

Applied Statistics A.A. 2019/2020

# Automated Chest X-Ray Diagnostics

Francesco Giammaria

Stefano Gusmeroli

Edoardo Peretti

Matteo Savino

Alessio Tranchida





# This is us



**Matteo Savino**



**Stefano Gusmeroli**



**Edoardo Peretti**



**Alessio Tranchida**



**Francesco Giammaria**



# DATA PRESENTATION



# DATA PRESENTATION

More than 60,000 patients

More than 180,000 images

14 pathologies



input



Deep Convolutional Neural Network  
(CNN)

output



Embeddings





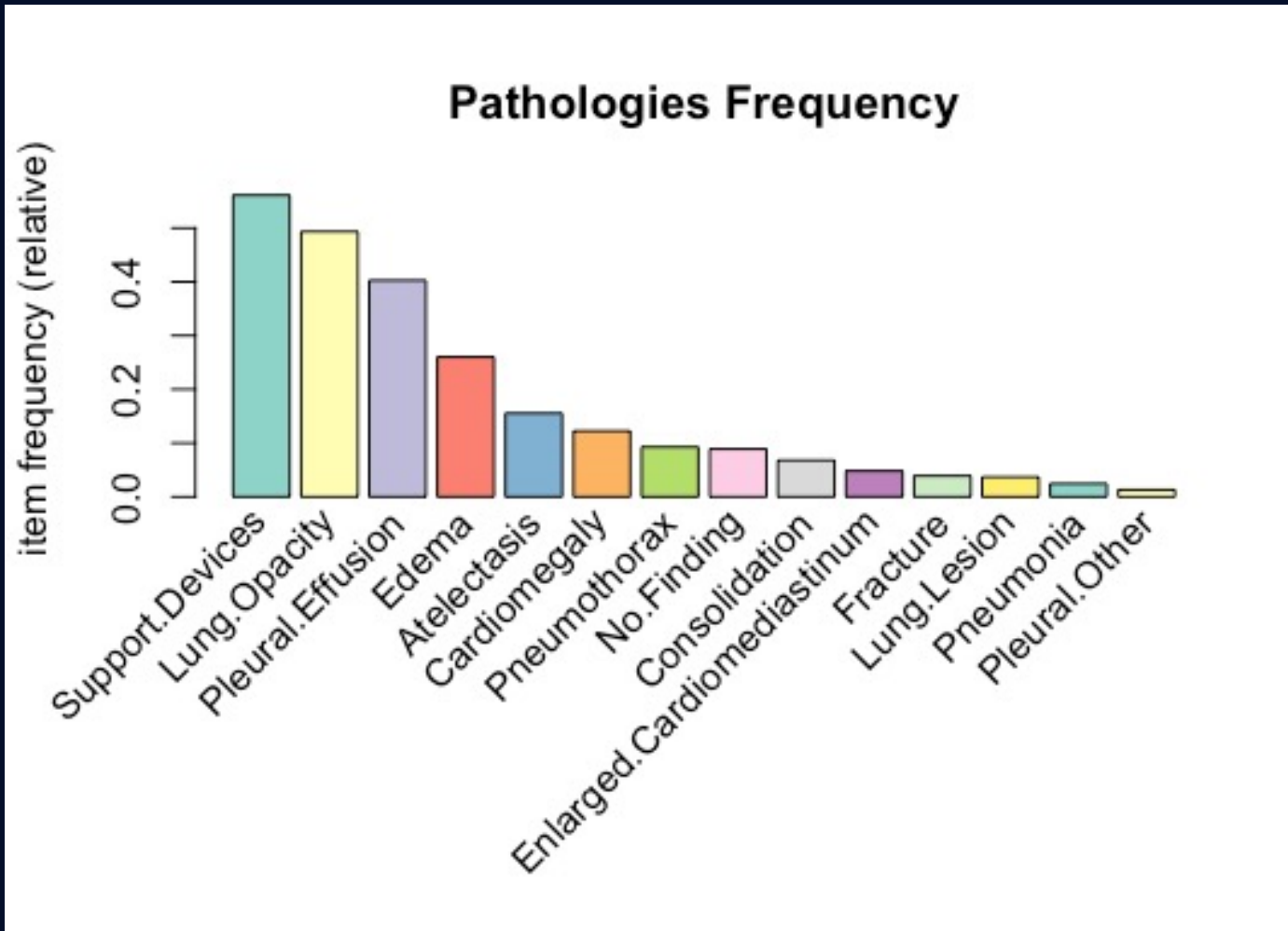
## Embeddings

(189116 x 931)

	input_data.Path	X2	X3	X4	X5	X6	X8
1	CheXpert-v1.0-small/tra...	5.044617e-08	0.009756592	0.005829324	0.13361263	0.6152759	0
2	CheXpert-v1.0-small/tra...	0.000000e+00	0.008479902	0.005943561	0.10519153	0.6668760	0
3	CheXpert-v1.0-small/tra...	0.000000e+00	0.009295503	0.007226741	0.10922829	0.5146037	0
4	CheXpert-v1.0-small/tra...	0.000000e+00	0.008557481	0.006199519	0.09658385	0.8638877	0
5	CheXpert-v1.0-small/tra...	6.659148e-06	0.004773823	0.005670027	0.14989142	0.5566312	0
6	CheXpert-v1.0-small/tra...	0.000000e+00	0.009678223	0.006153073	0.13729700	0.5397011	0
7	CheXpert-v1.0-small/tra...	0.000000e+00	0.008135295	0.005715157	0.05424061	0.3945267	0
8	CheXpert-v1.0-small/tra...	0.000000e+00	0.009958112	0.006867255	0.07581601	0.9344040	0
9	CheXpert-v1.0-small/tra...	5.897515e-06	0.007355212	0.006465926	0.17092736	0.5021766	0



