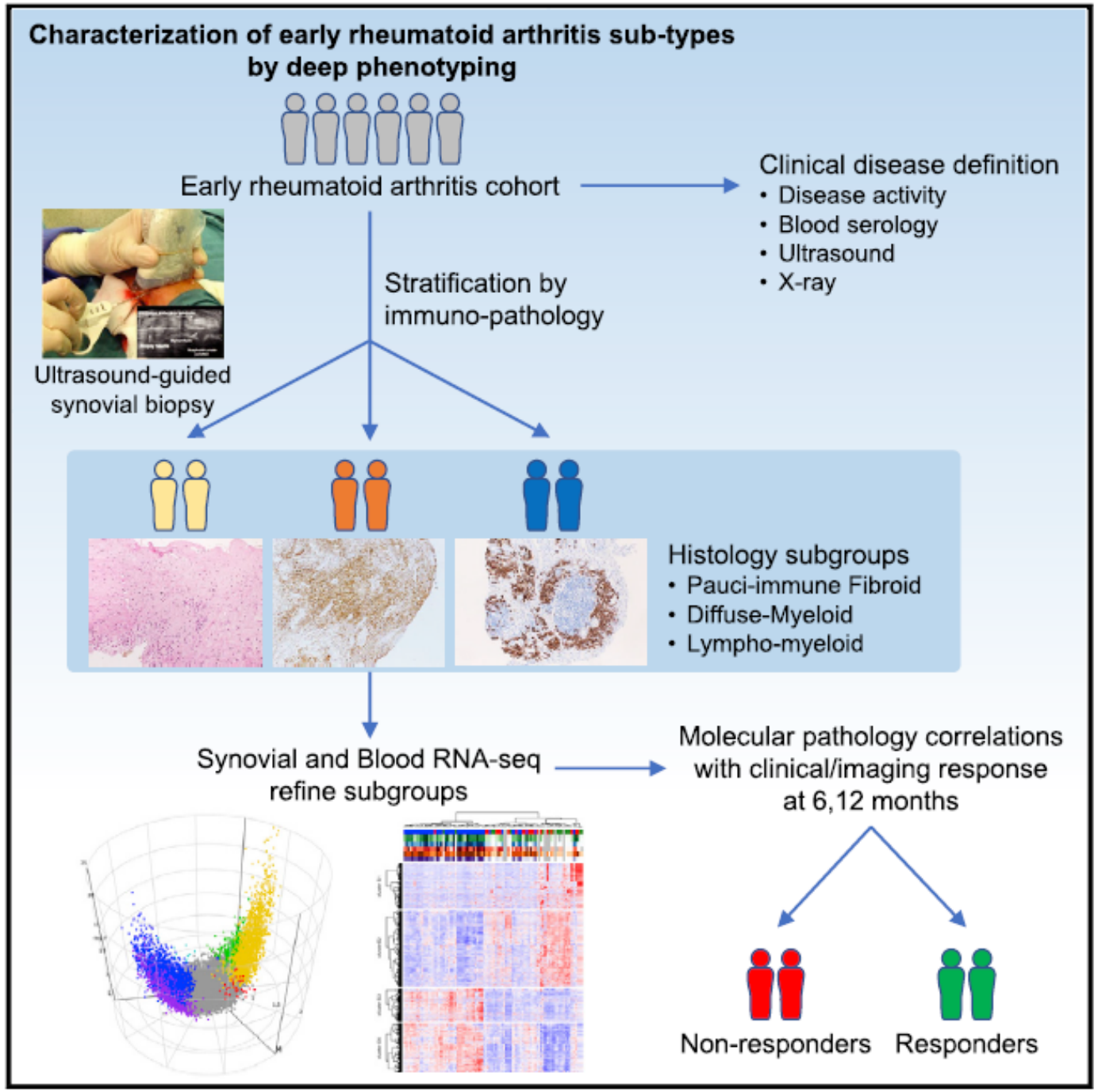
PEAC

Synovial biopsy sampling

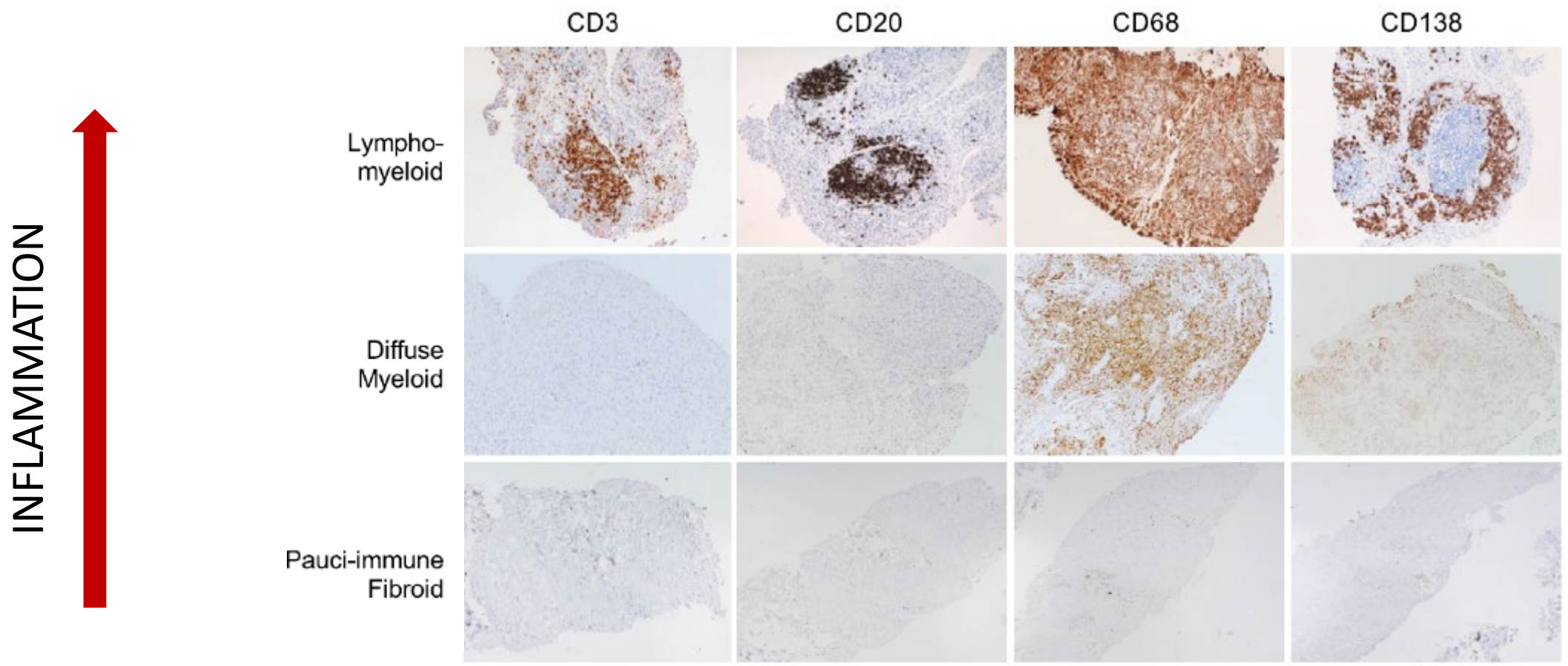
- A minimum of six synovial biopsies needs to be retrieved in large joints.
- A minimum of four synovial biopsies needs to be retrieved in small joints

- 6 biopsies for histology (stored in formalin)
- 6 biopsies for RNA extraction (in RNA preservative solution)

Lewis et al Cell Reports 2019



RA synovial biopsies pathotypes



[Humby et al ARD 2019]

Immune cells T cells B cells macrophages plasma cells

Aim

There is evidence histological pathotype is a predictor of drug response and disease progression in RA, a disease notable for its poor drug response and remission rate.

[1] Dennis et al Arthritis Research & Therapy 2014
[2] Lewis et al Cell Reports 2019
[3] Dana et al Arthritis & Rheumatology 2018
[4] Humby et al The Lancet 2021

There are numerous applications of deep learning frameworks to cancer histopathology, but to my knowledge there have been no published attempts to use a deep learning framework to classify histopathology slides by pathotype. The closest I could find was the work by Dana et al Arthritis & Rheumatology 2018.

Aim

The aim of this study is therefore to implement a simple, proof of concept, machine learning model to classify patients with early, treatment naive Rheumatoid Arthritis into three distinct pathotypes: pauci-immune, diffuse-myeloid and lympho-myeloid.

If successful this could be useful tool and benchmark for pathologist and clinicians seeking to further understand their patients pathotype.

