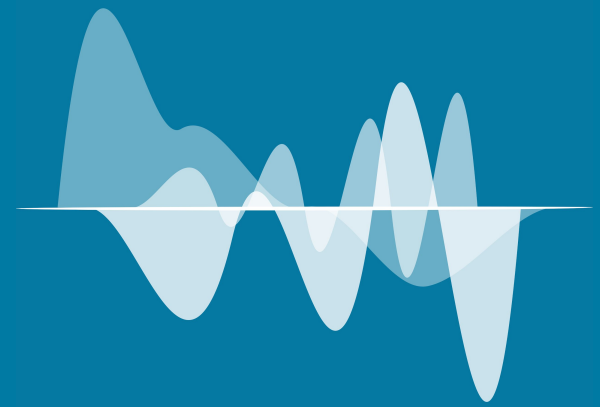# Artificial

Intelligence in

Medicine

Challenges in ML analysis implementation on RT data
part 2

Leonardo Ubaldi

Summary

- ## Datasets analysis

  - Analysis of the two datasets ('public' and 'proprietary')

- ## Features analysis

- ## Tasks:

  - Regression to predict survival time
  - Classification to predict different histologies (three categories)   } in progress

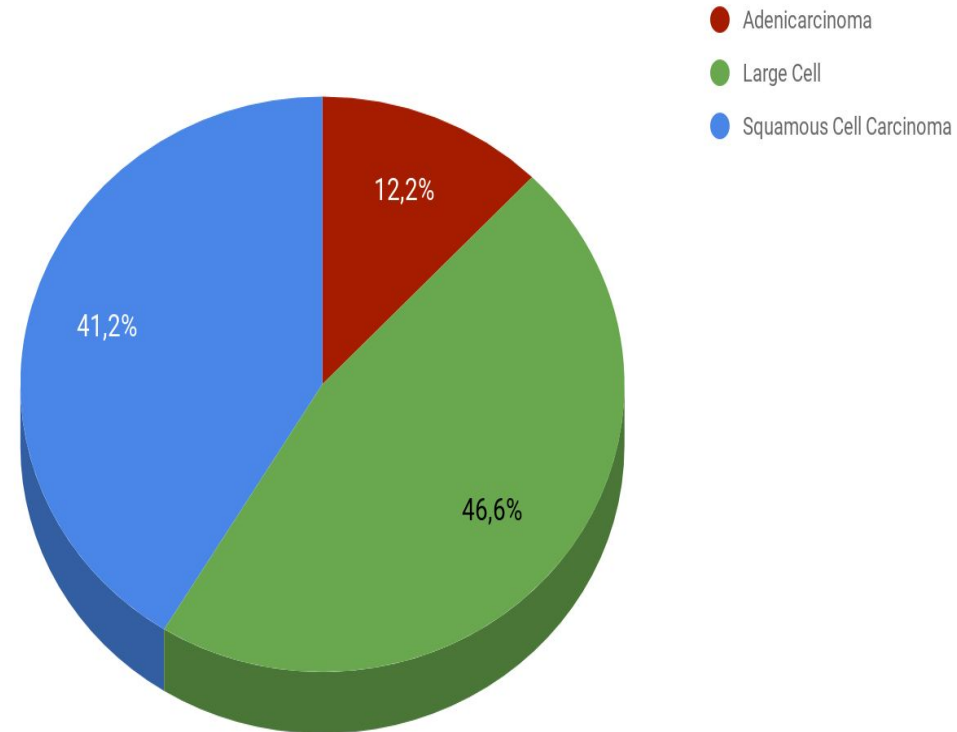# Public Dataset (Lung1 Maastro NSCLC)

131 subjects

Histology:

- Adenicarcinoma → 16
- Large Cell → 61
- Squamous Cell Carcinoma → 54

Overall stage:

- I   → 27
- II   → 13
- IIIa → 37
- IIIb → 53

Legend:
- Adenicarcinoma
- Large Cell
- Squamous Cell Carcinoma

12,2%
46,6%
41,2%

Aerts et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat Commun. 2014 Jun 3;5:4006. doi: 10.1038/ncomms5006.
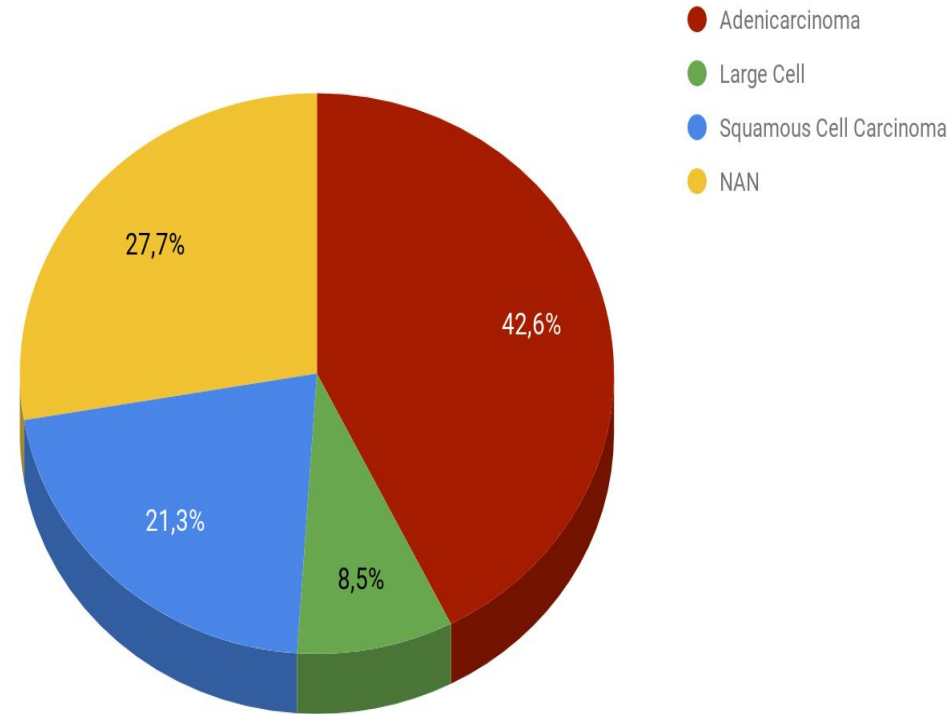
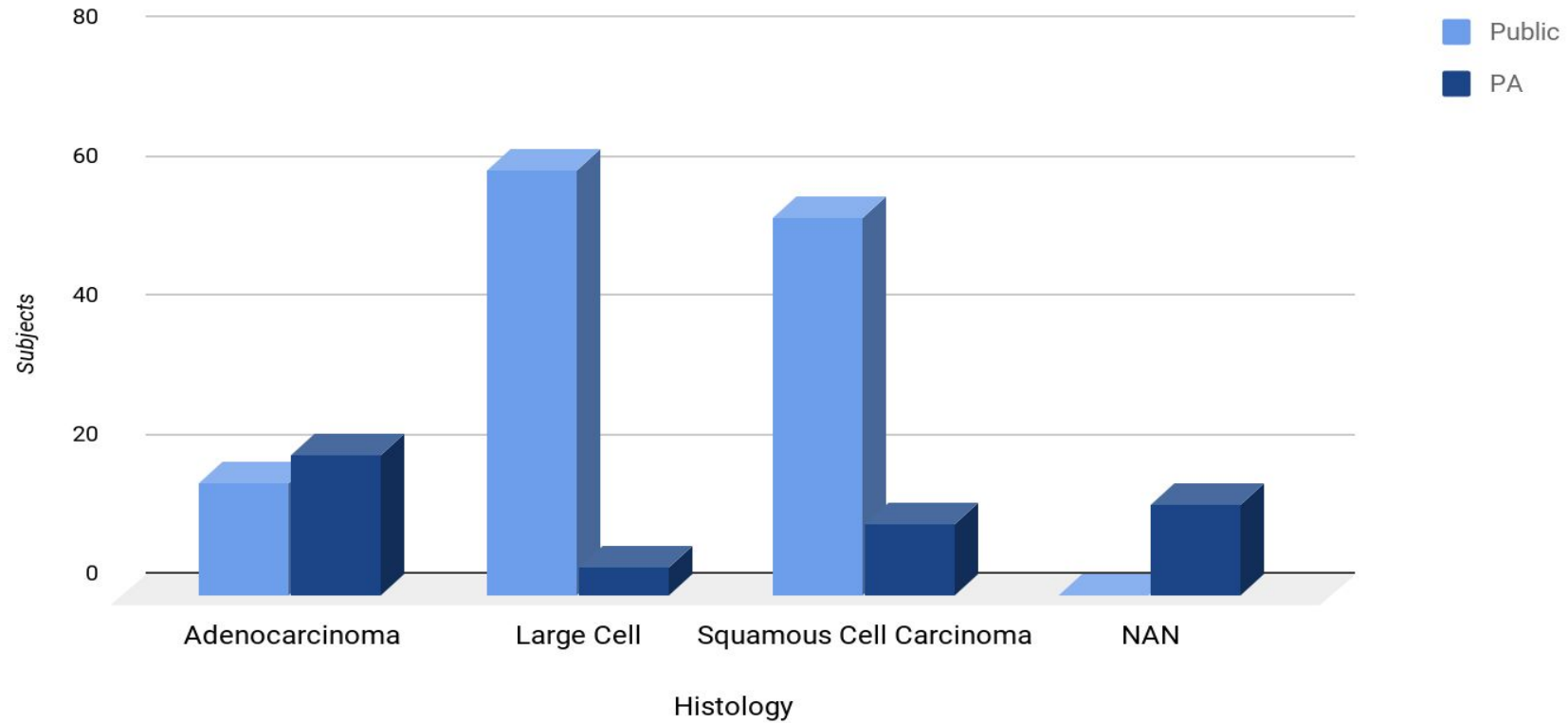# Dataset PA

47 subjects

Histology:

- Adenocarcinoma → 20
- Large Cell → 4
- Squamous Cell Carcinoma → 10
- NAN → 13

Overall stage:

- I → 42
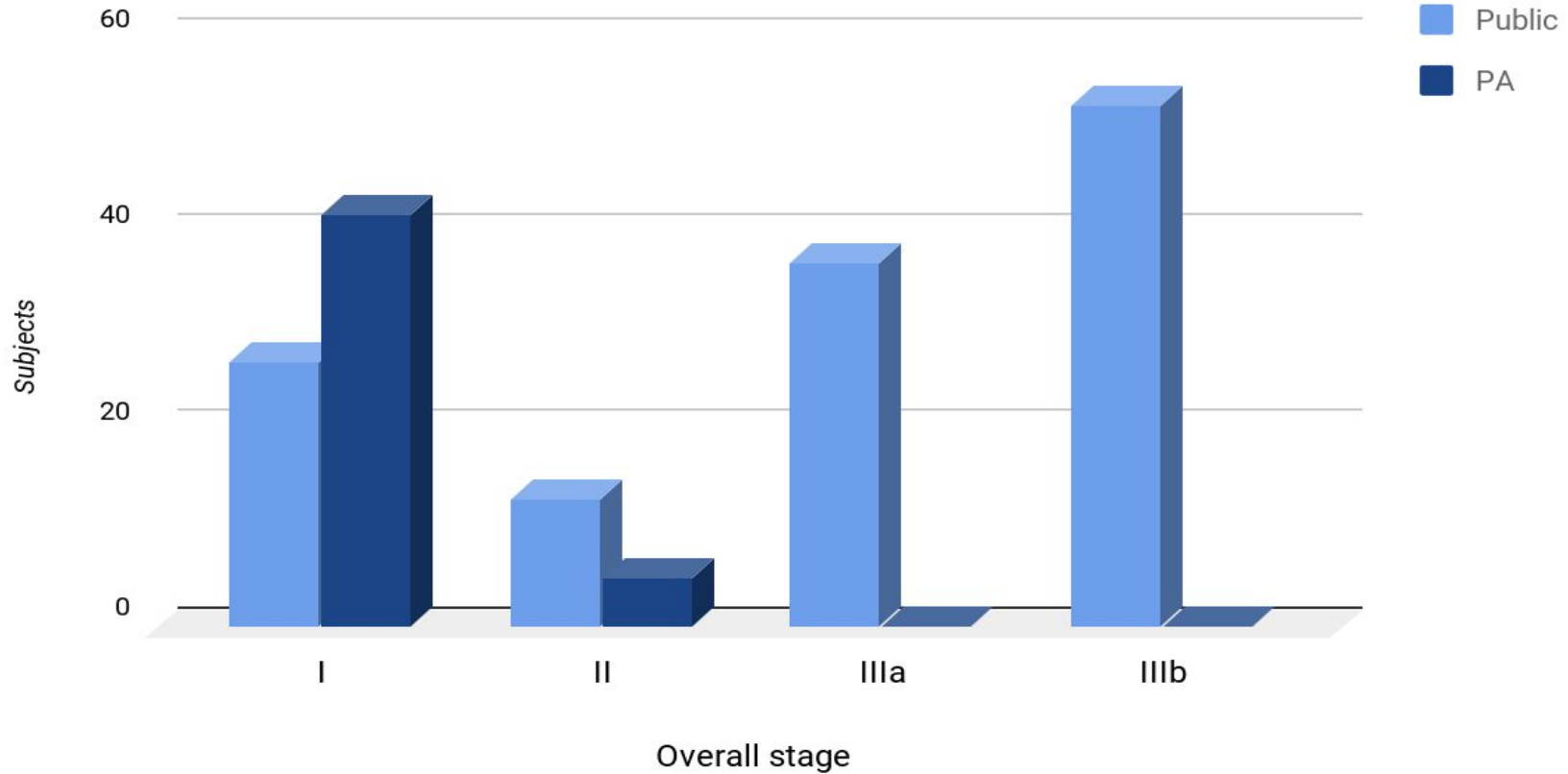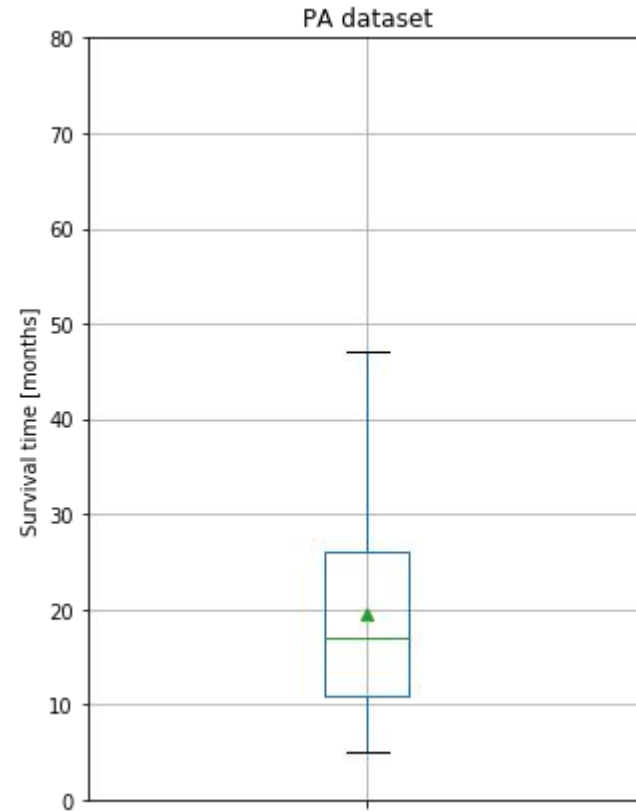- II → 5