## 14 pathologies







## Labels

**Radiologic Consultation Report**

REQUEST DATE: 8/5/2015  
PATIENT: NAME: [REDACTED] REPORT DATE: 8/5/2015 SERIAL #: 11212  
DOB: 11/07/2003 AGE: 12 GENDER: Female  
REFERRING BY: Dr. Smith  
CLINICAL: Evaluation of suspected rhabdomyosarcoma and assessment of tooth  
INFO: 17 and 18.  
TYPE OF IMAGE REPORTED: Both codes  
REQUEST TYPE: Dental Inspection C/CT

Dear Dr. Smith:

Thank you for referring Mr. [REDACTED] for CBCT volumetric imaging performed on 8/5/2015. Interpretation of the volumetric dataset was requested on 8/5/2015 and performed on 8/6/2015. This evaluation is provided to assist you in the management of these lesions for your patient. Thank you for this opportunity to assist you with your patient's diagnosis. Please

**RADIOLOGIC FINDINGS**

**GENERAL FINDINGS:** The scan demonstrates a dentate child with developing maxillary and mandibular third molars, and impacted tooth 16. No endodontic dental implants are present. No teeth have root canal treatment. There is no evidence of periodontal pathology associated with any of the teeth. The marginal alveolar bone height is within normal limits.

**SPECIFIC FINDINGS:** Tooth 16 is completely developed, immotivated and mesio-angulated in almost horizontal position. The crown of 16 is in the middle of the alveolar ridge and is buccally positioned to the apex of tooth 17. The apex of 17 shows significant angulation, which may be related to tooth rotation or initial periodontal resorption by 16. The crown of 16 is located palatally to the apex of tooth 18. In the present time, signs of radiolucency of 16 are not observed. Signs of cystic degeneration of tooth 16 are not observed. The nasopharyngeal duct is in its lower position to the crown of 16 with partial filling and is laterally displaced (see axial view). The root of 16 is approximately 1.5 mm superior to the apex of the palatal root of 15. The apex of tooth 16 is located superiorly at the level of tooth 14 and is displaced buccally. In the mandible, mild overall displacement of teeth 422 and 27 is noted.

**INCIDENTAL FINDINGS:** The TMJ articulations are not within the field of view. The walls of the maxillary sinus demonstrate mild generalized mucosal thickening bilaterally. A small paranasal cyst is noted on the floor of the right sinus. The maxillary canals are blocked. There is mild mucosal thickening of the ethmoidal air cells. Sphenoidal sinus is clear. There are bilateral concha bullosa of the middle concha. There is mild nasal septal deviation to the L, reducing the associated lumen of the nasal fossa. The nasopharyngeal airway space dimension within the field of view is within normal limits. The cervical spine is outside the field of view.

**RADIOLOGIC IMPRESSION**

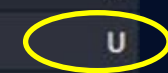
1) Tooth 16 is immotivated and impacted, and in contact with tooth 17 and 18. Mild resorption of the apex of 17 is suspected. Endodontic evaluation and follow-up of tooth 17 is recommended.  
2) Generalized chronic sinusitis. In the presence of symptoms, medical evaluation is recommended.  
3) Developing maxillary and mandibular 3rd molars in initial phase of radiolar formation.

Dr. [REDACTED] D.D.S., B.S., (Board) Dental: 8/6/2015  
Dental Imaging Consulting Services to Oral and Maxillofacial Radiology Only  
PLEASE NOTE: The radiologic findings and impressions are developed in consultation with Dr. Warren H. Shuman, DDS, MS, Dip. ABOMR, a board certified Oral and Maxillofacial Radiologist. The information and recommendations contained herein is based upon the provided history and imaging material, images and volumetric data set and is for consultation purposes only. In with all diagnostic imaging, some form of CT has diagnostic limitations. Diagnostic, medical advice and treatment is the responsibility of the treating physician or dentist.



	Cardiomegaly	↕	Lung.Opacity	↕	Edema	↕
1		0		1		1
2		0		1		0
3		0		1		0
4		0		U		0
5		0		1		U
6		1		1		0
7		0		1		0
8		1		1		1
9		0		1		0
10		U		1		0

