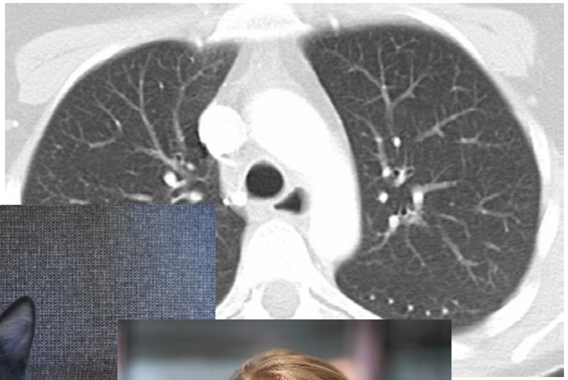
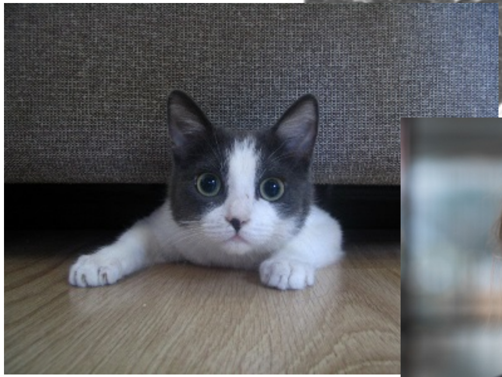
Technische Universiteit
Eindhoven
University of Technology