Methods

1. Applying Convolutional Neural Networks (CNNs) to the image data
2. Using Transfer learning
3. Data augmentation
4. Ensembling of results through voting systems

Methods

1. All code written using Python 3.7 in Spyder
2. Neural Networks, transfer learning and data augmentation implemented using PyTorch
3. Other packages used included Numpy, Scikit-learn, Pandas, Torchvision, SciPy, …
4. DELL G5 Desktop Intel(R) Core(TM) i7-10700F CPU @ 2.90GHz 2.90 GHz, 16 GB RAM, GPU NVIDIA® GeForce® GTX and RTX2070

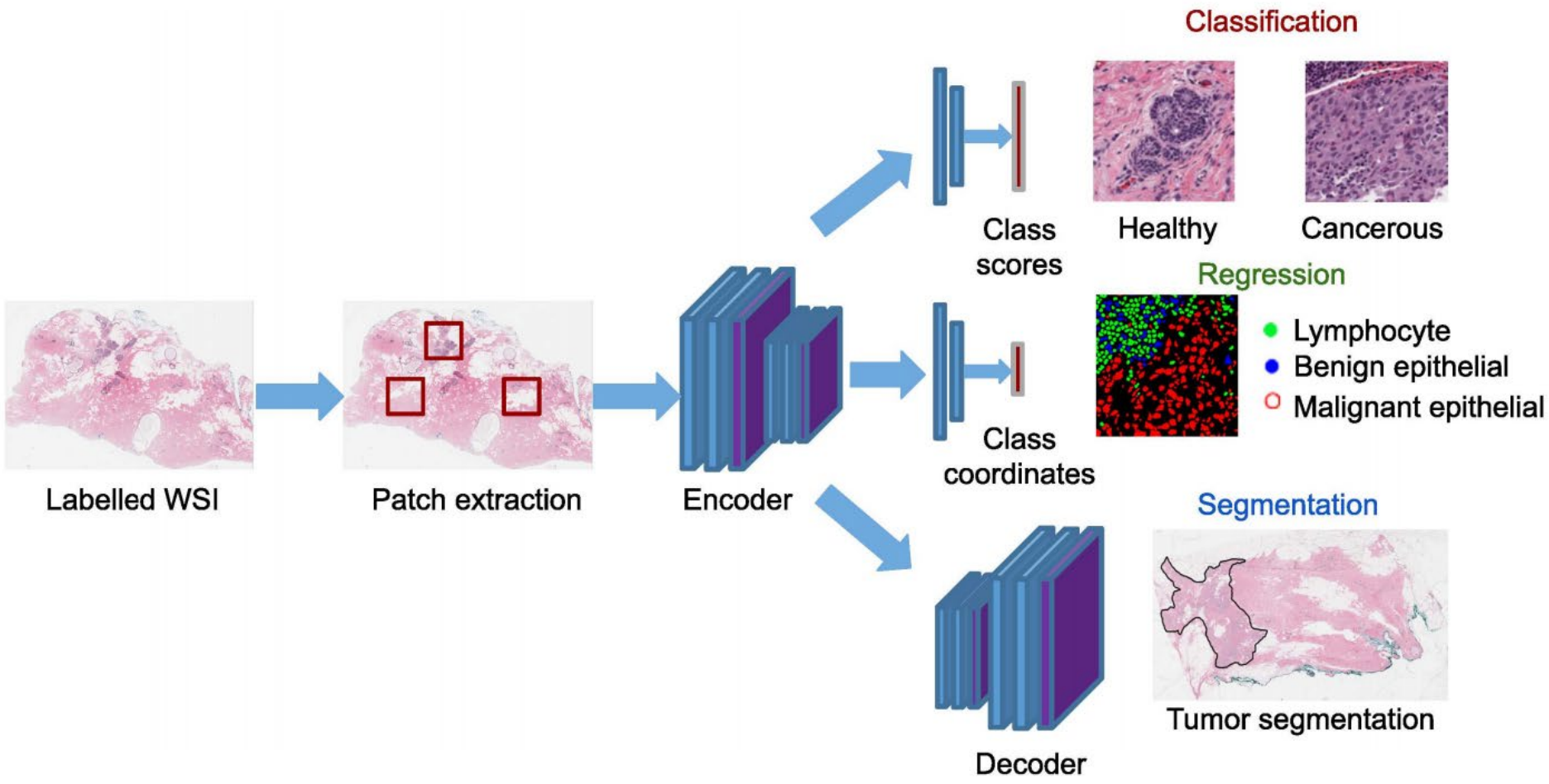
Supervised learning

- When training a supervised learning algorithm, the training data consists of inputs paired with ground truth labels. During training, the algorithm will search for patterns in the data that correlate with the desired outputs. After training, the objective of a supervised learning model is to predict the correct label for newly presented input data.

- Supervised learning can be used to implement classification, regression and segmentation tasks.

- Supervised learning consists of machine learning (RF, SVM, Logistic) and Deep Learning, which uses neural networks to address the same tasks.

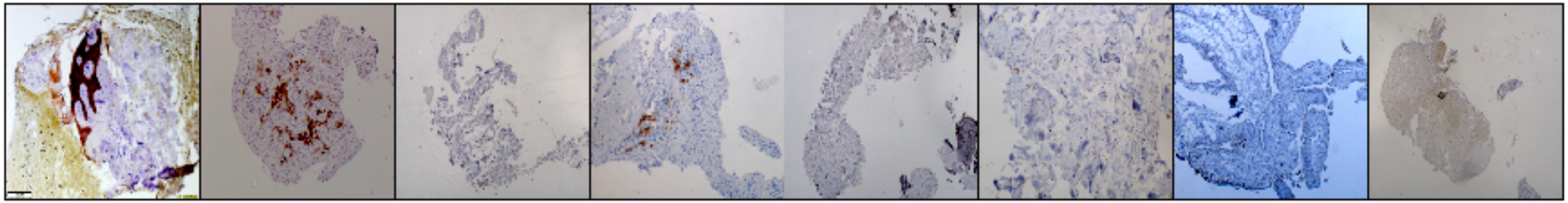
Supervised learning tasks



Supervised machine learning

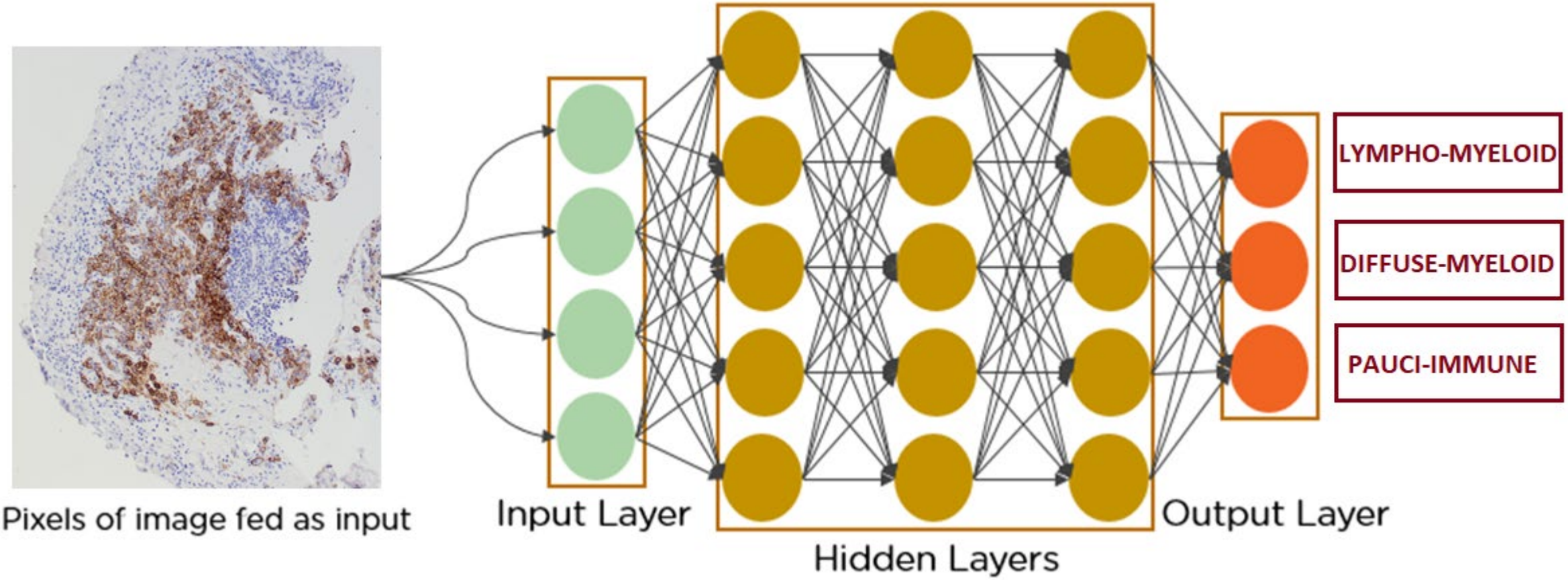
- In our case the histopathology slides are labelled with the patients pathotypes:

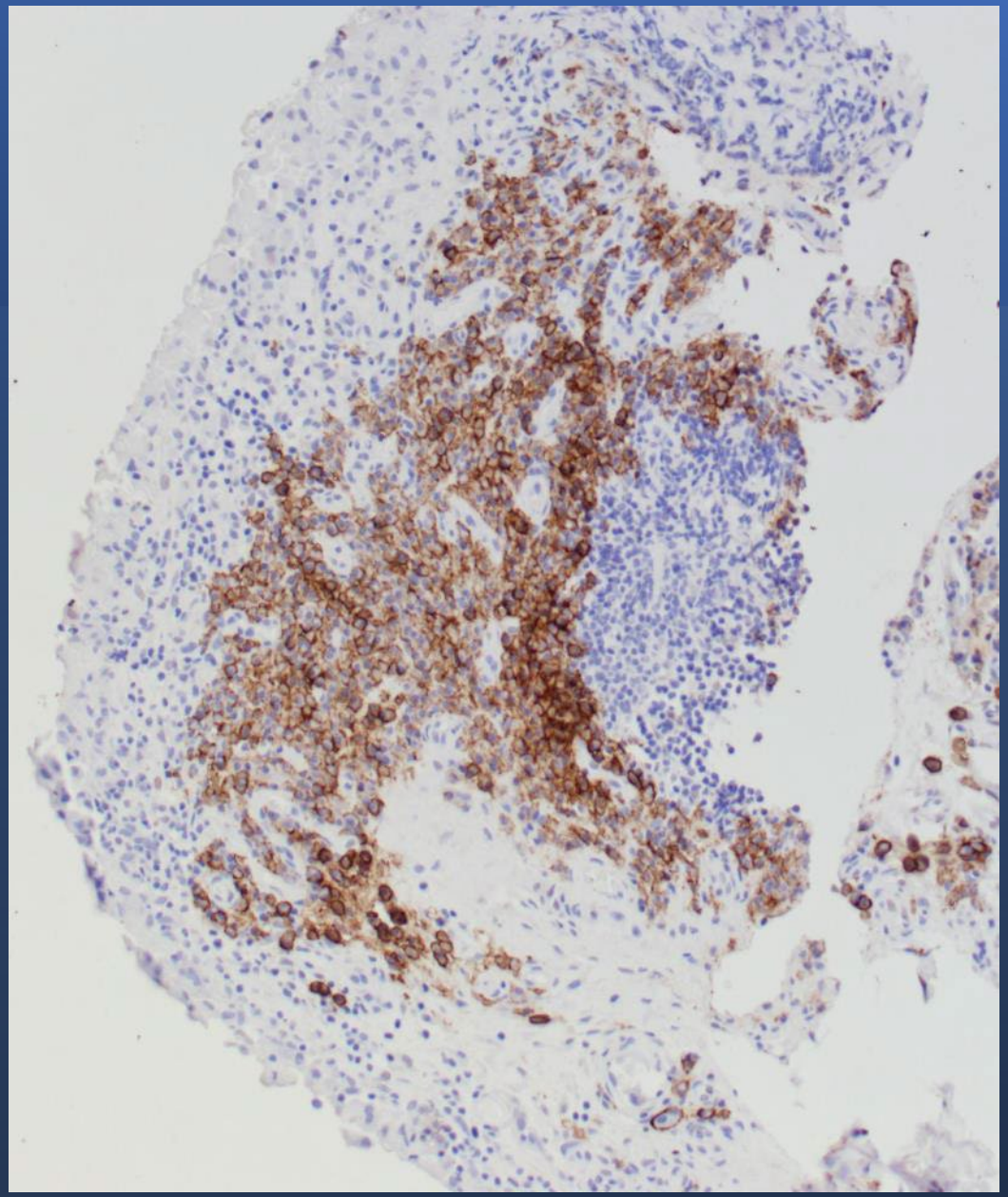
['Pauci-Fibroid', 'Lympho-Myeloid', 'Diffuse-Myeloid', 'Diffuse-Myeloid', 'Pauci-Fibroid', 'Pauci-Fibroid', 'Diffuse-Myeloid', 'Diffuse-Myeloid']