## Uncertainties



# Guidelines

**Embeddings Analysis**

**Treatment of Uncertainties**

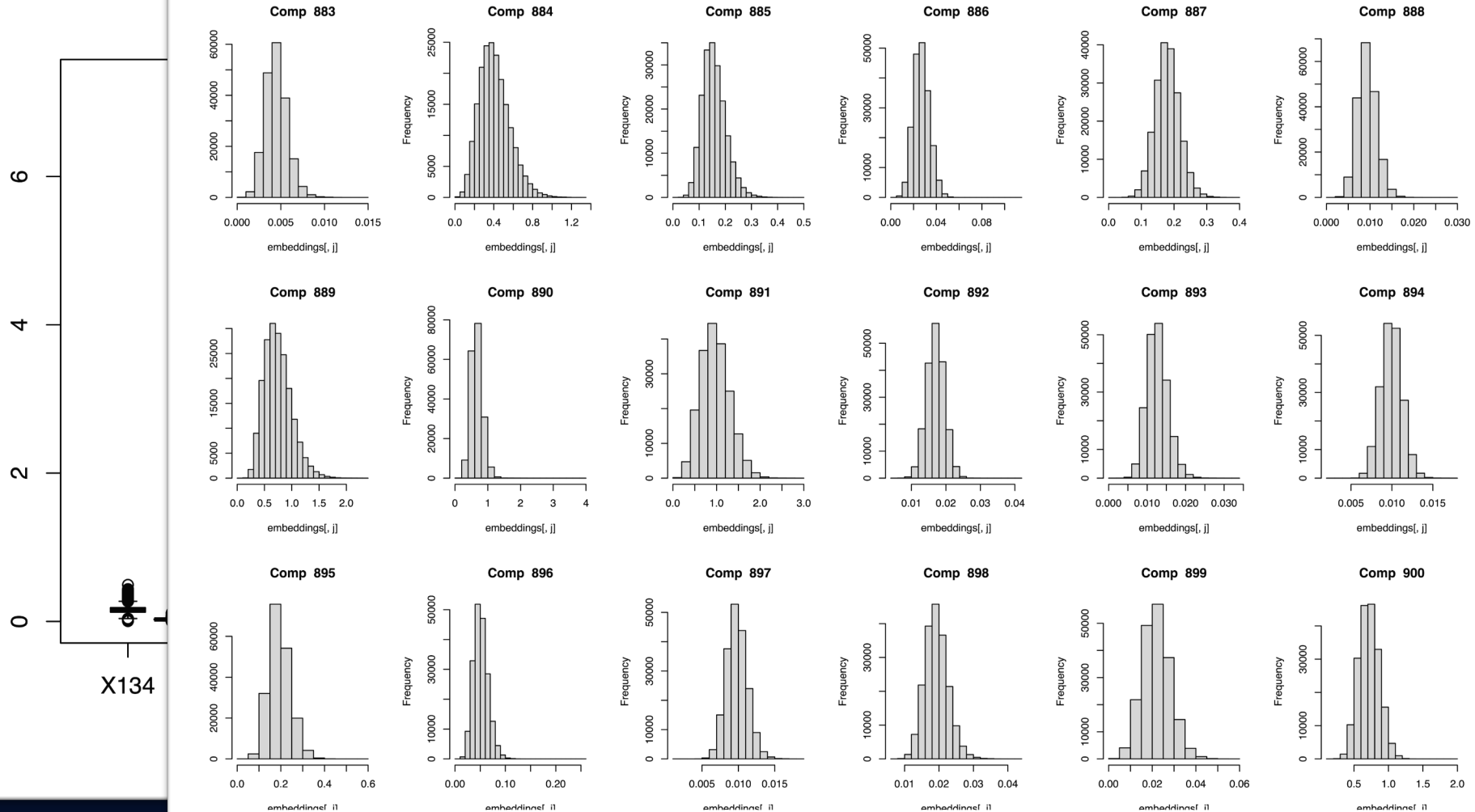
**Pathologies Relationships**





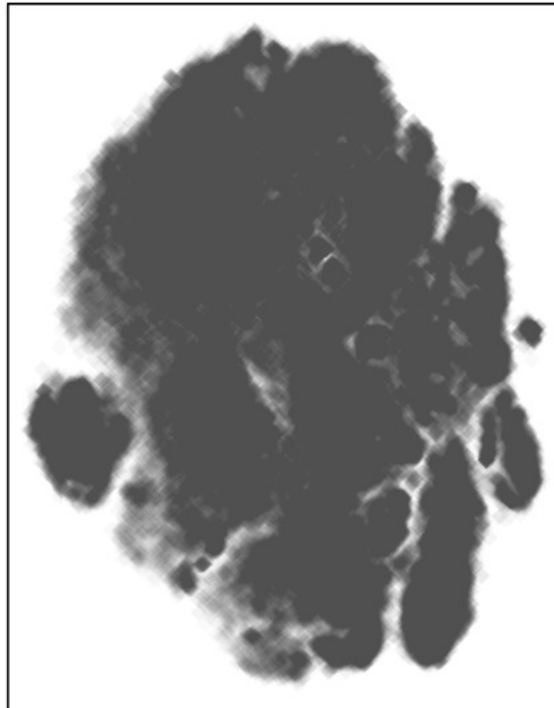
# GRAPHICAL REPRESENTATION



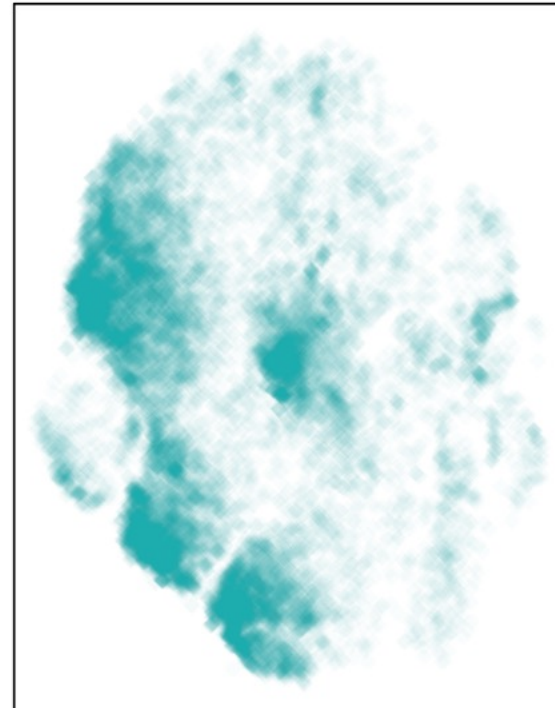


## t-SNE (t-distributed stochastic neighbor embedding)

4 Cardiomegaly negative

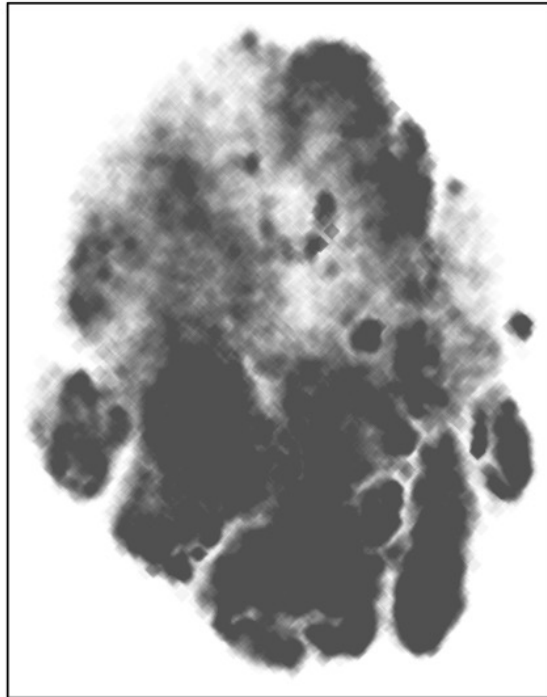


