

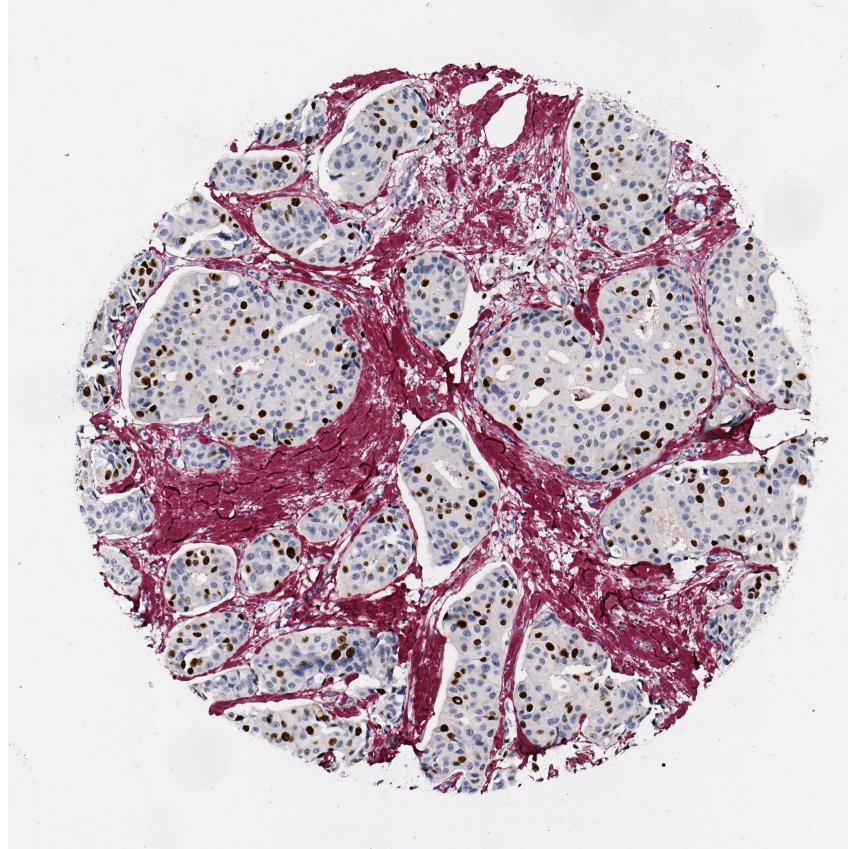
Collagen-based survival analysis in the Hematoxylin and Eosin stained histological images.

Project Supervisors: Dr. Povilas Treigys (VU Mathematics Faculty), Mindaugas Morkūnas (Pathology Centre)

Marius Latinis

Project Overview

I am given a list of images (92), one image corresponds to one cancer patient:



Attached to an image is the survival data for that patient:

the patient's death time, measured as the number of months until the patient's death starting from the time an image was acquired