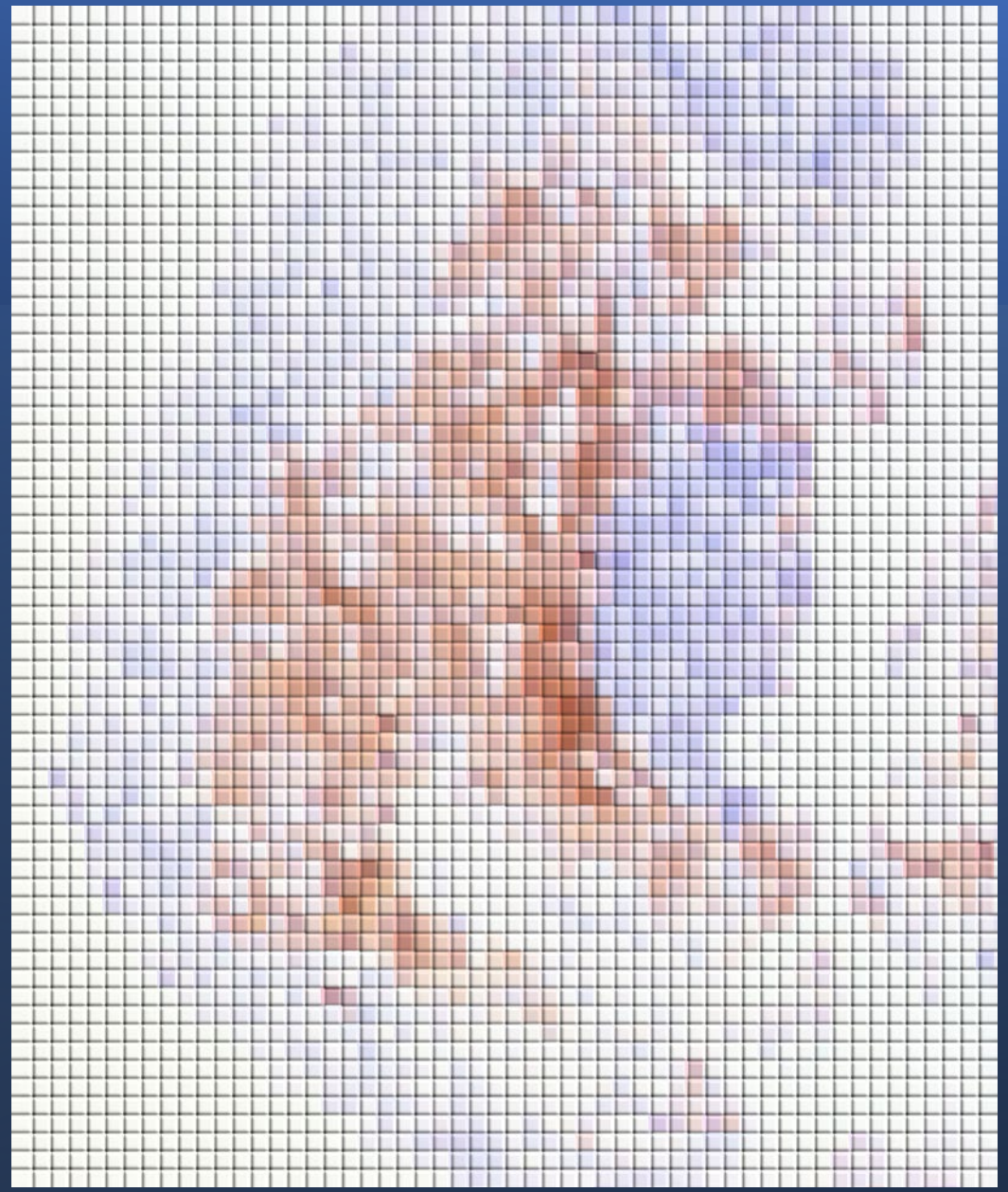


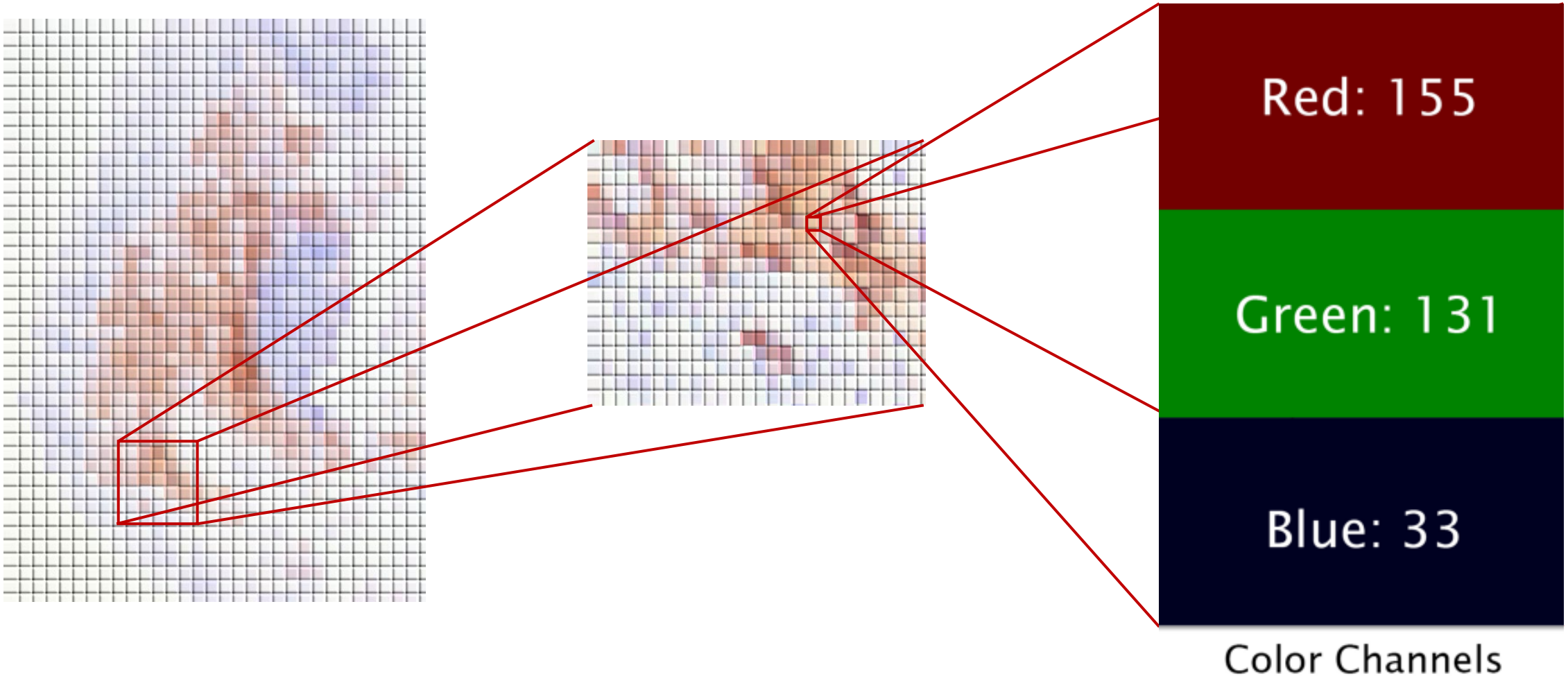
- It’s a classification task, where the images are classified into three categories.