- Adenicarcinoma
- Large Cell
- Squamous Cell Carcinoma
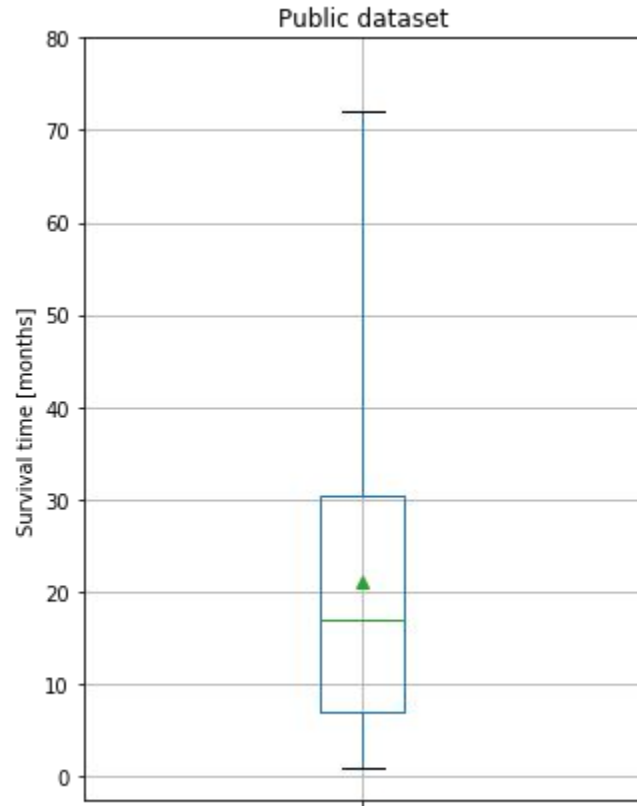- NAN

