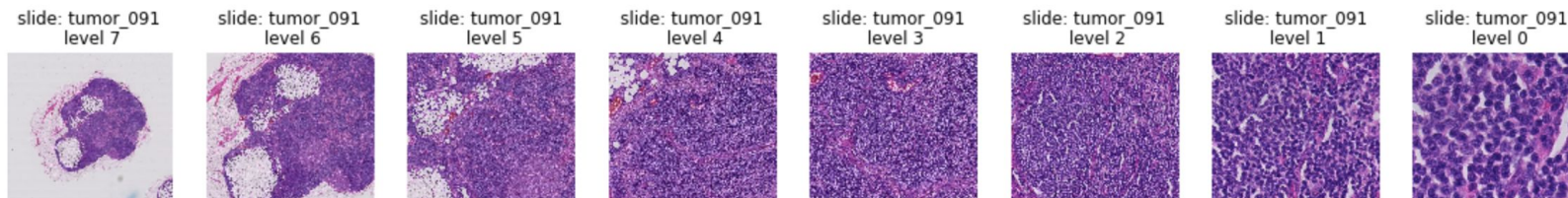


# Project Image Classification Cancer Tumors

## The Data

1-) Whole-slide images (WSI) of hematoxylin and eosin (H&E) as as TIFF images with 8 levels of magnification.

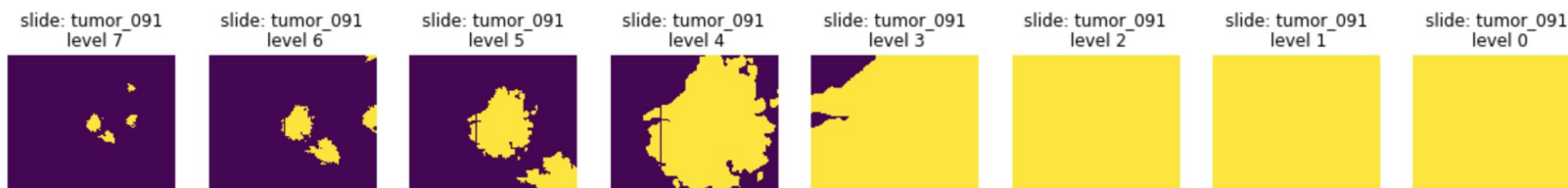


2-) Lesion-level annotations in XLM format with (X, Y) pixel coordinates at the highest resolution.

```
<ASAP_Annotations>
<Annotations>
  <Annotation Name="_0" Type="Polygon" PartOfGroup="_0" Color="#F4FA58">
    <Coordinates>
      <Coordinate Order="0" X="52519.4" Y="140936" />
      <Coordinate Order="1" X="52521.4" Y="140936" />
      <Coordinate Order="2" X="52529.2" Y="140936" />
    ...
  ...

```

3-) Mask file extracted from the annotation file where we can observe the location of the lesion (in yellow)

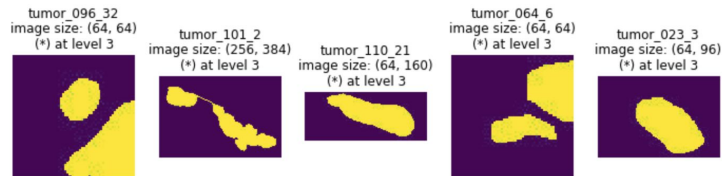


# Data Preparation

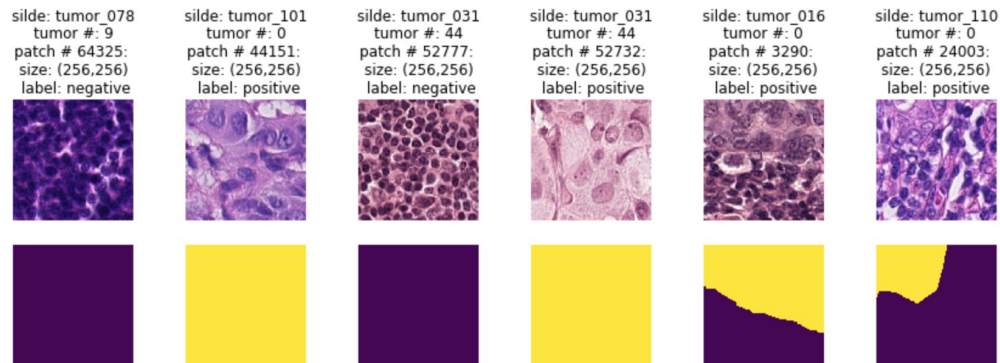
**Goal:** Extract 256x256 patches at the highest magnification with areas of lesion (positive cases) and lesion free patches (negative cases)