4 Cardiomegaly positive

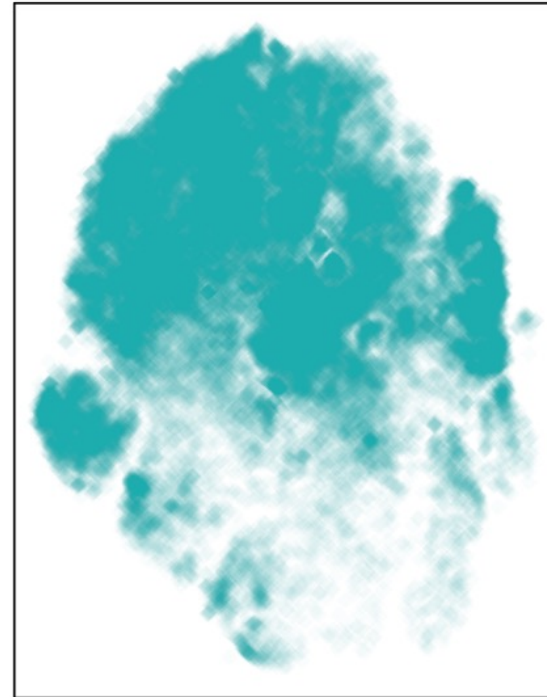


## t-SNE

12 Pleural.Effusion negative



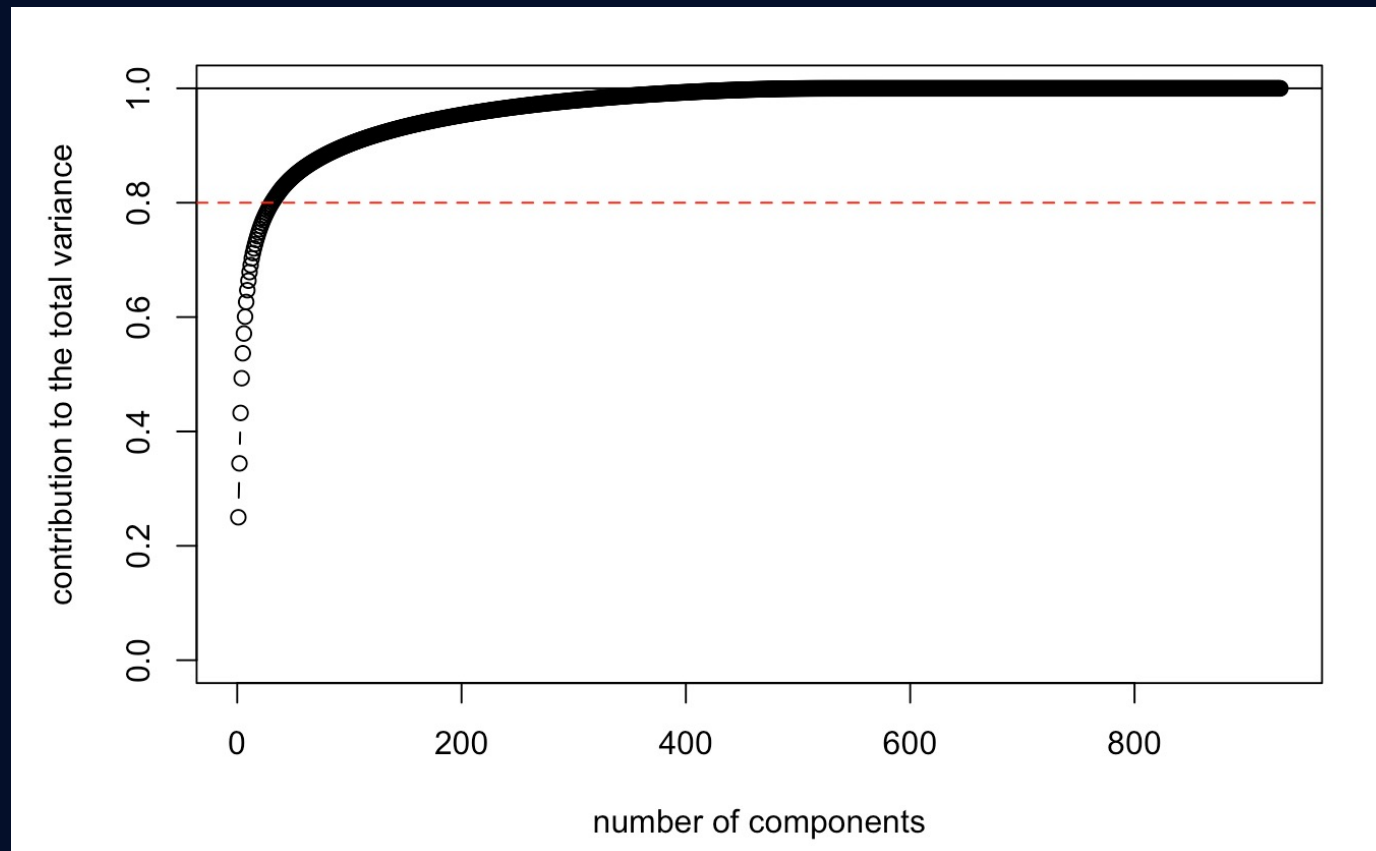
12 Pleural.Effusion positive



# CLASSIFICATION MODELS

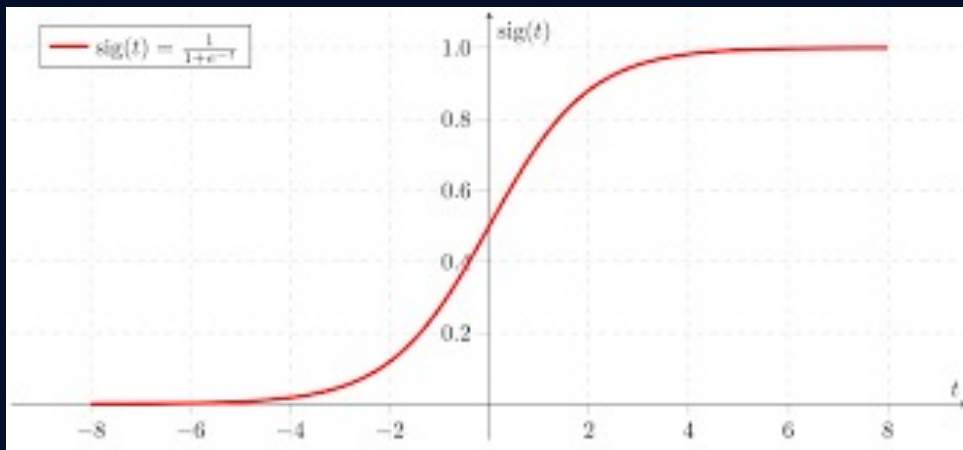


# DIMENSIONALITY REDUCTION → PCA



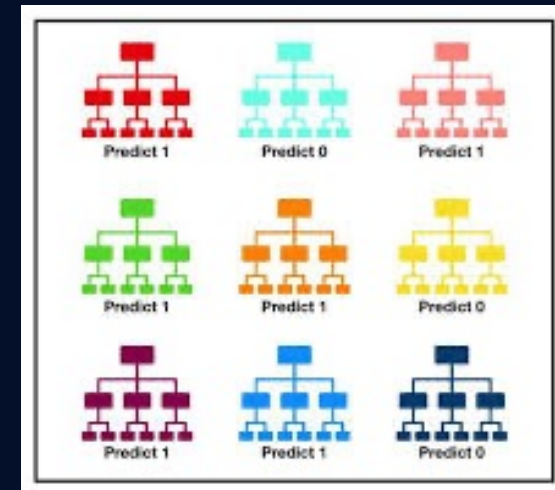
## LOGISTIC REGRESSION

- Bagging Model



## RANDOM FOREST

- 100 trees