Deep Neural Networks

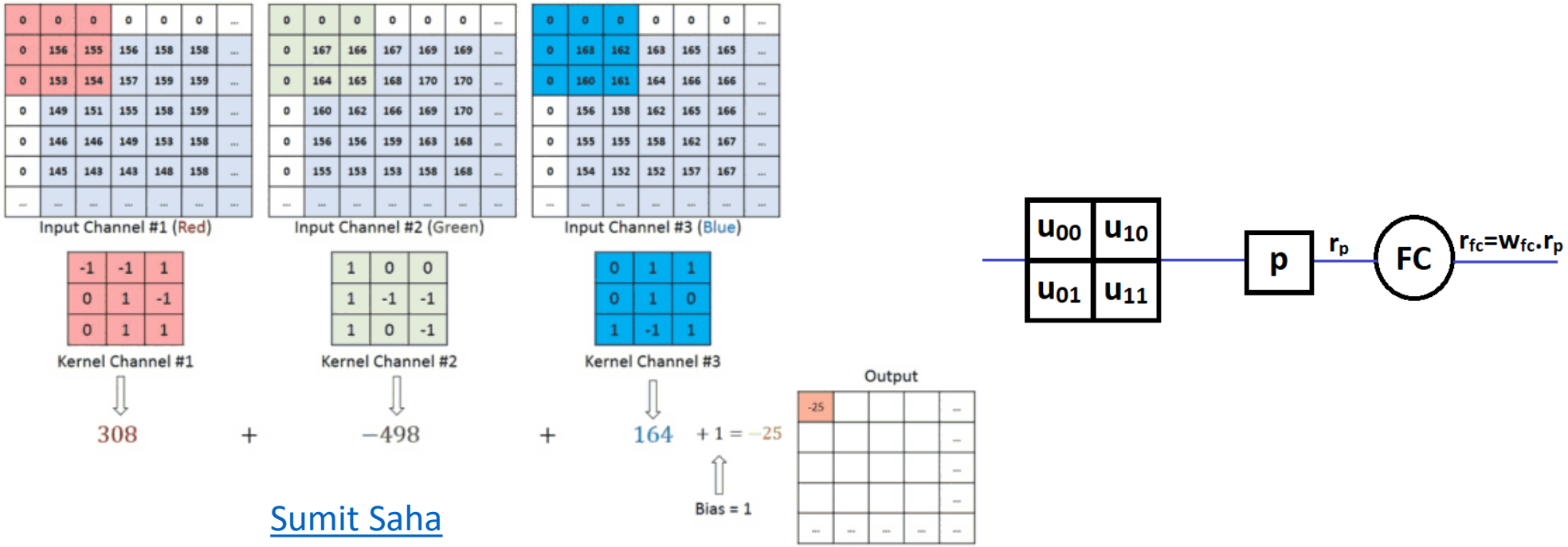






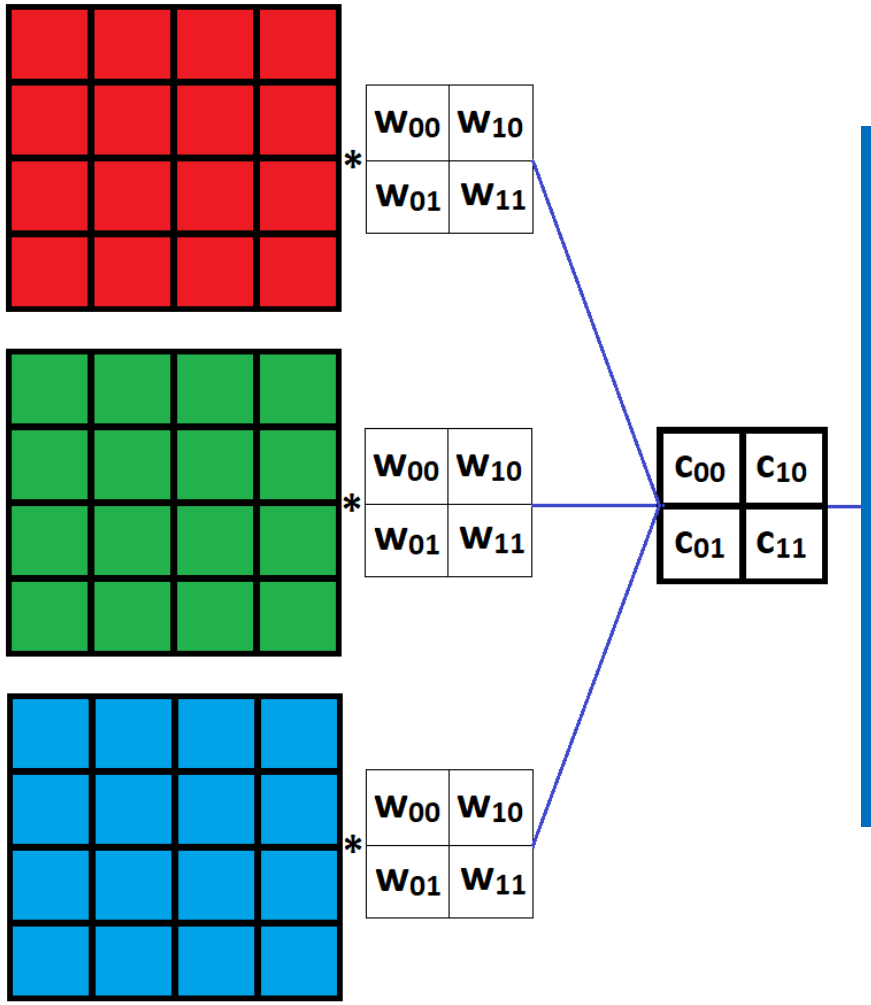

Red: 155
Green: 131
Blue: 33
Color Channels

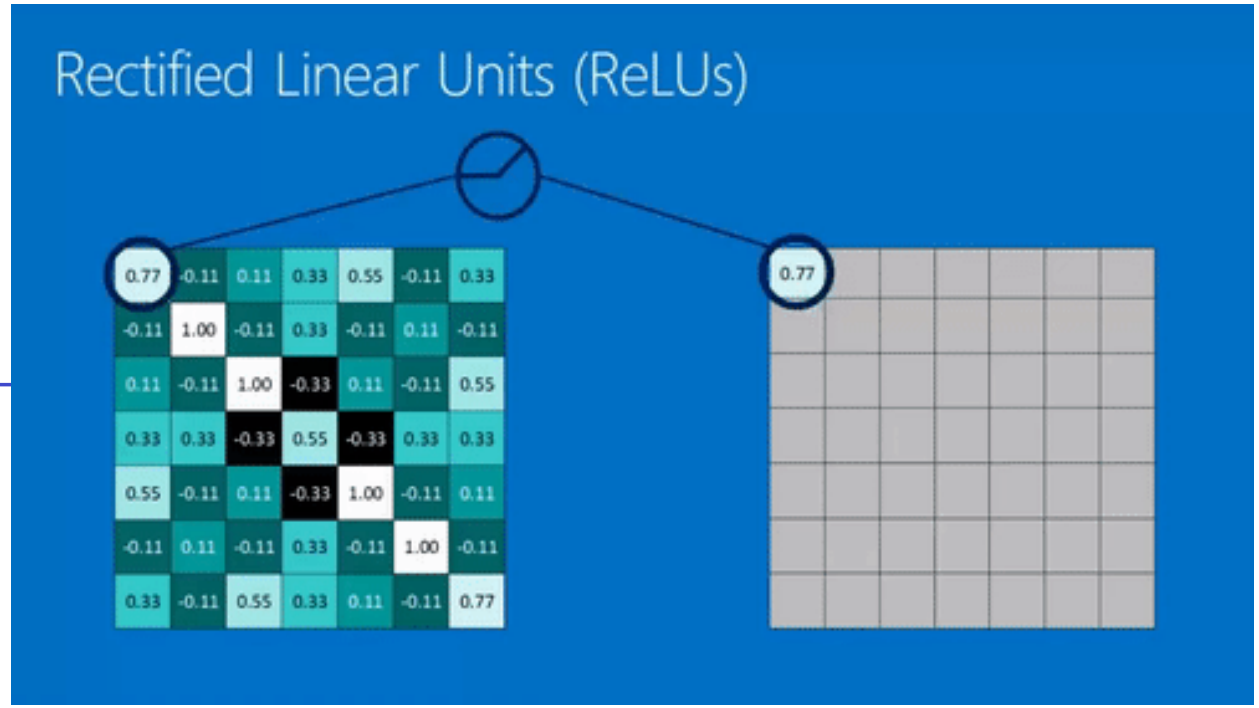
Convolutional Neural Networks



Input Filter(F) Convolution ReLu Pool Fully Connected

Convolutional Neural Networks

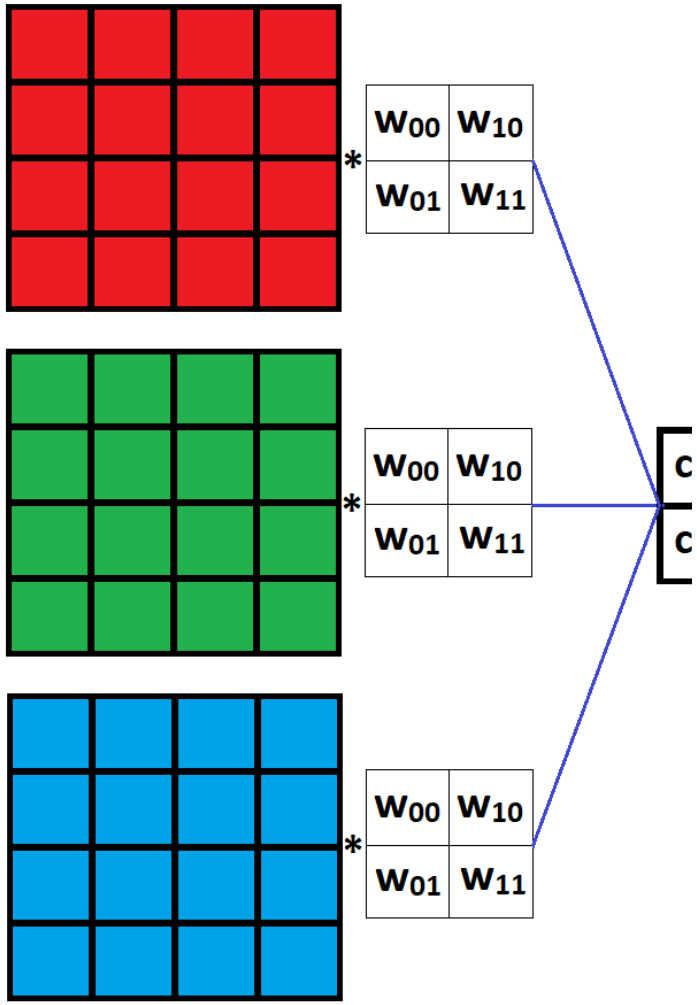




[Brandon Rohrer]