**How:**

1- Extract coordinates and size of the tumor area using XML annotation file



2- Extract 256x256 patches from the slide file restricted to the tumor areas



## Data FACTS

Total number of slides: **21**

Total number of tumors: **361**

Minimum number of tumors in slide: **1**

Maximum number of tumors in slide: **52**

Total train negative images: **20994**

Total train positive images: **24325**

Total validation negative images: **3455**

Total validation positive images: **4183**

Total test negative images: **6671**

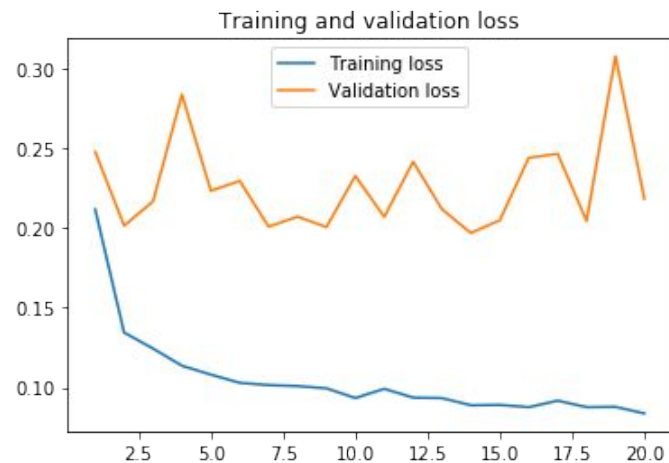
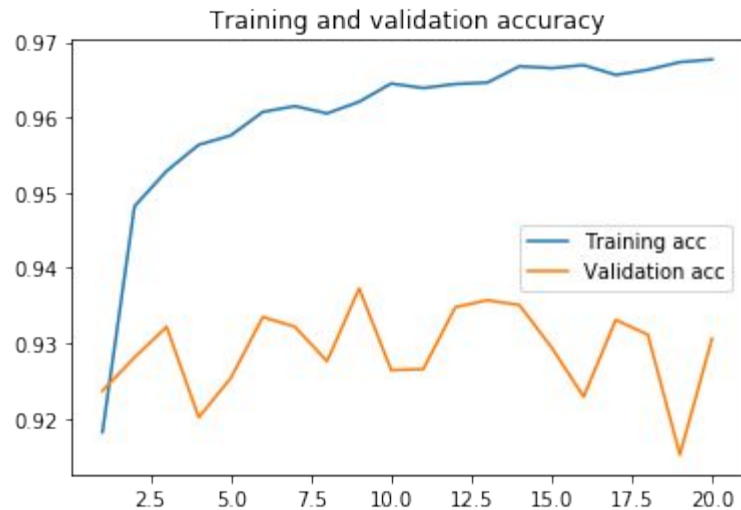
Total test positive images: **6167**

slide	Patches	Tumors						
0	tumor_001	552	2					
1	tumor_002	77	1					
2	tumor_005	148	4	10	tumor_059	42	1	
3	tumor_012	72	1		11	tumor_064	4761	7
4	tumor_016	4802	3			12	tumor_075	2892
5	tumor_019	148	3		13		tumor_078	35011
6	tumor_023	202	6	14		tumor_081	48	1
7	tumor_031	6539	52		15	tumor_084	1451	29
8	tumor_035	36	3	16		tumor_091	1751	5
9	tumor_057	190	19		17	tumor_094	1780	2
19	tumor_101	5830	8	18		tumor_096	1554	37
20	tumor_110	30219	13					