- features random selection
- Quality of split: Gini index



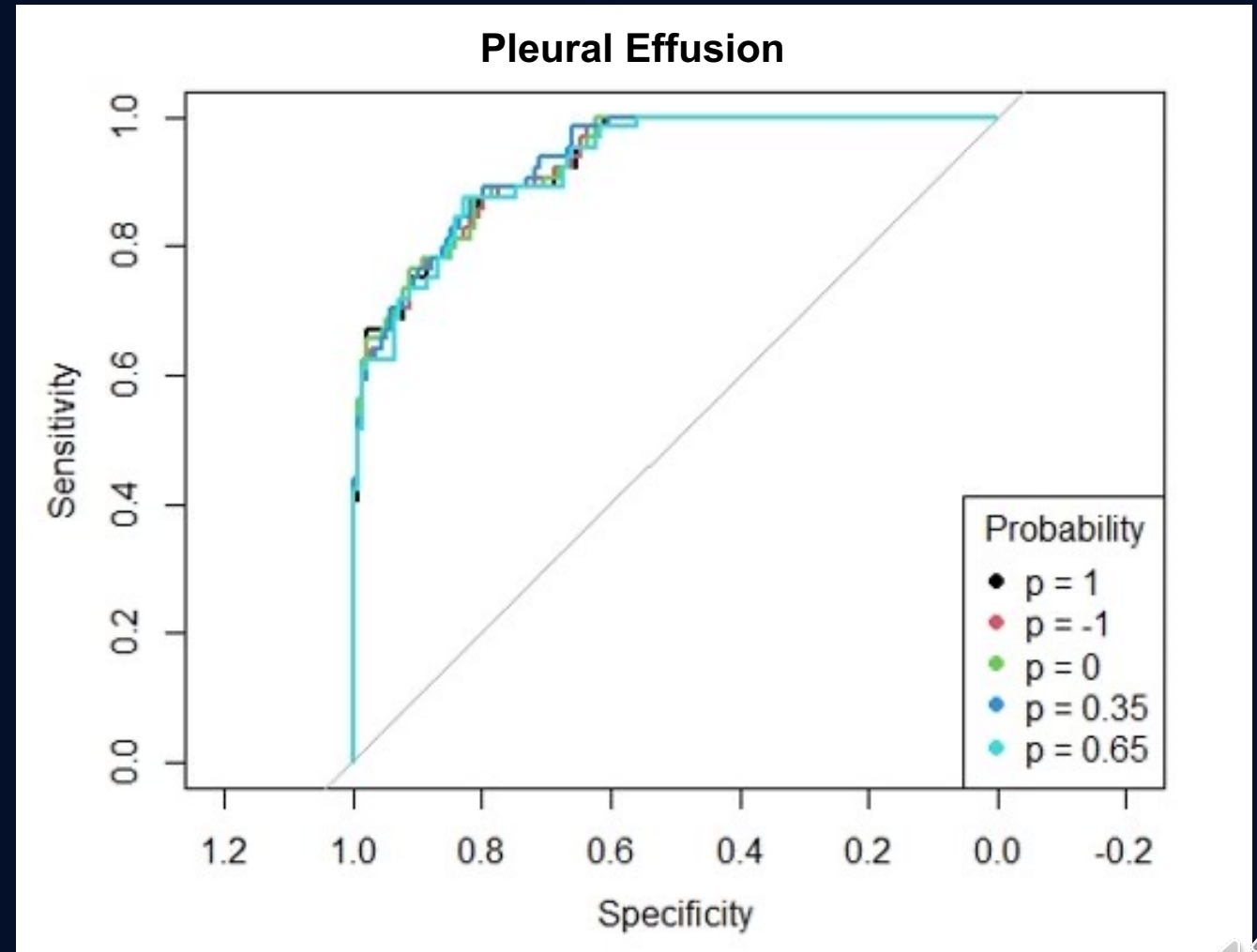


# Treatment of Uncertainties

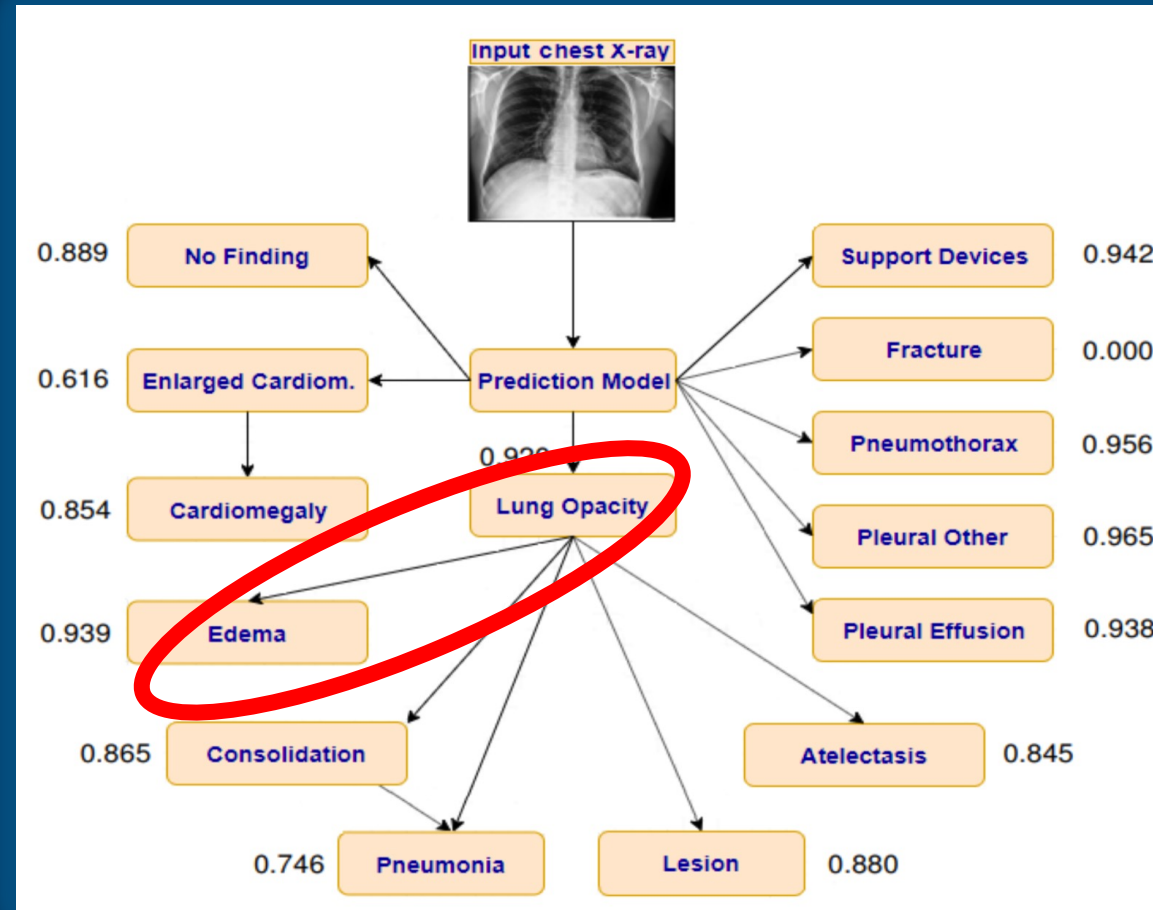
Uncertain values can be:

- Discarded ( $p = -1$ )
- All assigned to 0 ( $p = 0$ )
- All assigned to 1 ( $p = 1$ )
- Assigned to 1 with probability  $p$

(Logit Models for different uncertainties)



# ASSOCIATIONS



[Pham et al. 2019]



# Conditional Models

LUNG.OPACITY → EDEMA

EDEMA									Lung.Opacity
X345	X346	X347	X348	X349	X350	X351	X352	X353	
0.005	0.034	0.94	0.56	0.23	0.66	0.26	0.001	0.569	1
0.567	0.734	0.005	0.745	0.234	0.657	0.678	0.392	0.004	1
0.178	0.167	0.001	0.070	0.788	0.167	0.281	0.192	0.182	0
0.121	0.431	0.379	0.781	0.112	0.011	0.087	0.965	0.289	0
0.078	0.256	0.389	0.719	0.001	0.781	0.228	0.118	0.801	1