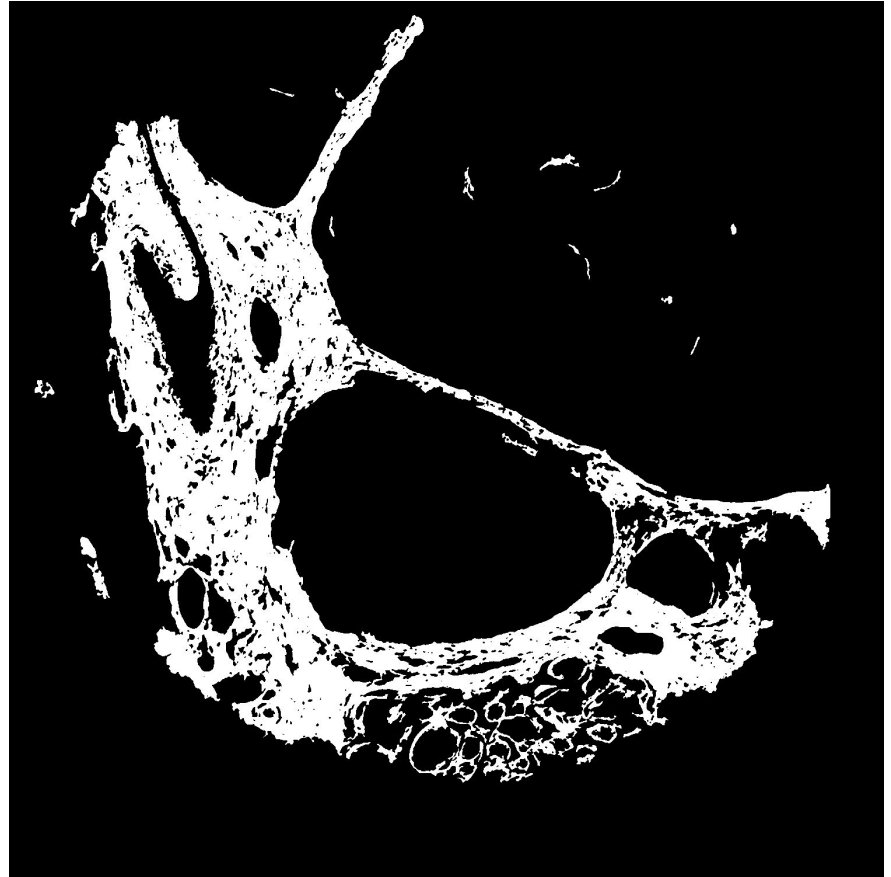
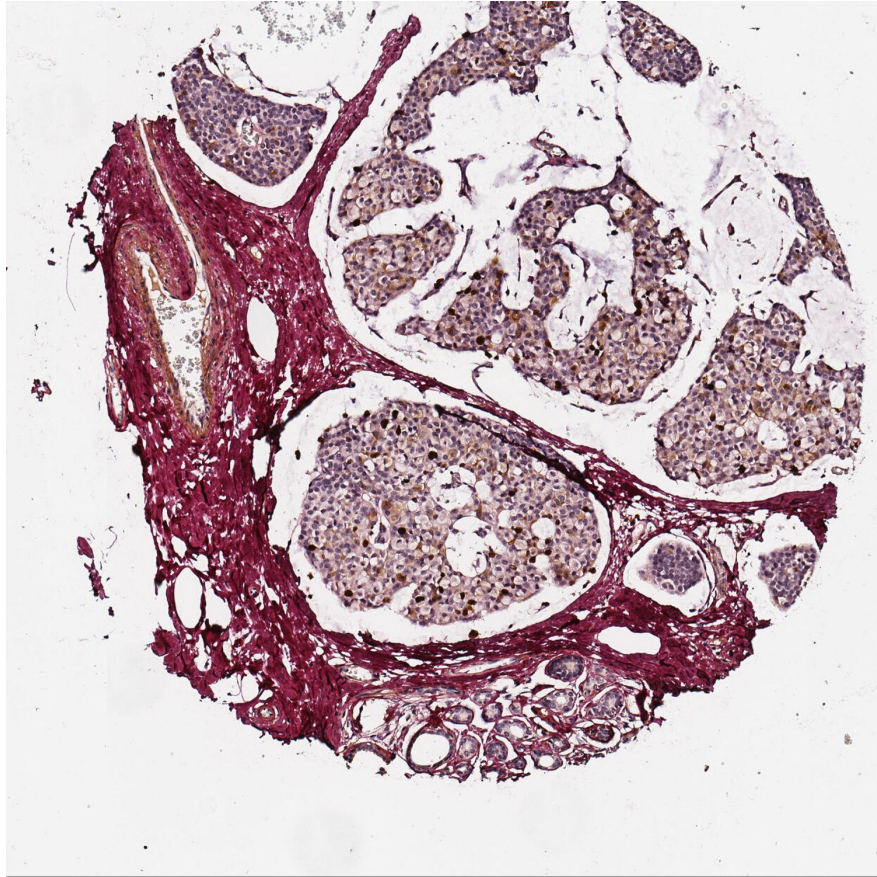
or

a censoring event, if at the end of the study the patient was still alive.