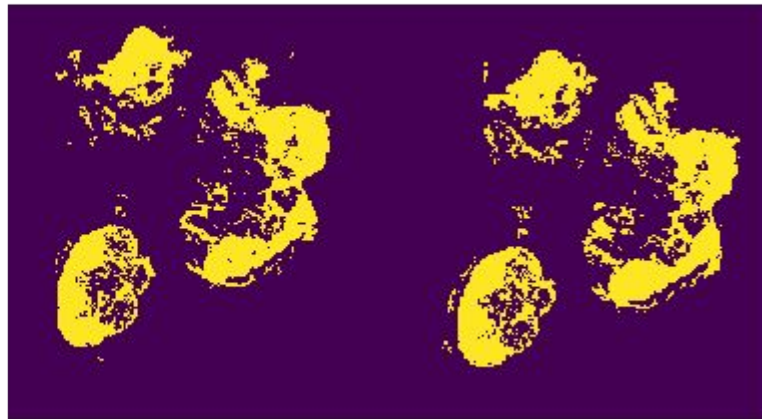
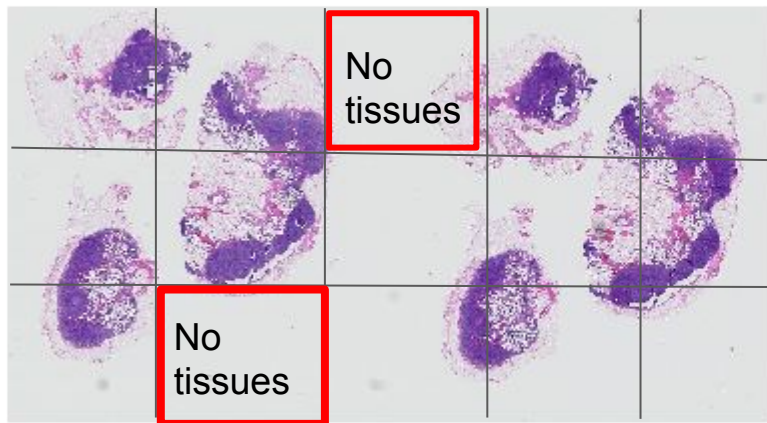
# The Model

Model: "sequential"

Layer (type)	Output Shape	Param #
=====		
vgg16 (Model)	(None, 4, 4, 512)	14714688
=====		
flatten (Flatten)	(None, 8192)	0
=====		
dense (Dense)	(None, 256)	2097408
=====		
dense_1 (Dense)	(None, 1)	257
=====		
==		
Total params: 16,812,353		
Trainable params: 9,177,089		
Non-trainable params: 7,635,264		