42,6%
27,7%
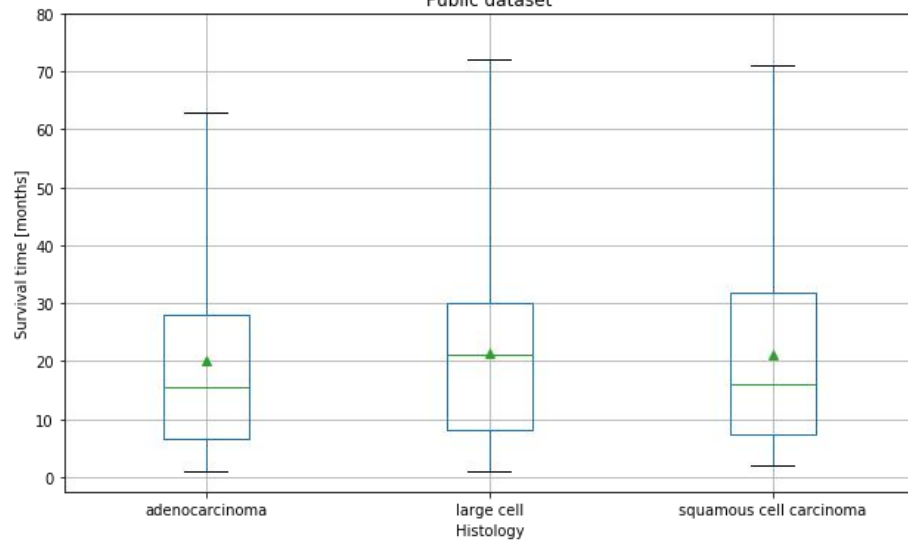21,3%
8,5%

Histology

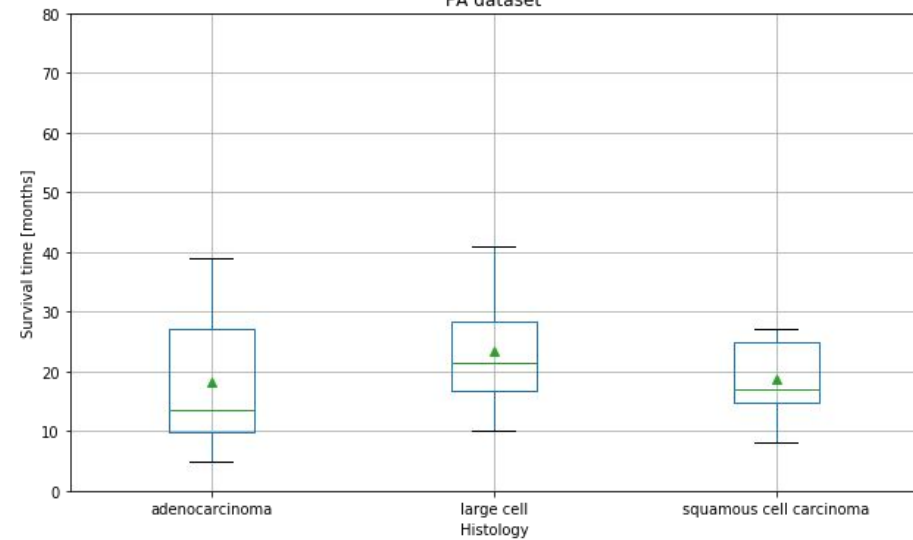# Overall stage

Boxplot Survival time
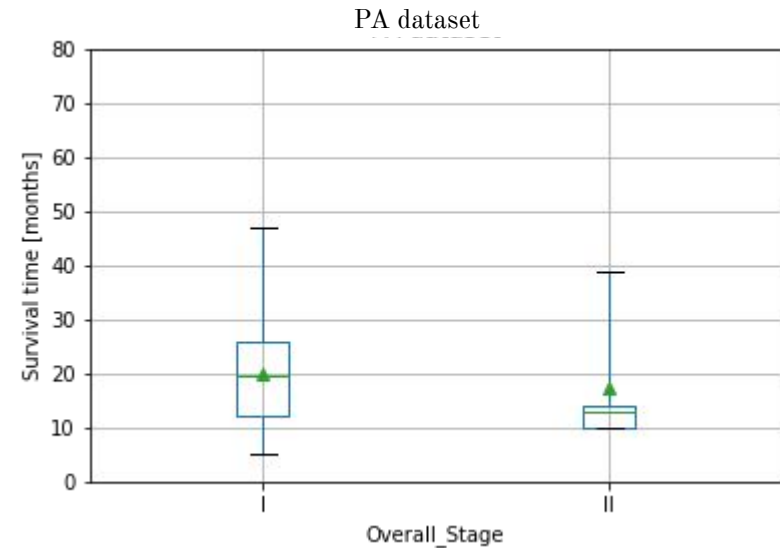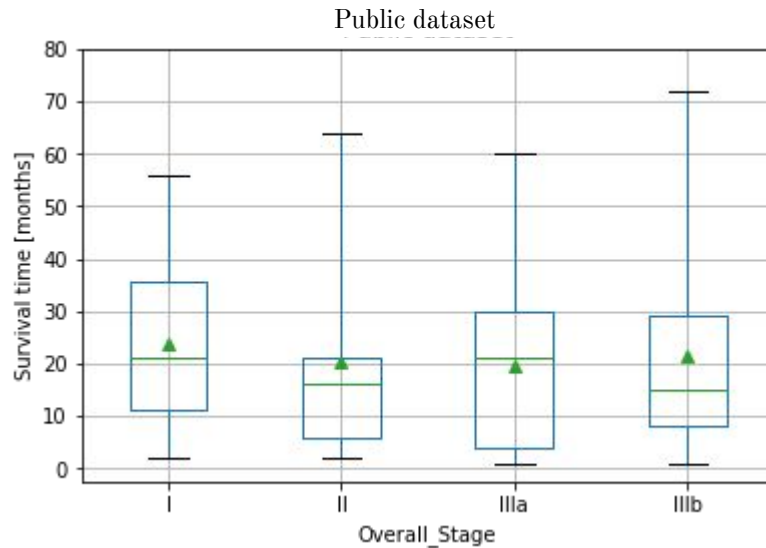
Boxplot Survival time by Histology

Boxplot grouped by Histology
Pubilc dataset
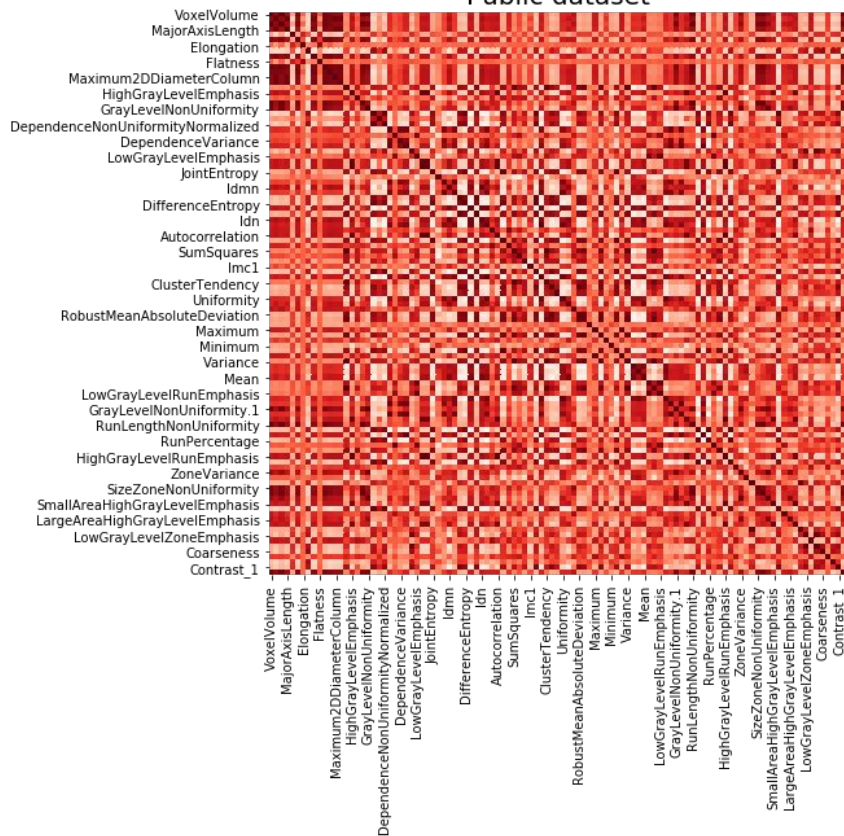