Input Filter(F) Convolution ReLu Pool Fully Connected

Convolutional Neural Networks



Type: max pool - Stride: 1 Padding: 1

0	0	0	0	0	0	0
0	4.3	5	12	3.7	11	0
0	12	12	6	11	13	0
0	8.5	8.4	7.6	6	10	0
0	3.9	11	5.7	3.6	11	0
0	8.3	5.8	9.7	13	7.1	0
0	0	0	0	0	0	0

Input

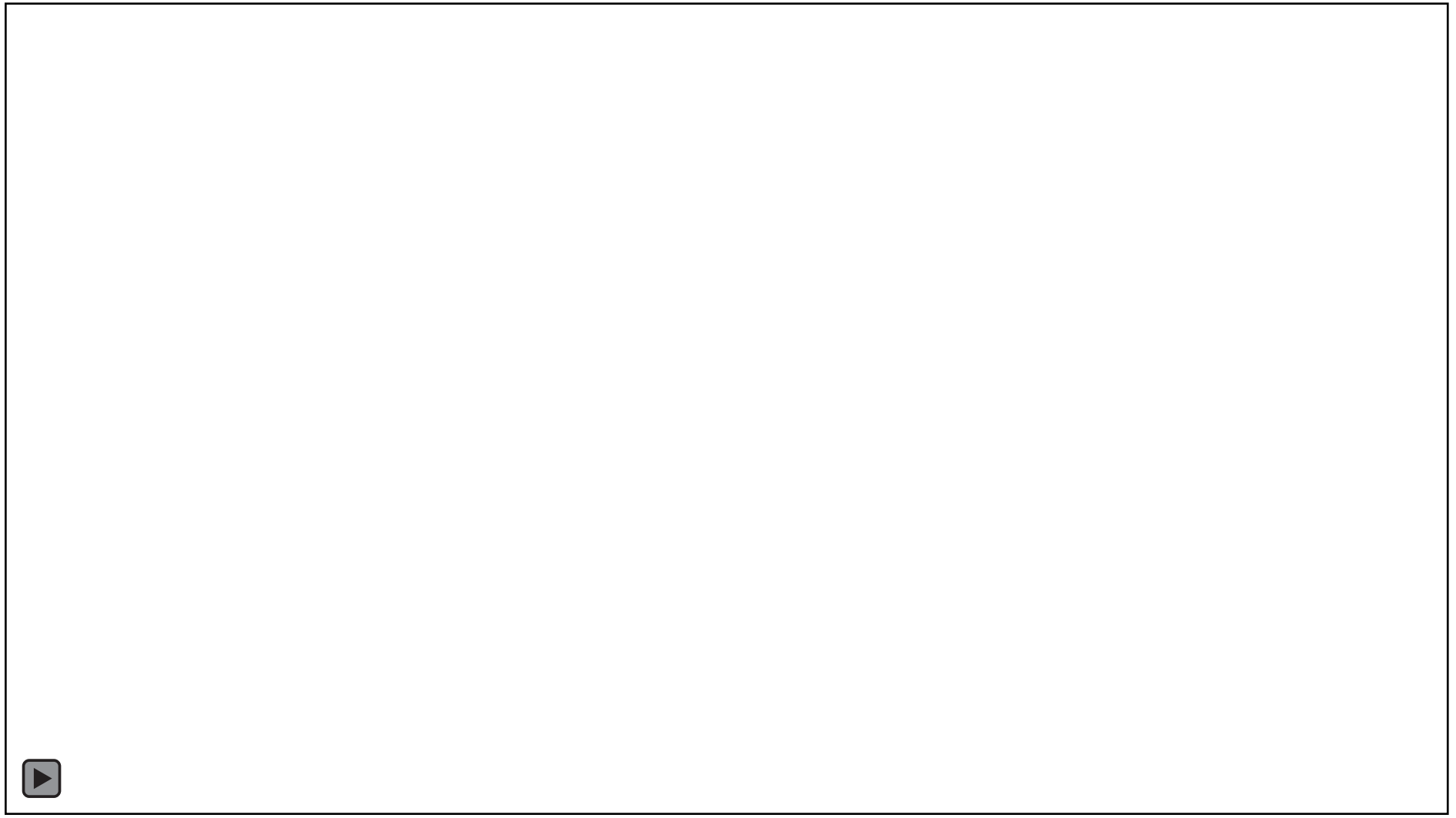
12	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Output

Aqeel Anwar

Input Filter(F) Convolution ReLu Pool Fully Connected

Convolutional Neural Networks



3Blue1Brown

Transfer learning

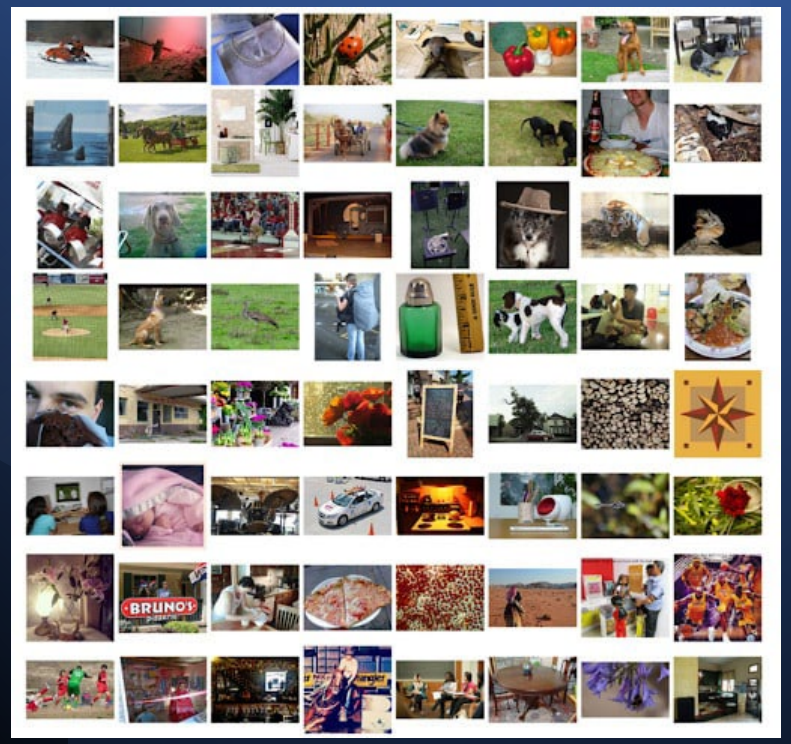
- CNNs have a very large number of parameters, hence training them from scratch is very data intensive.

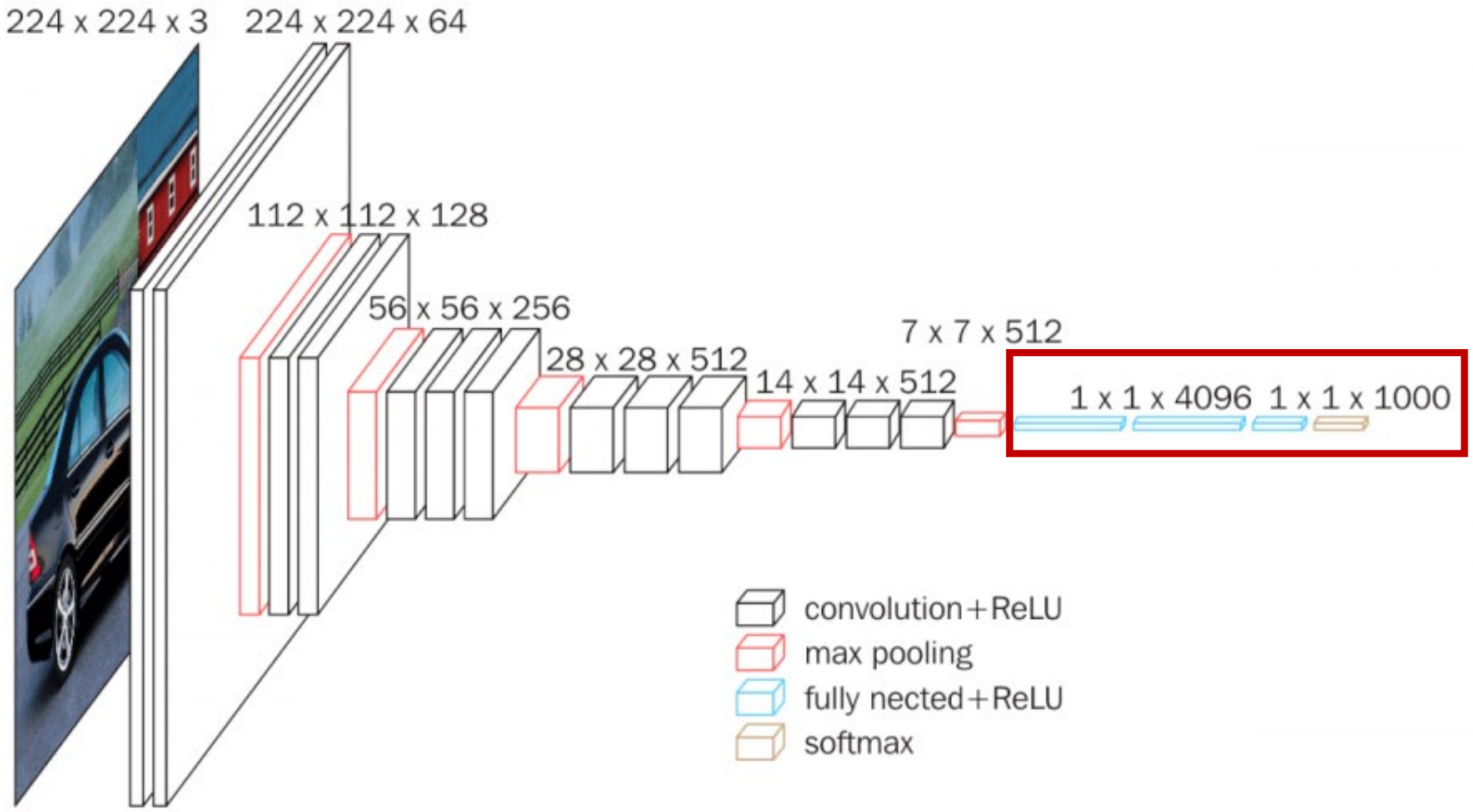
- In our case, approx. 200 patient with on average 25 images is nowhere near enough to train a network from scratch.