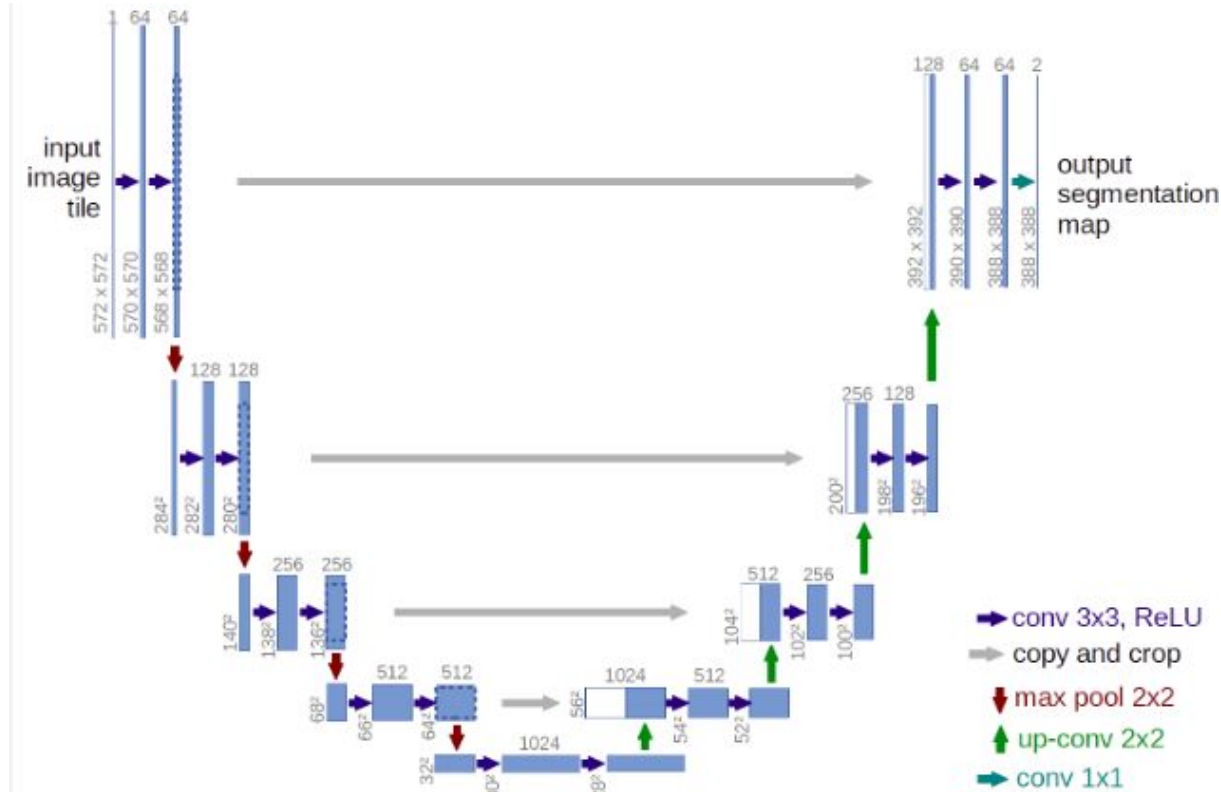
patId	id	ost_m	status01
1	28979	120	0
2	23721	76	1

The project aims to write an automated pipeline that takes an image and predicts survival data correlating with the known data for that image.