Boxplot grouped by Histology
PA dataset

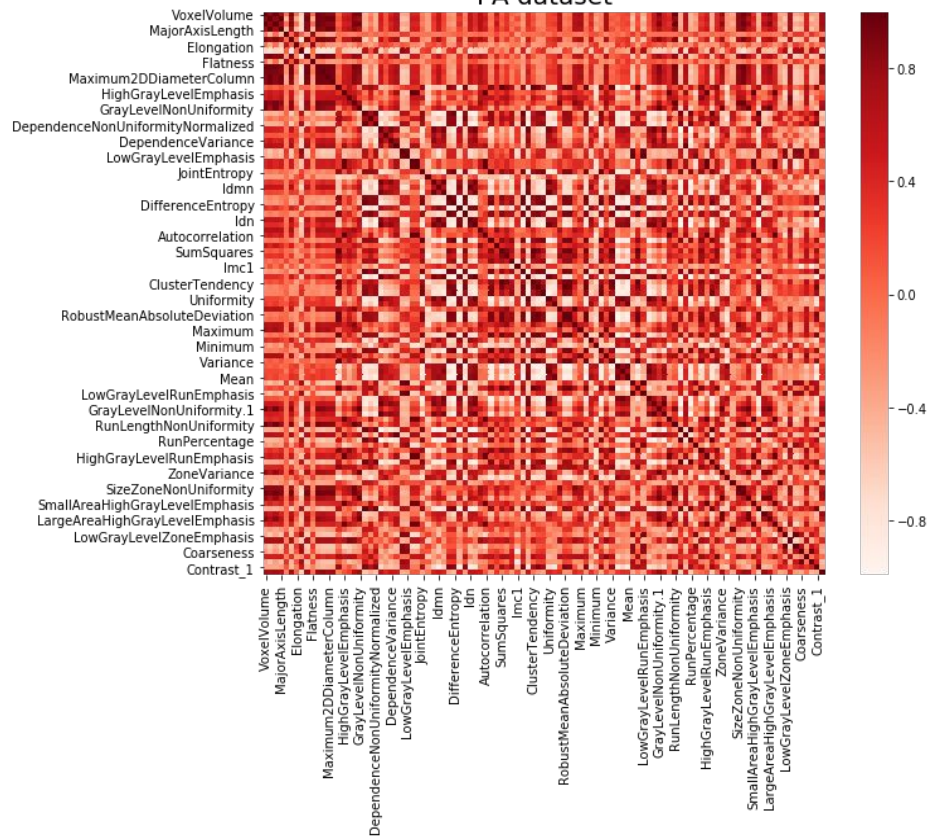# Boxplot Survival time by Overall Stage

Pearson correlation matrix
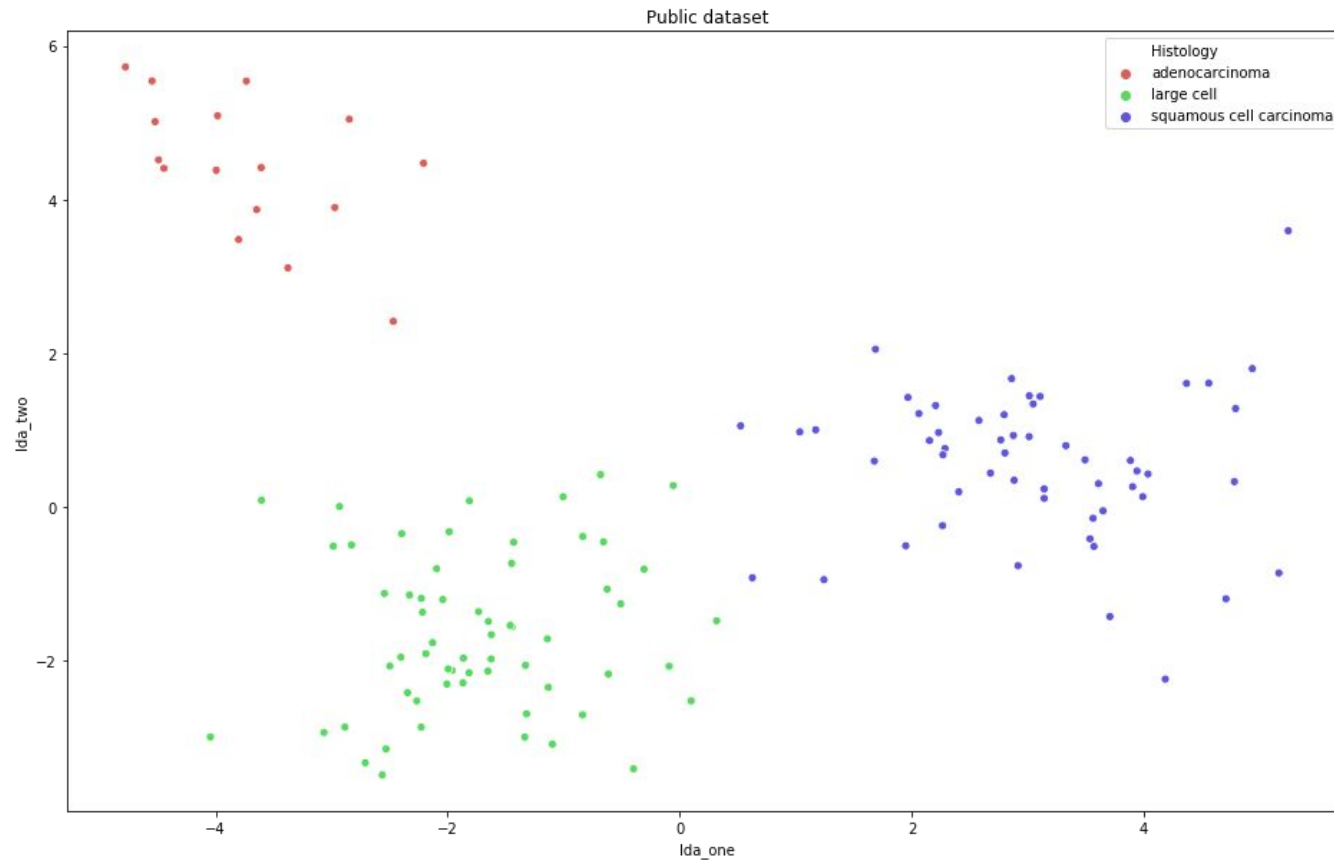

Public dataset

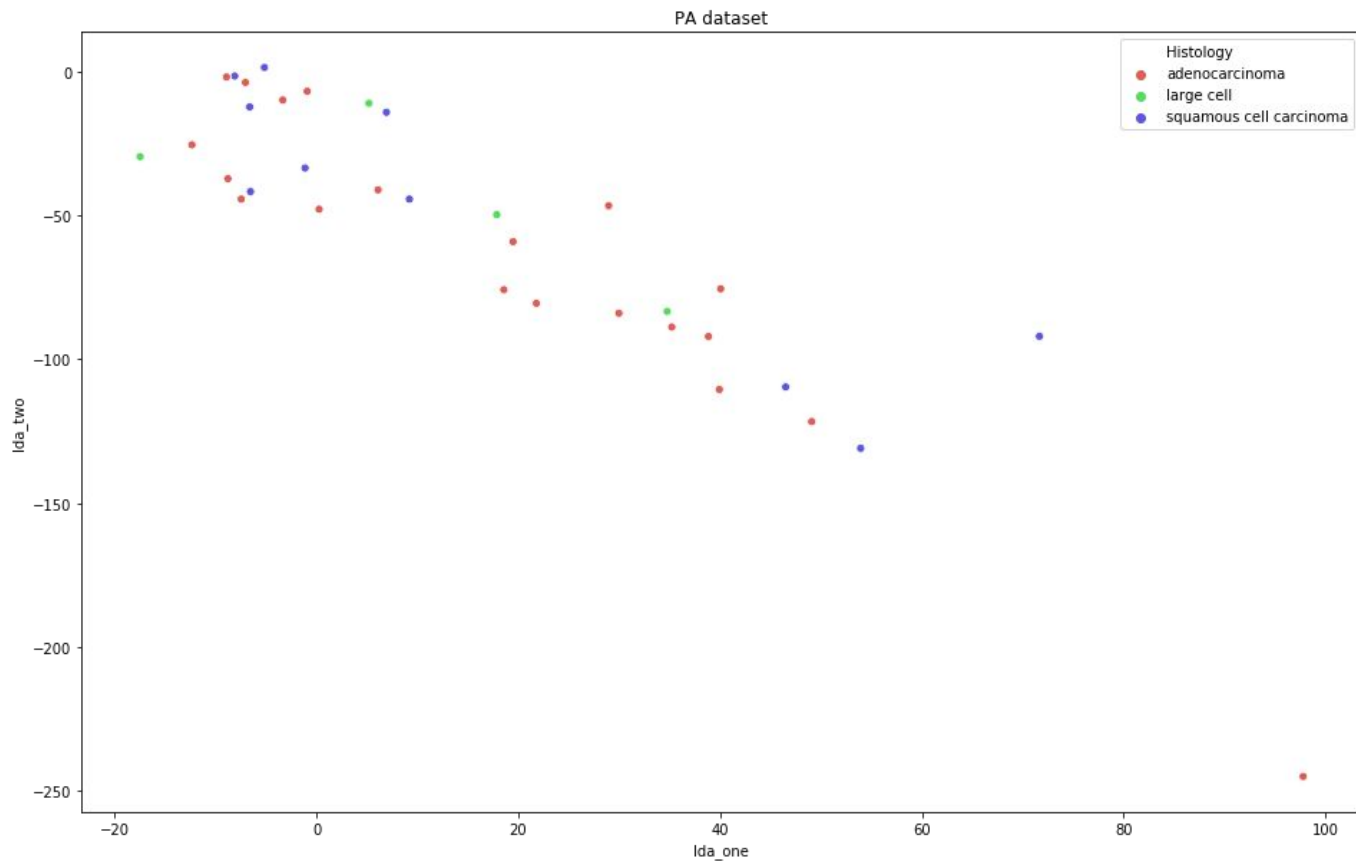PA dataset

Dimensionality reduction

- ## Regression to predict survival time
  - ○ dimensionality reduction with PCA