- To overcome this problem, we use transfer learning, which are networks which have been previously trained on millions of unrelated image data. The trained weights from this network are then reused on our data and only the last layers of the network are trained.

VGG16

A CNN architecture pretrained on ImageNet (10 million images)



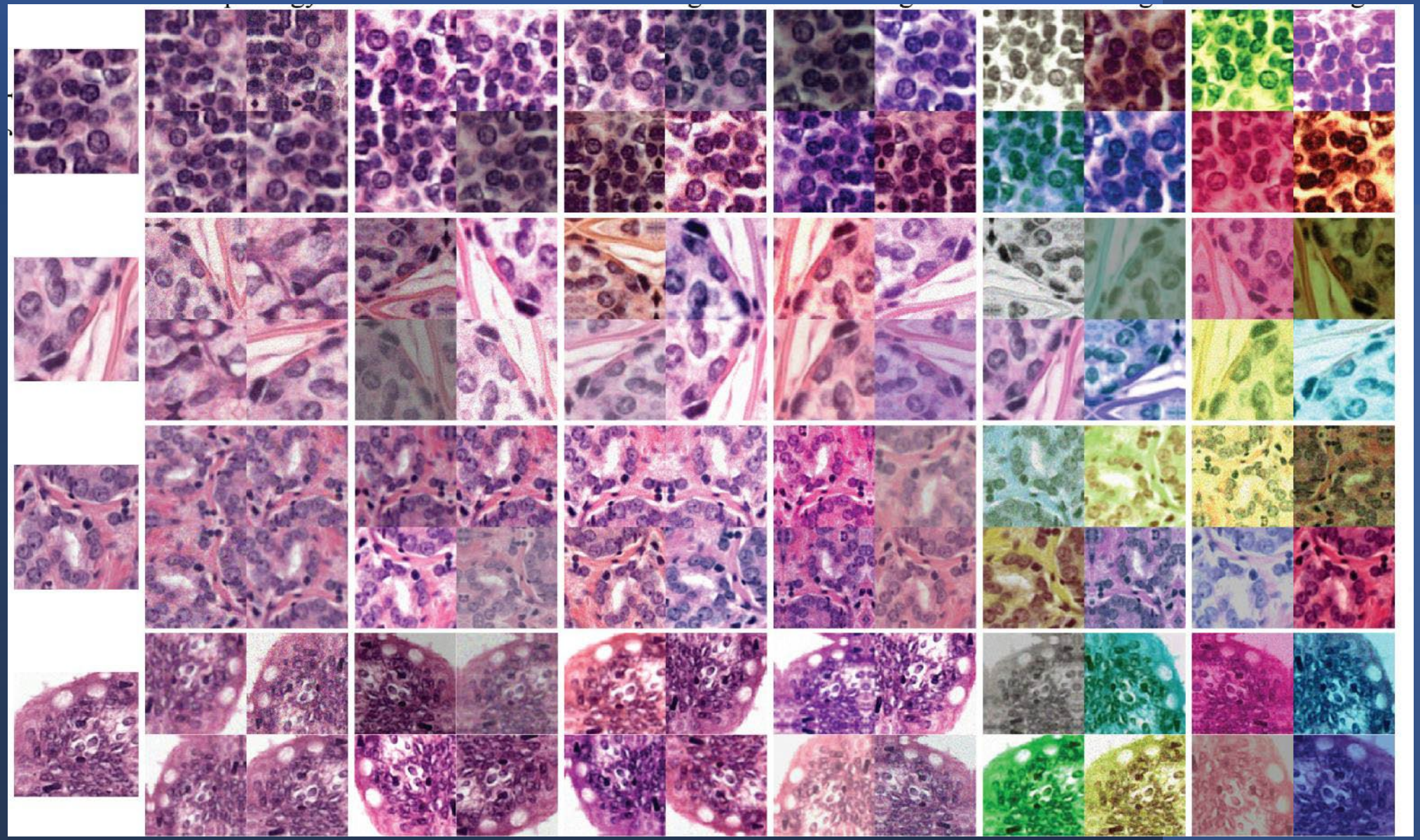


138 million parameters

Data augmentation

- Another technique aimed at circumventing the lack of medical image data is data augmenting.

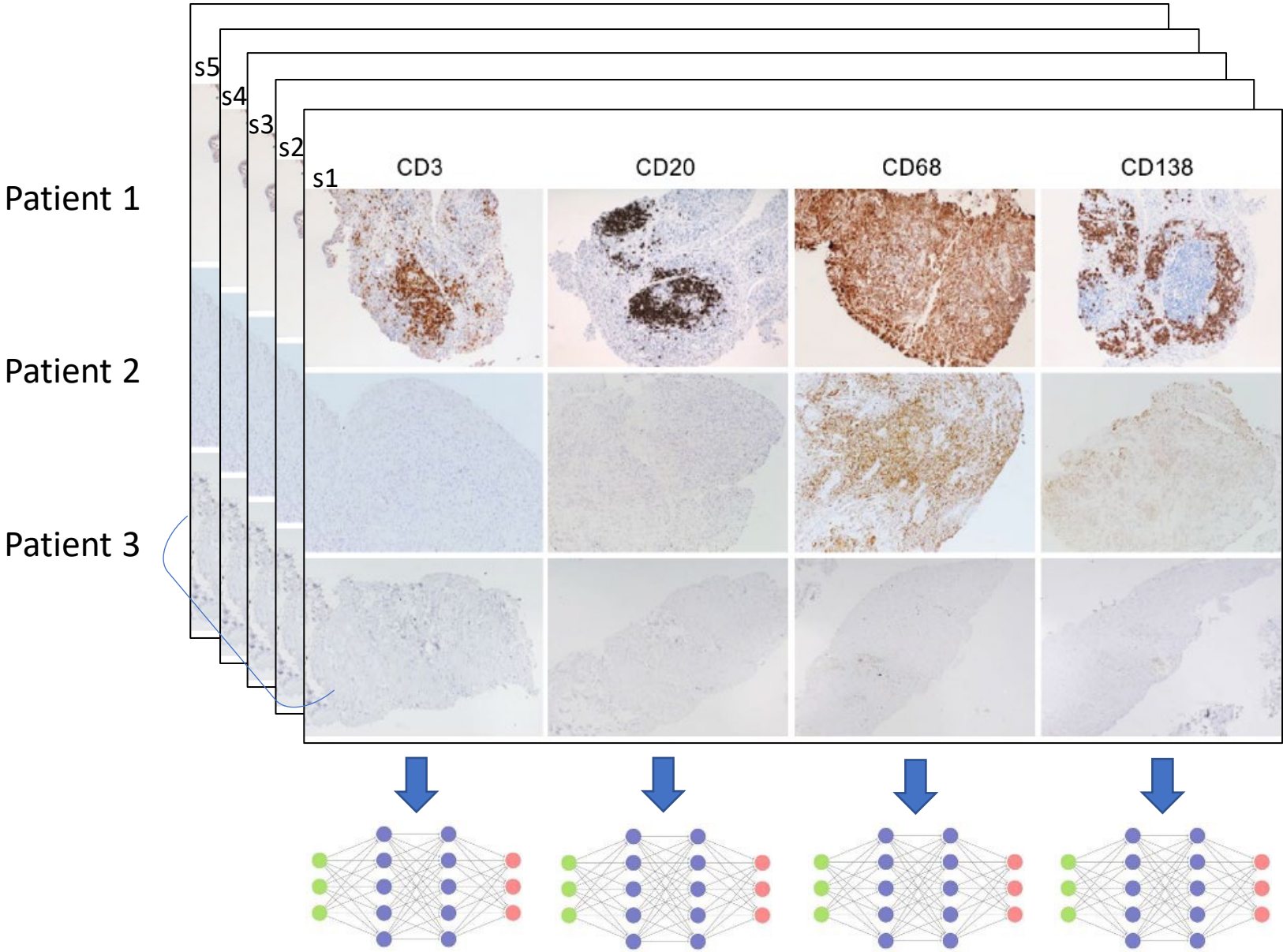
- Basically we make loads of new images by applying many small transformations to our data



Tellez et al Medical Image Analysis 2019

Ensembling results

- Each patient had several samples taken for each immune cell marker.
- For the purpose of training the network each image was considered individually
- A network was trained for each immune cell type with all images for this type