## Corrected Labels

Edema	Lung. Opacity	Cardiomegaly
1	u	1
0	u	1
0	0	1
1	0	0
1	1	1



Edema	Lung. Opacity	Cardiomegaly
1	1	1
0	u	1
0	0	1
1	1	0
1	1	1



## Conditional models:

increase in AUC up to 10÷20%, e.g. Cardiomegaly and Edema

Corrected labels: further increase of 1.5%





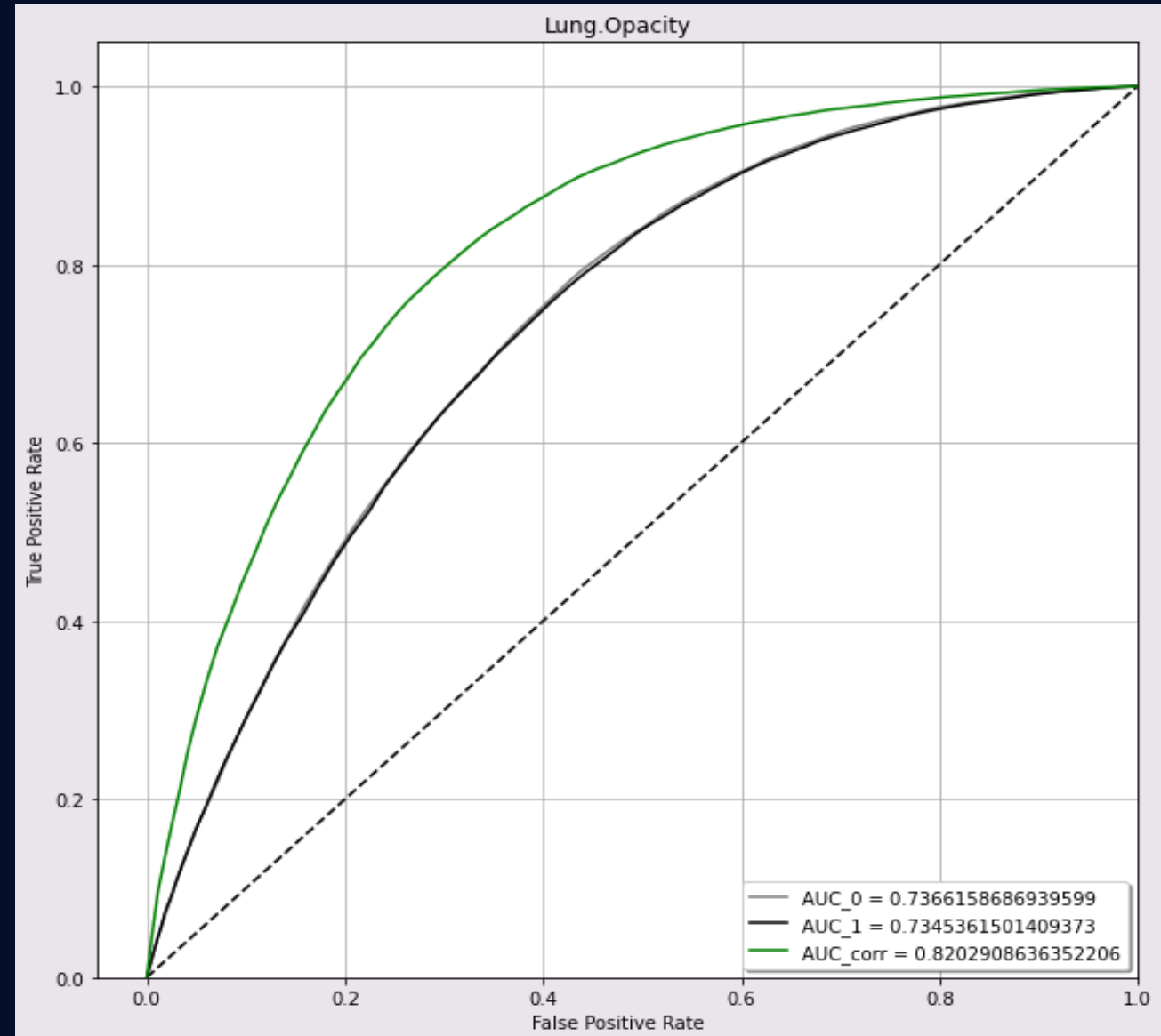
# RANDOM FOREST

**Lung.Opacity:**

U = 0

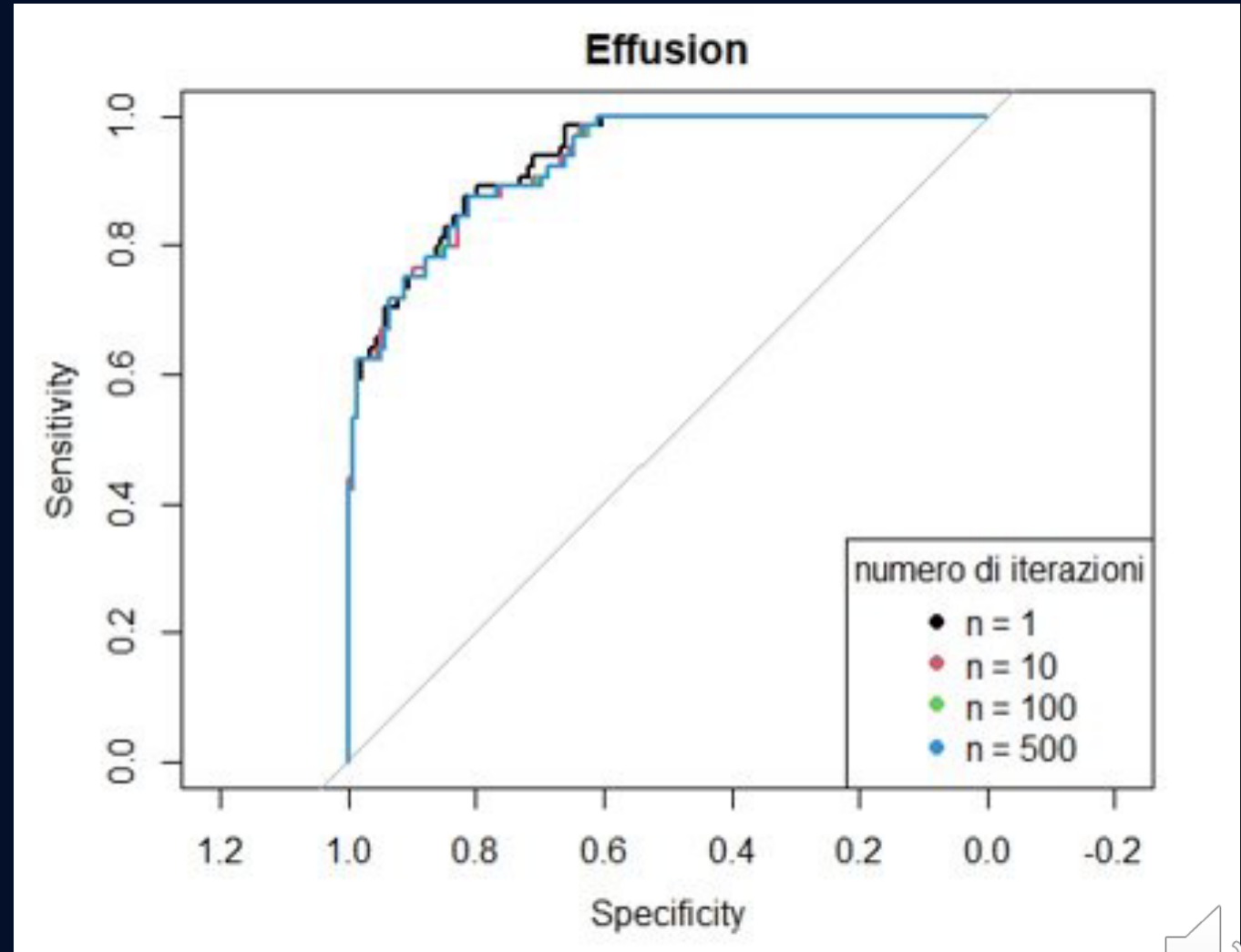
U = 1

U = corrected



# Bagging

- **Bagging**: sample with replacement (bootstrap) + aggregation
- Ensemble model with bootstrapped training sets
- At each iteration **uncertainties are resampled**