    - ■ The first 9 PCA components can explain $\sim 90\%$ of the variance of dataset

- ## Classification to predict different histologies
  - ○ dimensionality reduction with LDA:
    - ■ Linear and supervised,
    - ■ Maximize separation between classes
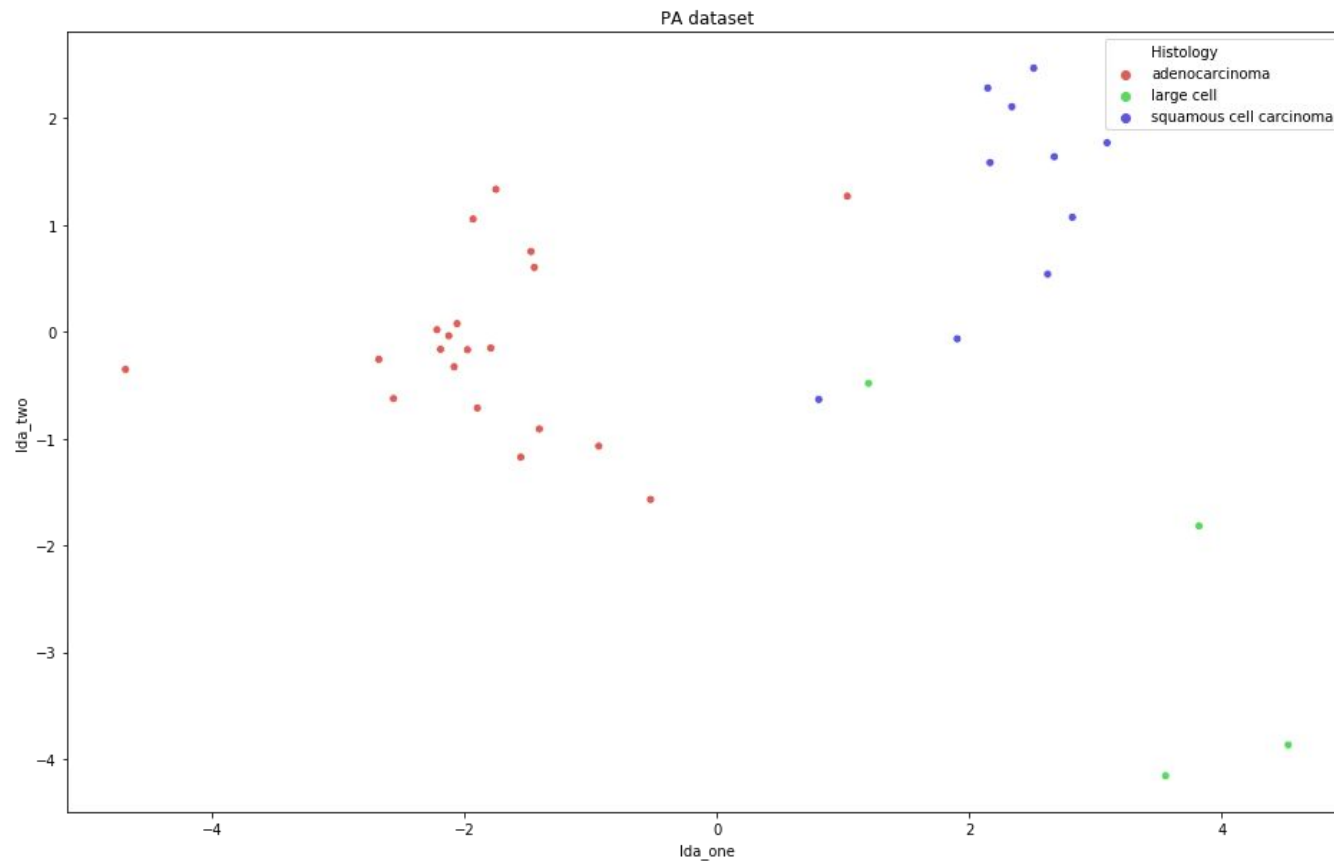   - ■ 2 components (max using 3 classes in Scikit-learn)