Collagen Extraction

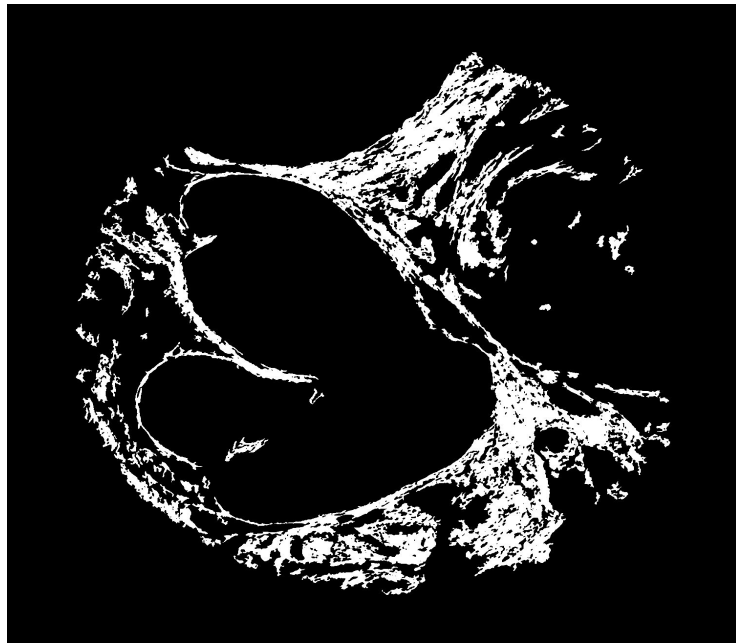


Collagen detection viewed as an image segmentation problem, solved using the U-net architecture



Binary classification problem

Whole image as features input vector



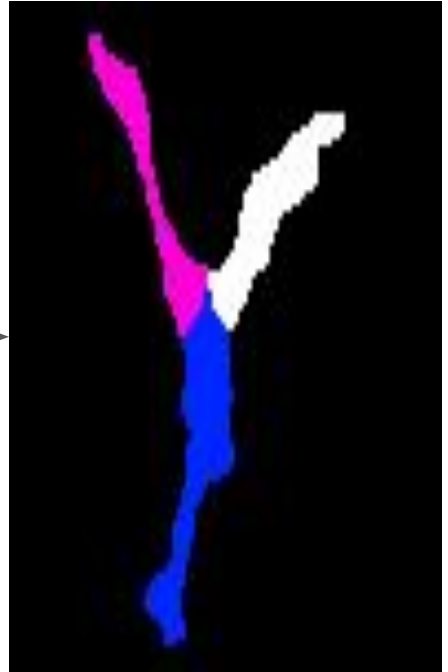
Death or censored
status as binary output

Convolutional layers followed by a dense layer were constructed to solve the problem.

We report 55% accuracy on validation set.

For completeness, a full coloured image was taken as an input vector and a larger convolutional network was created to predict the binary output. No accuracy improvement was seen.

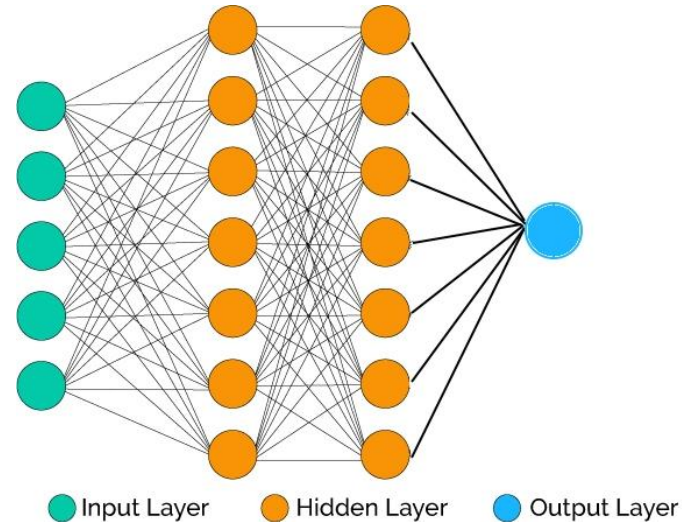
Computing features by hand: average collagen fiber length, width, etc. 16 features in total.