Labelling issues

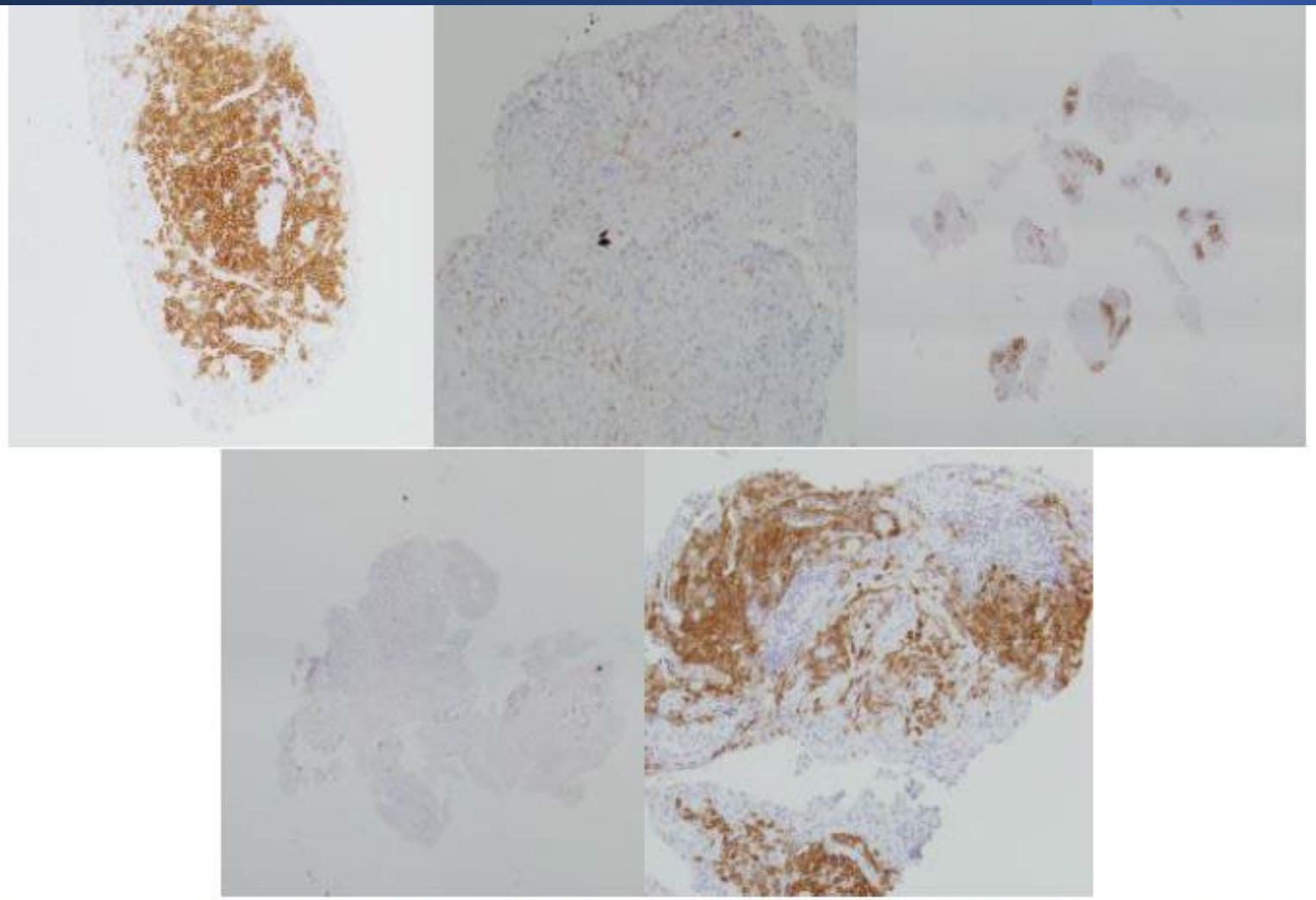


Figure 3.5: Biospy samples stained for CD138^{+} plasma cells from a single patients. Some sample.

Voting systems

Samples from the same patient often showed different pathotypes, so several voting systems were implemented to disentangle results:

- majority voting
- soft voting
- mixed type voting

Results for each stain type were further aggregated via voting system. Every majority voting permutation for stain type was run and the vote combining results for CD20, CD68 and CD138 stain types were consistently better than any other combination.

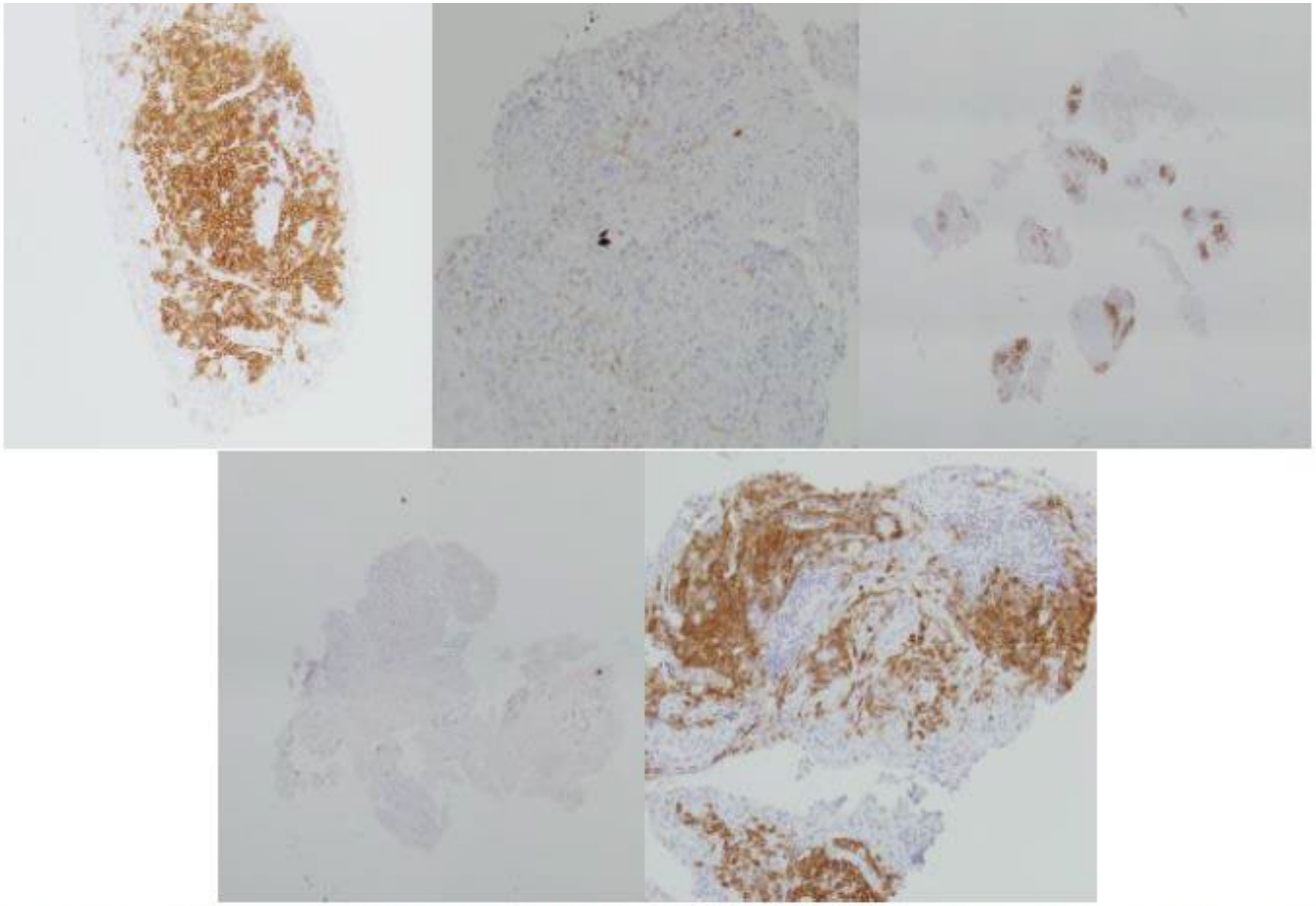
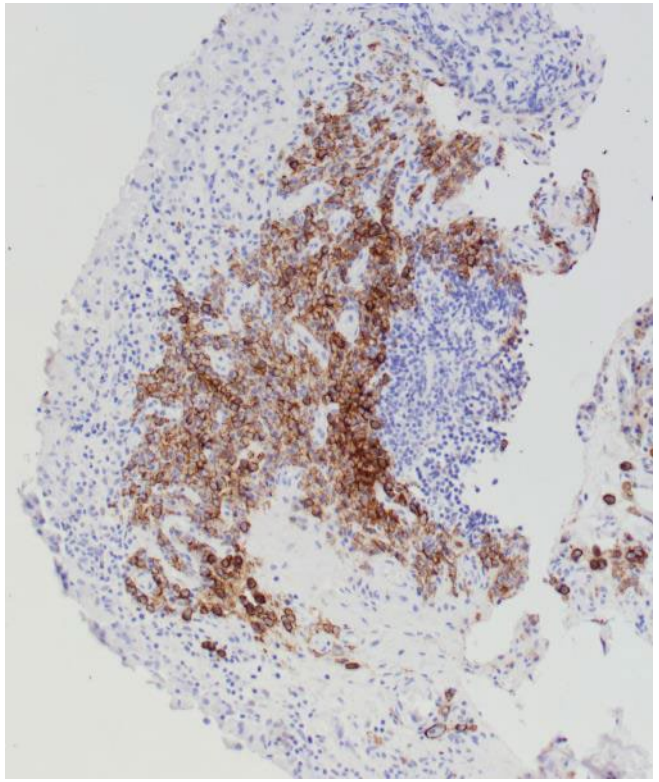
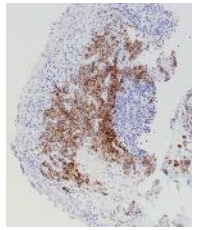


Figure 3.5: Biospy samples stained for CD138+ plasma cells from a single patients. Some sample

