

# Data for the IA2-Project: Diagnosing oral cancer

**Data is not for public distribution!**

Patient	Glass	Diagnosis	Cells, size 80x80
1	3	Healthy	1965
1	4	Healthy	1851
2	5	Healthy	1382
2	6	Healthy	956
3	7	Healthy	863
3	8	Healthy	738
4	12	Cancer	226
5	36	Cancer	534
6	37	Cancer	963
6	38	Cancer	796

Do use the following splits of training and test data (where no patient exists in both training and test set):

Set	Glasses used for training	Glasses used for evaluation
Set 1	3,4,5,6,12,37,38	7,8,36
Set 2	3,4,5,6,36,37,38	7,8,12
Set 3	3,4,5,6,12,36	7,8,37,38

The data set 43MB (**Not for public distribution!**) can be downloaded here (user: ia2project, pass: CancerCells): <http://www.cb.uu.se/~joakim/OralData/slides.zip>

The data consists of cut out cells of size 80x80 pixels. Each glass is stored in one hpf5 file, which can be read in Matlab using `cells=h5read('glass3.hdf5','/data');`

Ask if there are any problems whatsoever. Discussions are of course also very welcome! :-)

Good luck!

/ Joakim