# CONCLUSION

# CONCLUSION

- ✓ Conditional models are very useful
- ✓ Corrected labels are helpful
- ✗ Treatment of uncertainties (because uncertainties  $\ll$  observations)
- ✗ Bagging for logistic regression

# THANK YOU

We look forward to answer your question!!!

If you want, join us at the following link:

<https://github.com/edpere/CheXpert-analysis>

**CheXpert competition:** <https://stanfordmlgroup.github.io/competitions/chexpert/>

## Bibliography:

- *Interpreting chest X-rays via CNNs that exploit disease dependencies and uncertainty labels*, Hieu H. Pham, Tung T. Le, Dat Q. Tran, Dat T. Ngo, Ha Q. Nguyen 2019
- *CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison*, Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, et al.









DATA PRESENTATION



GRAPHICAL  
REPRESENTATION



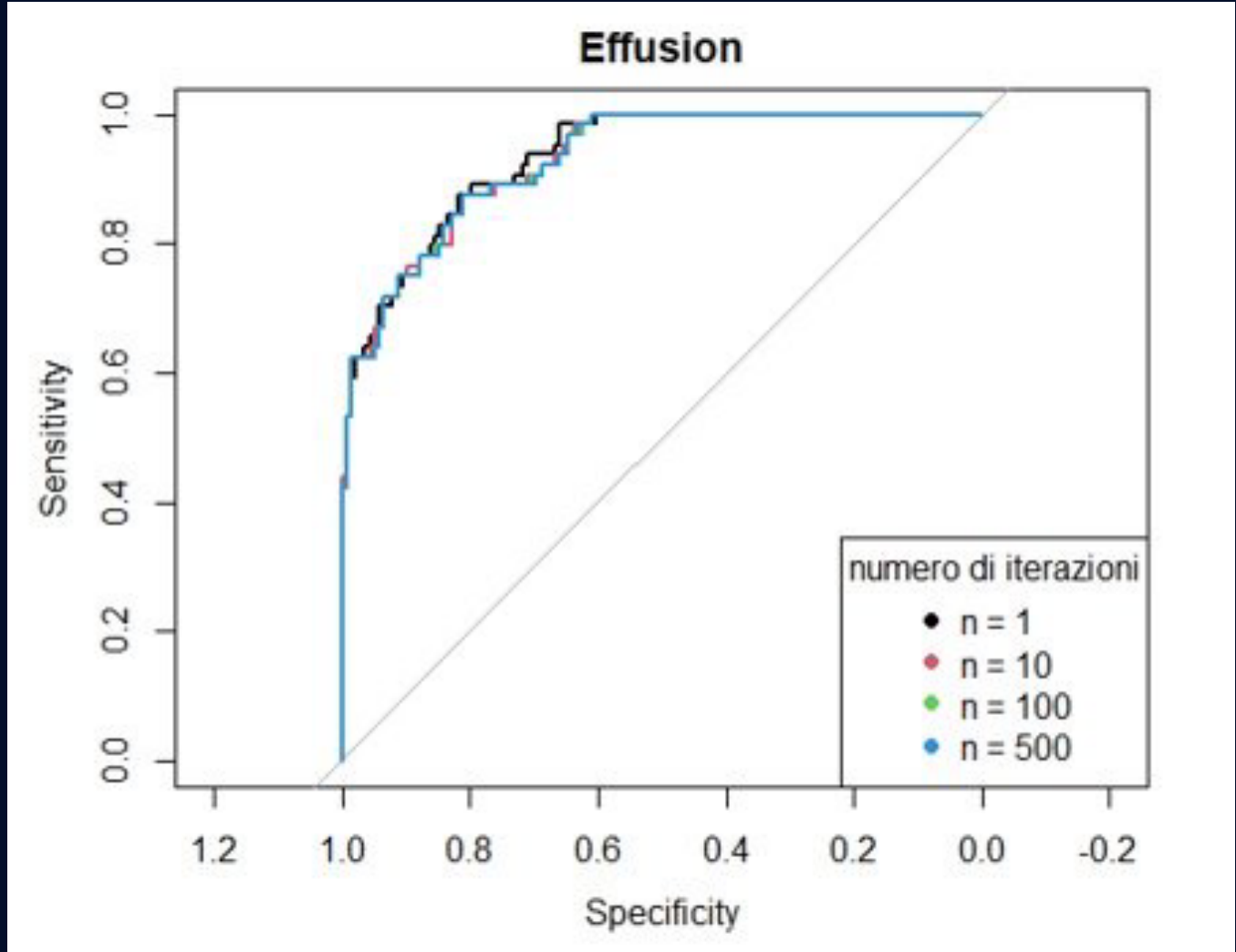
CLASSIFICATION  
MODELS



CONCLUSION

## Lasso

- We looked for the optimal hyper-parameter  $\alpha$  using the **validation set**
- In the end we do not see great improvements





## Conclusions

- ✓ Conditional models are very useful
- ✓ Corrected labels can be helpful
- ✗ Treatment of uncertainties (because of big dataset)
- ✗ Bagging for logistic regression
- ✗ Regularization (e.g. lasso) (no overfitting)