Visualization using LDA

Public dataset

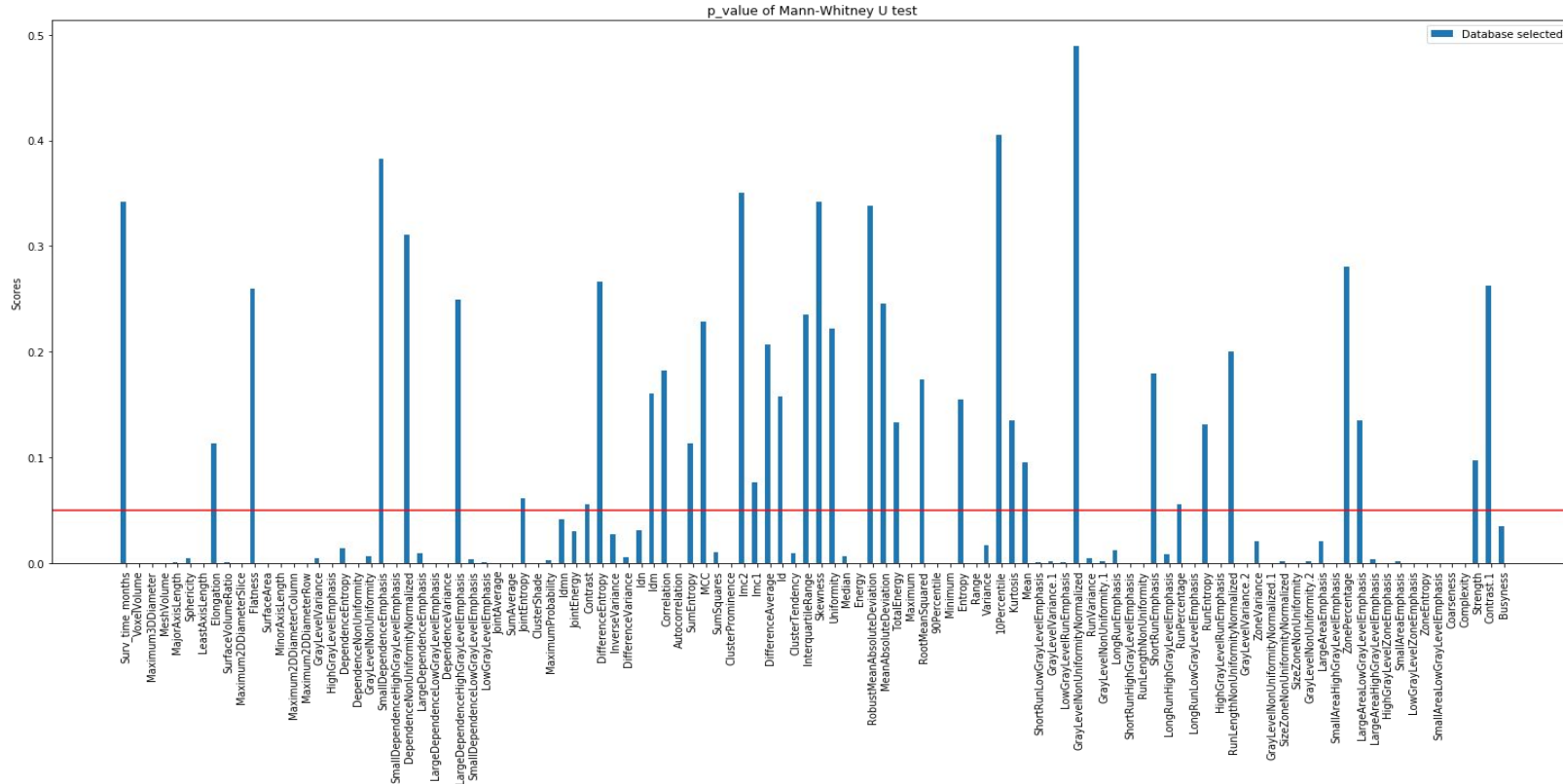# Visualization using LDA

PA dataset

# Visualization using LDA (new fit)

PA dataset

Can datasets be merged? Mann-Whitney U test

- Overall stage I and II selected from public dataset (40 subjects)
- The Mann-Whitney U test provided a $p > 0.05$ only for $37/108$ features

p_value of Mann-Whitney U test

Mann-Whitney U test with Bonferroni correction

- The Mann-Whitney U test with Bonferroni correction provided a $p > 0.05/108$ for 86/108 features
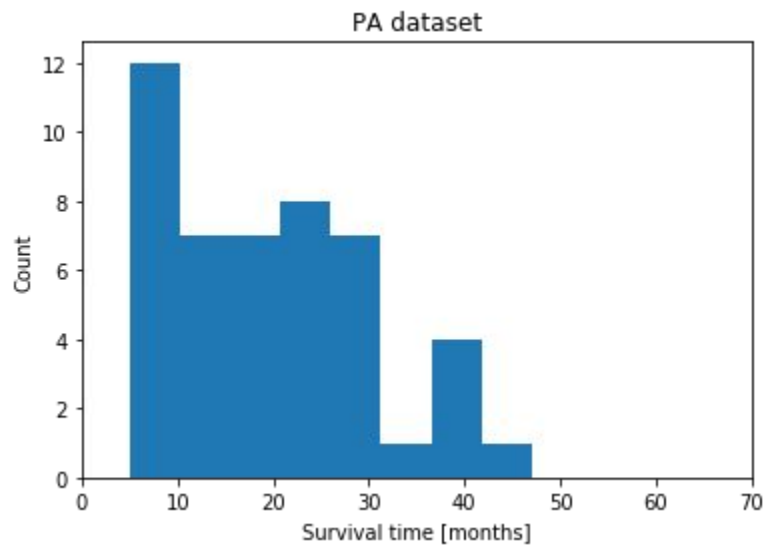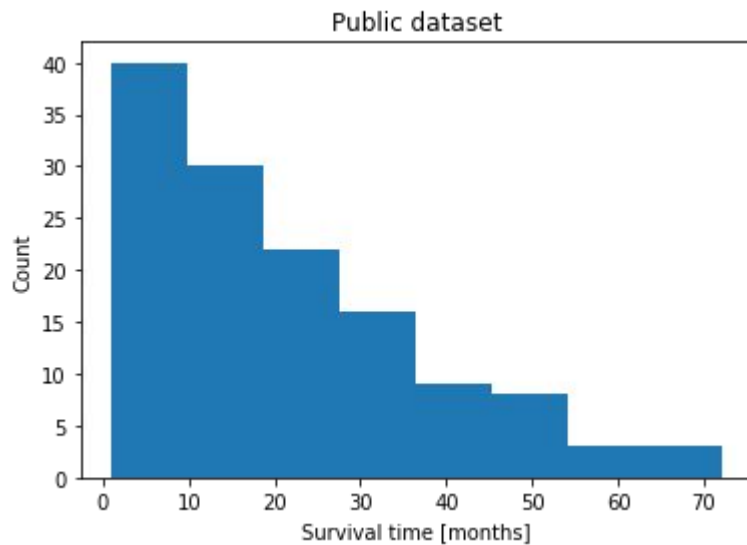

