

Video: Sampling

~ 10K or 100K
Patients with
Labels

image	patient_id	label	...
Xray10.jpg	15	0	
Xray23.jpg	24	0	
Xray31.jpg	20	0	
Xray41.jpg	86	1	

Sample Test set
then Validation
the Train

Sample Testset

In Studies
Testset with atleast X% of
minority class
Sometimes set to 50%

Dataset

10%

Testset

image	patient_id	label	...
Xray4.jpg	17	0	
Xray30.jpg	39	1	
Xray5.jpg	11	0	
Xray8.jpg	18	1	

Sample Validation Set

In Studies
Sample to have same
distribution of classes
as the test set.

Dataset

10%

Validation Set

image	patient_id	label	...
Xray2.jpg	15	0	
Xray12.jpg	26	0	
Xray37.jpg	31	1	
Xray55.jpg	41	1	

Sample Train Set

In Studies
Remain Patients in
Training Set

Dataset

Train Set

image	patient_id	label	...
Xray16.jpg	12	0	
Xray3.jpg	21	0	
Xray13.jpg	25	1	
Xray43.jpg	36	0	