Complete Pipeline

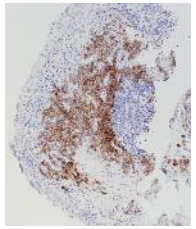
Load → Resize → Split → Augment



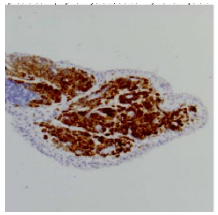


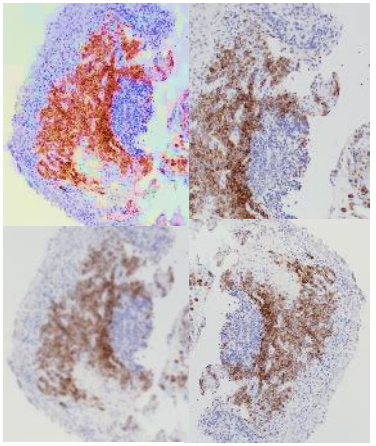
224 x 224

Training set – 67 %



Testing set – 33 %

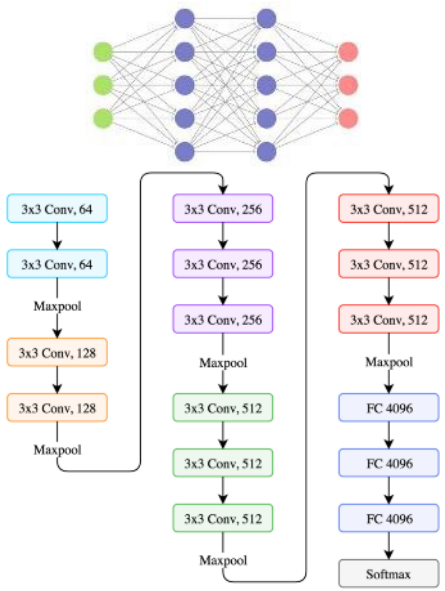




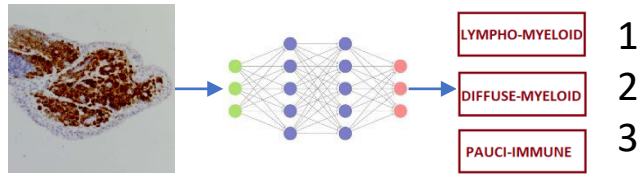
Non-overlapping
stratified random 3 fold split

Complete Pipeline

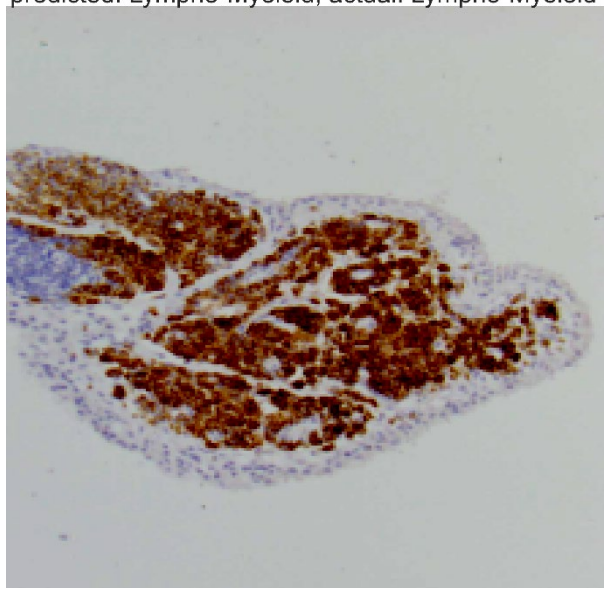
Train → Test → Vote

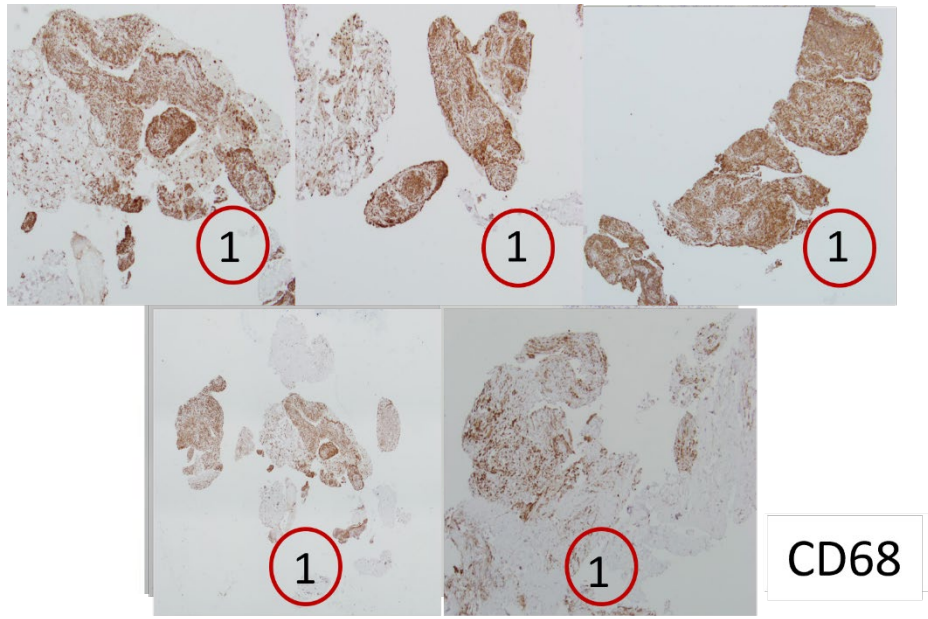


30 epochs
3 non-overlapping train/test sets
10 hours



predicted: Lympho-Myeloid, actual: Lympho-Myeloid





CD68

PATIENT → LYMPHO-MYELOID

Results

	Precision	
	Average	StDev
Diffuse-Myeloid	0.68	0.04
Lympho-Myeloid	0.82	0.08
Pauci-Fibroid	0.62	0.06
weighted avg	0.72	0.02

Table 3.6: *The precision, recall and f1-score results for the majority voting results of the ensembled CD20, CD68 and CD138 networks are presented by class and by overall weighted average.*

Results

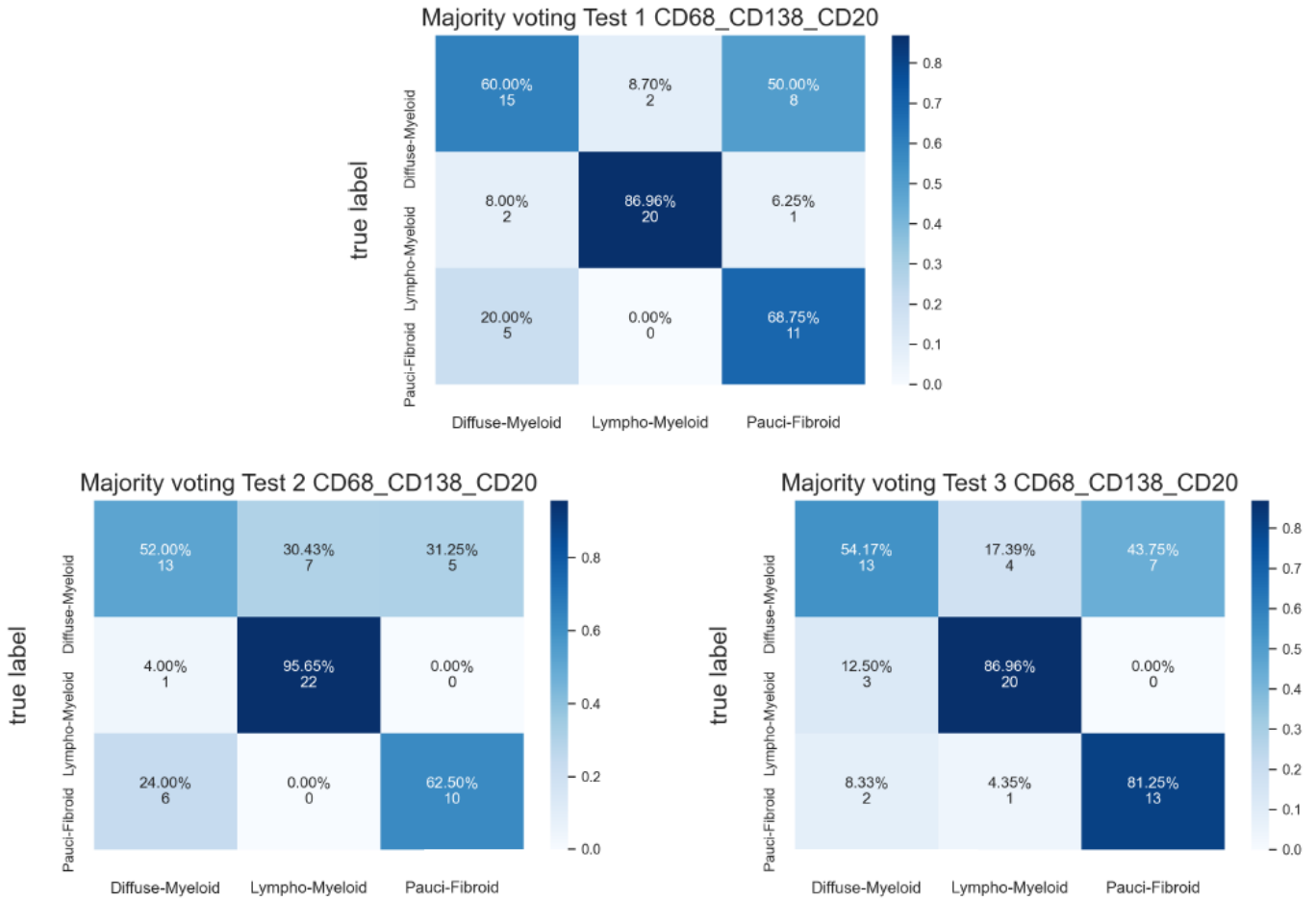


Figure 3.7: Confusion matrices for each test in the 3-fold cross validation of the majority vote for CD68, CD20 and CD138 VGG16 netowrks.

Results

	Average Accuracy	Standard Deviation
Lympho-Myeloid	0.90	0.02
Pauci + Diffuse	0.86	0.05

Table 3.7: Average accuracy for the merged Pauci-Fibroid and Diffuse-Myeloid classes versus Lympho-Myeloid.

Future work

- Adapt this pipeline to images from the R4RA trial using RNAseq B-cell rich/poor clustering of samples as labels?
- Design a pipeline combining histopathology and eco-doppler images
- Apply a pipeline for segmentation of ROIs?
- Could this be used to do automatic scoring?
- Explore potential usefulness of the model in clinical practice and consider developing an app or computer software package for ease of use?