Back to binary classification problem:

this time used 16 features as an input vector per patient (rather than the whole image).

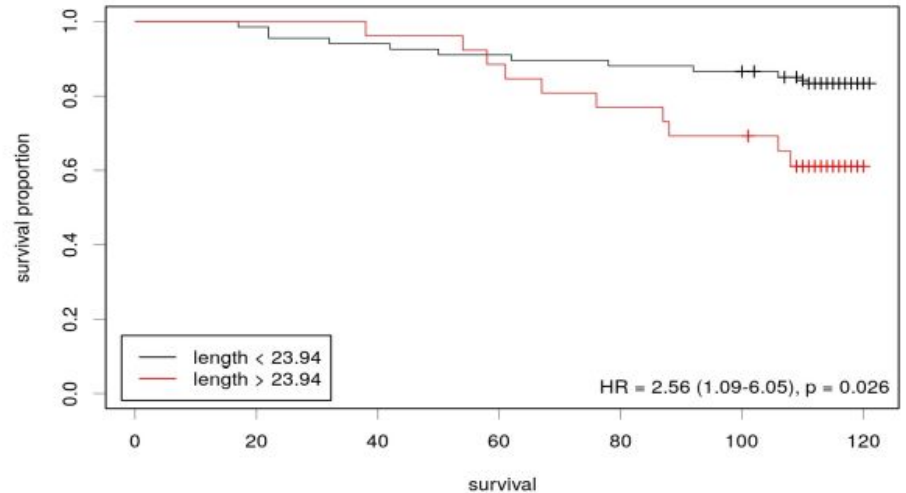
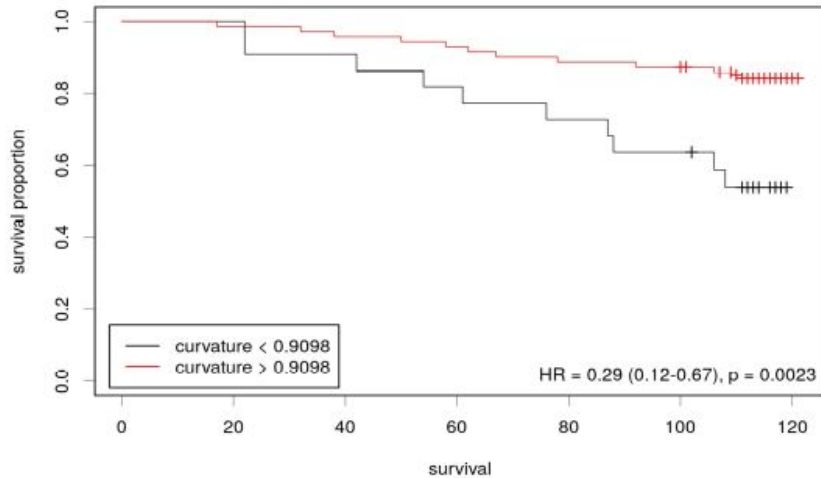
Designed a dense neural network to solve the binary classification problem:



Unfortunately, the accuracy did not increase this time either.

Regression Problem

We performed a separate Kaplan-Meier analysis for each feature to report that the curvature and length separate the dataset into two groups with p-value < 5%.



To test how the two features jointly influence the survival outcome, we performed the Cox regression using the two variables.

We report the following results:

Likelihood ratio test = 8.87 on 2 df, p=0.01
Wald test = 9.58 on 2 df, p=0.008
Score (logrank) test = 10.84 on 2 df, p=0.004

	coef	exp(coef)	se(coef)	z	Pr(> z)
curvature	1.0342	2.8129	0.4810	2.150	0.0315
length	-0.5479	0.5782	0.4813	-1.138	0.2549