p_value of Mann-Whitney U test
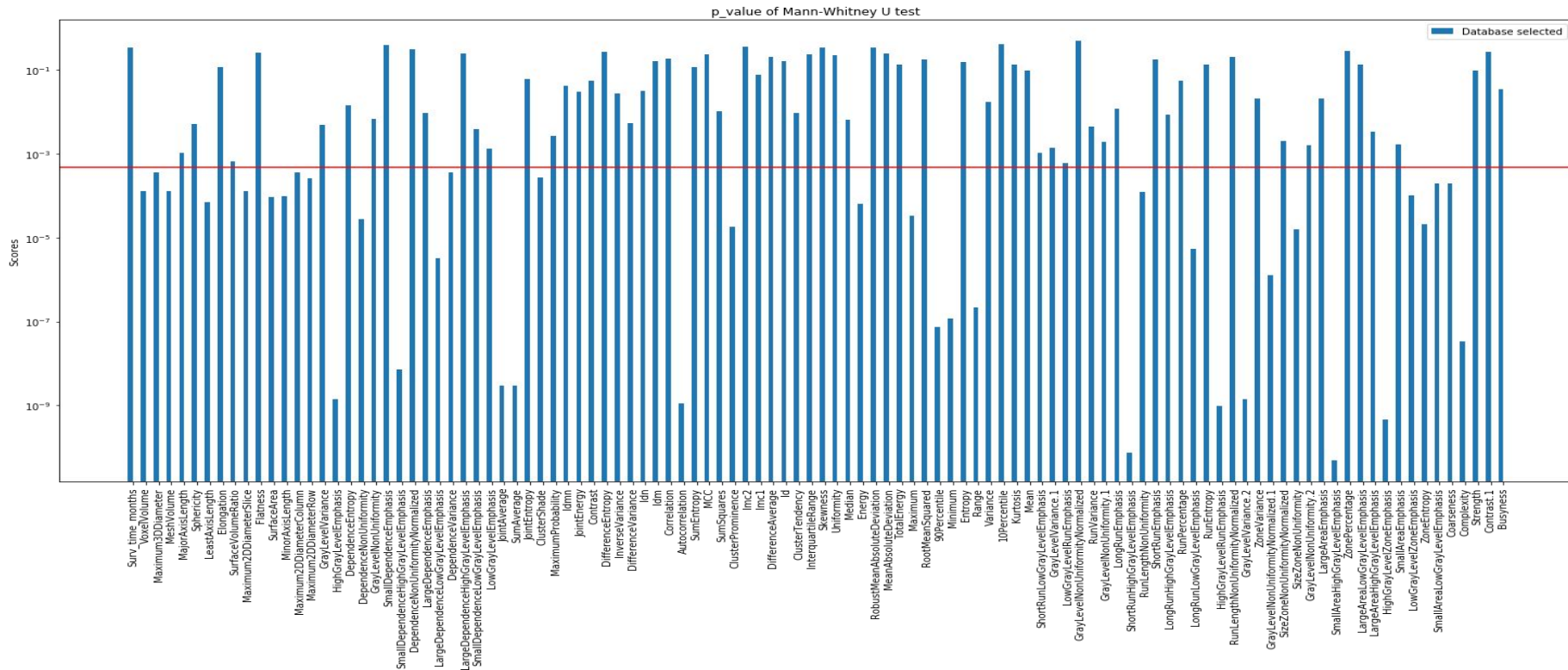
Next steps:

- Regression to predict survival time
  - Using neural network

- Classification to predict different histologies
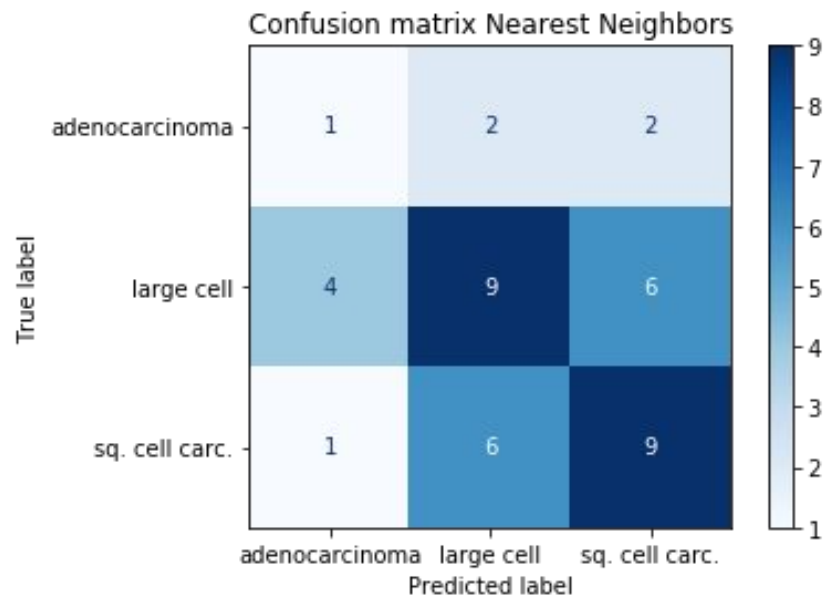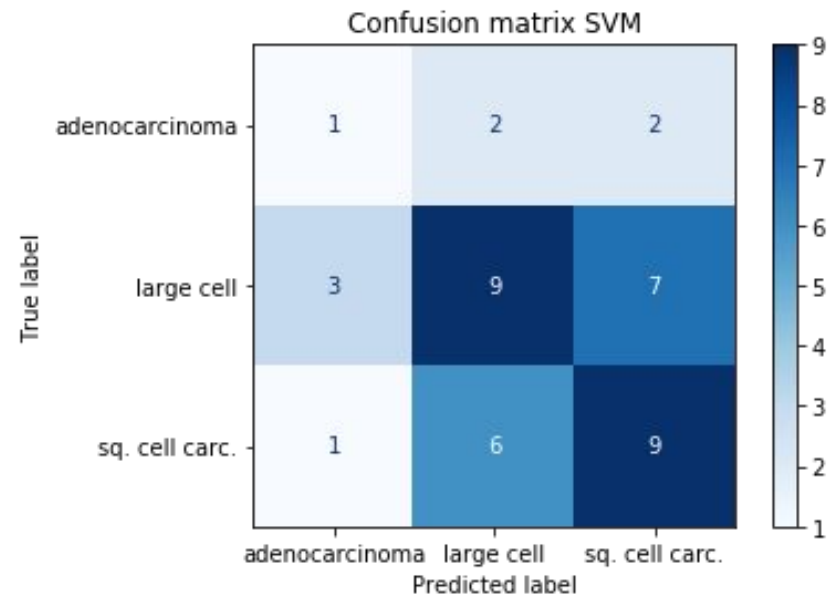 - Using Nearest Neighbors, SVM, neural network

thank you

Histograms Survival time

Confusion Matrix Public Dataset

Accuracy 0.475

Accuracy 0.475

70% training (91 subjects), 30% test (40 subjects)