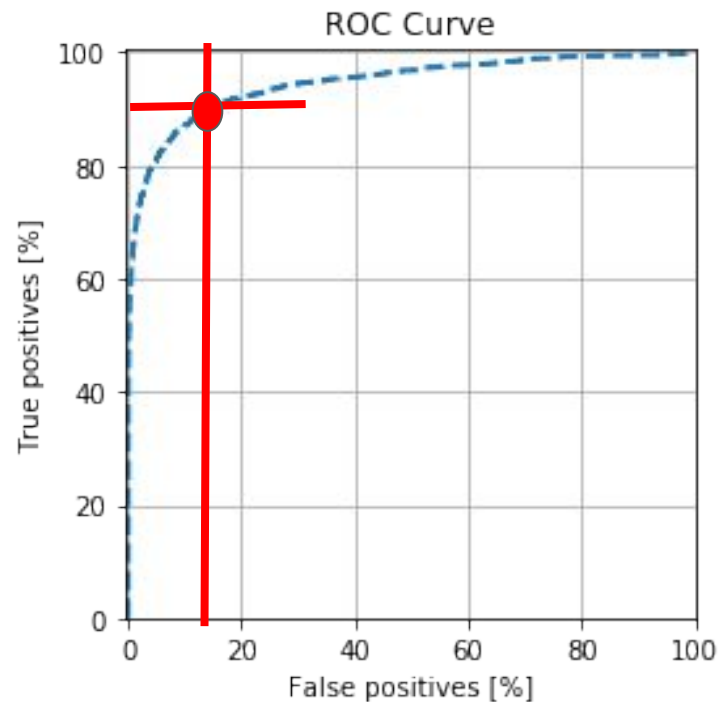
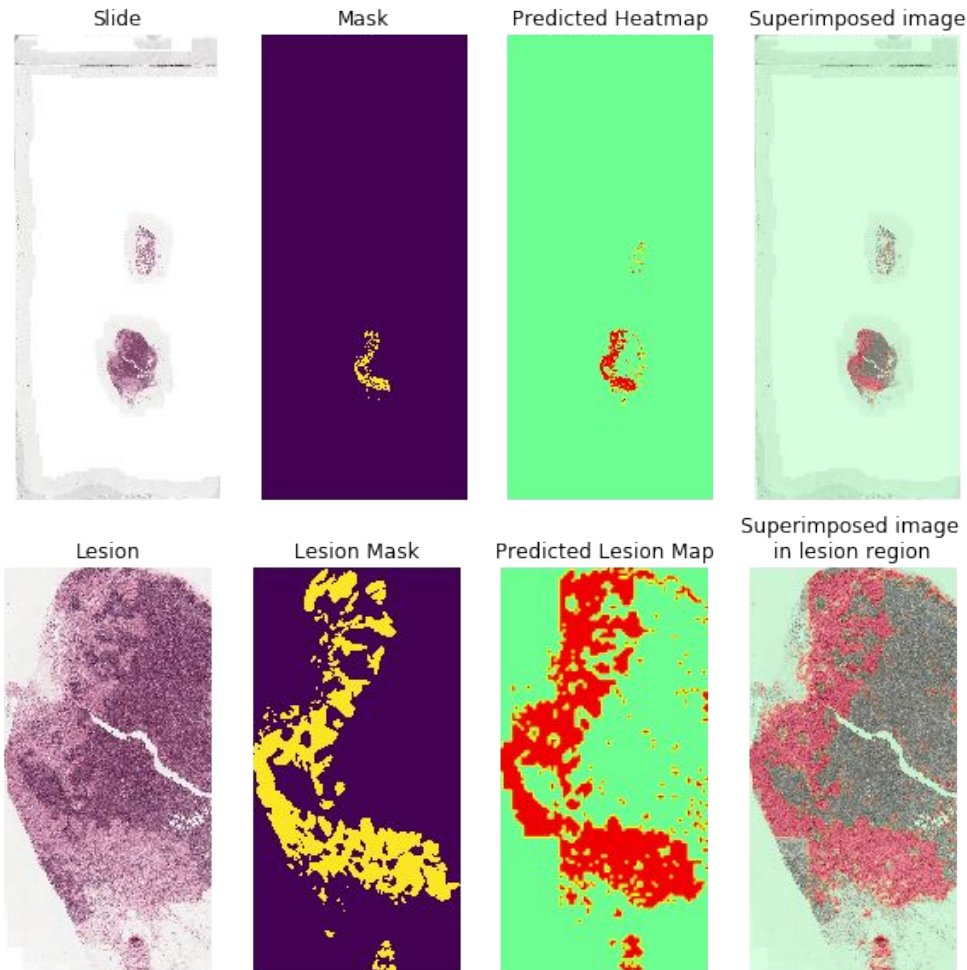


## Extracting patches from new slide. Discard regions with no tissues to reduce number of patches to evaluate



```
Start Time 18:30
Level 6 , Previous level total regions: 8
Level 5 , Previous level total regions: 25
Level 4 , Previous level total regions: 63
Level 3 , Previous level total regions: 201
Level 2 , Previous level total regions: 672
Level 1 , Previous level total regions: 2259
Level 0 , Previous level total regions: 7385
End Time 19:02
--- 1931.282926082611 seconds ---
(1024, 560)
Processing slide: tumor_096 with 21486 patches
```

# Evaluation



The images were generated using threshold = **0.5**  
True Positive Rate is: **90%**  
False Positive Rate is: **14%**