Not-so-supervised learning of algorithms & academics





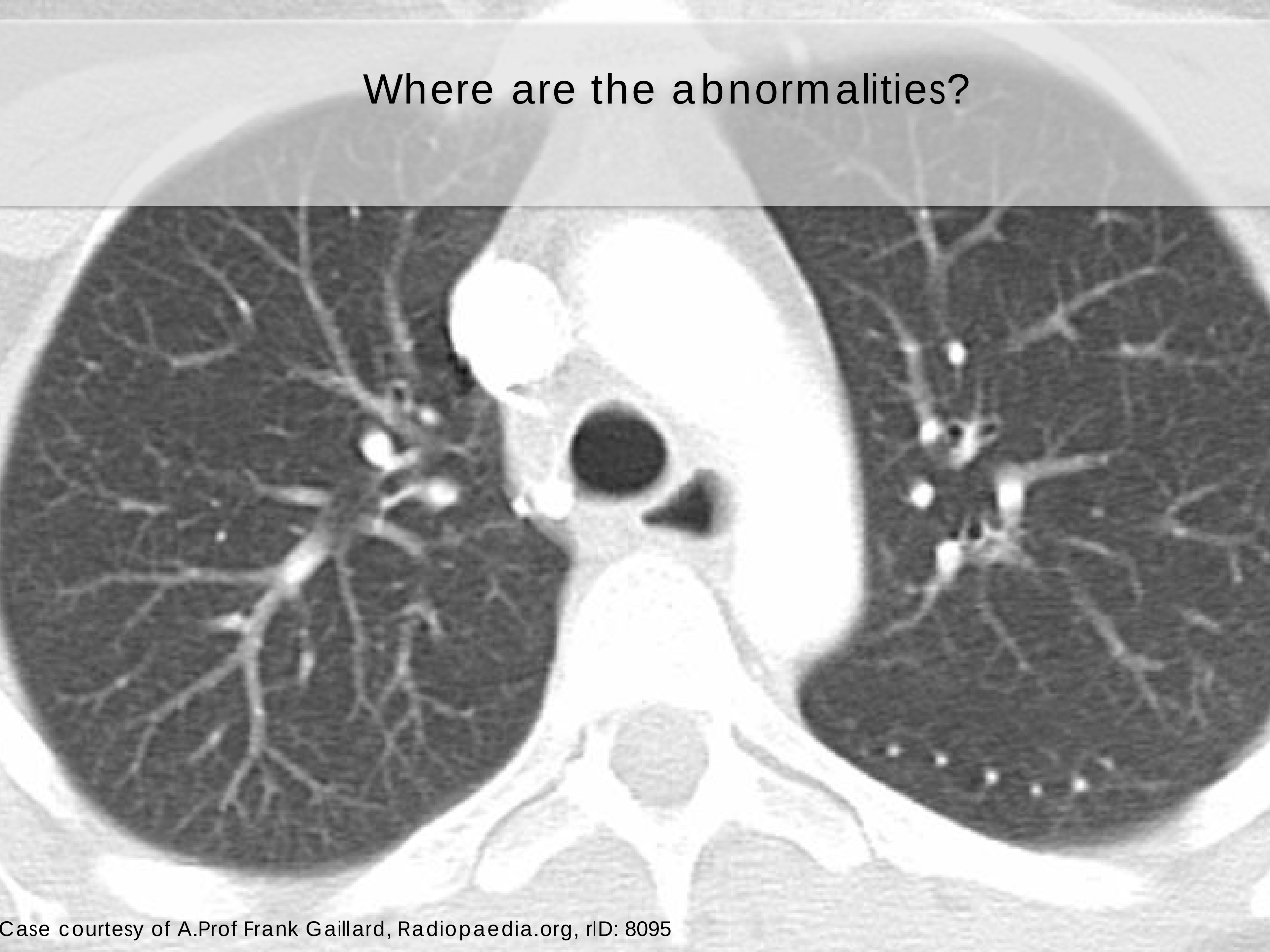
Veronika Cheplygina

@drveronikach

http://www.veronikach.com

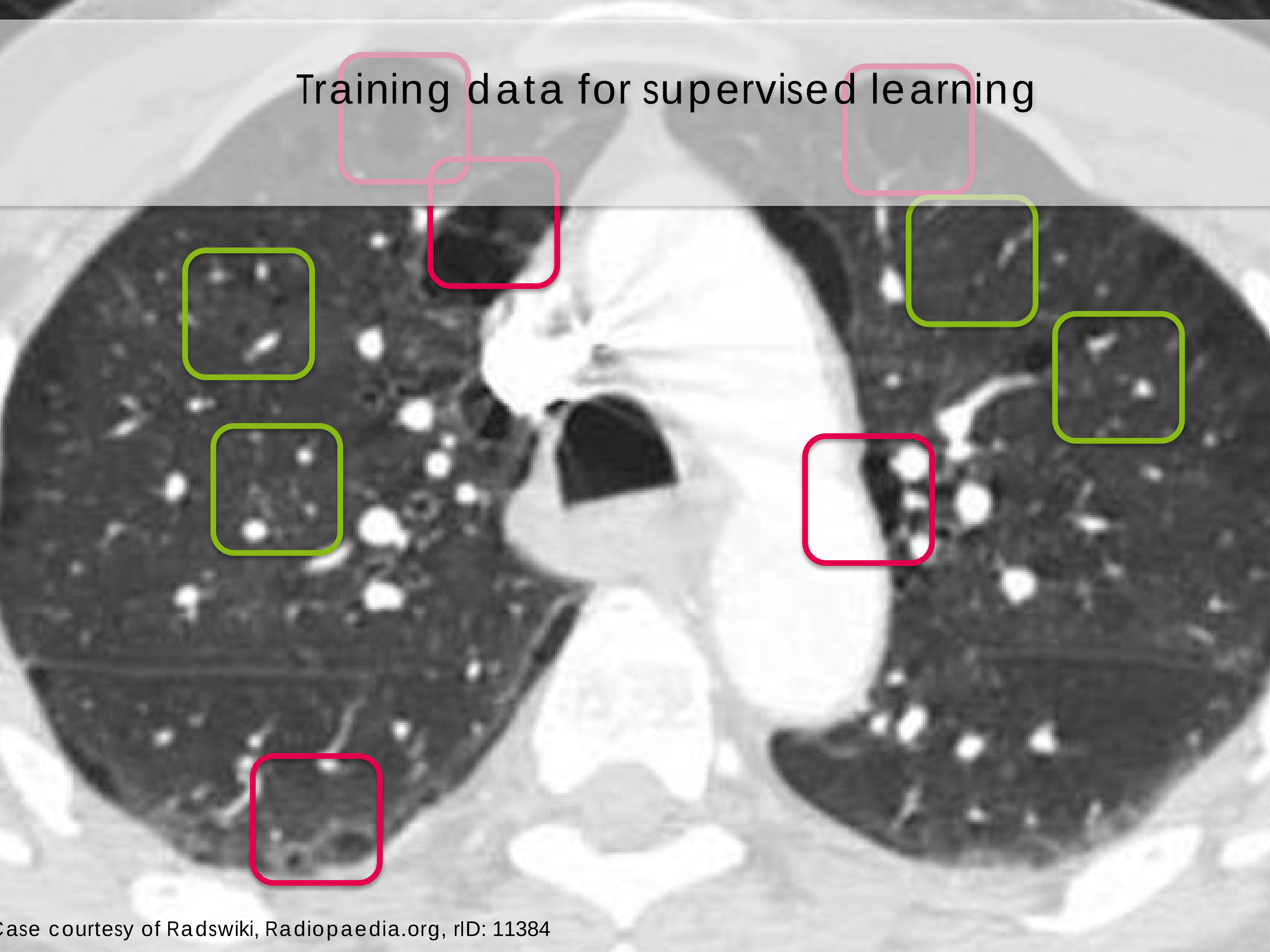

Data?

Where are the abnormalities?



Case courtesy of A.Prof Frank Gaillard, Radiopaedia.org, rID: 8095

Training data for supervised learning



Case courtesy of Radswiki, Radiopaedia.org, rID: 11384

Training data we get

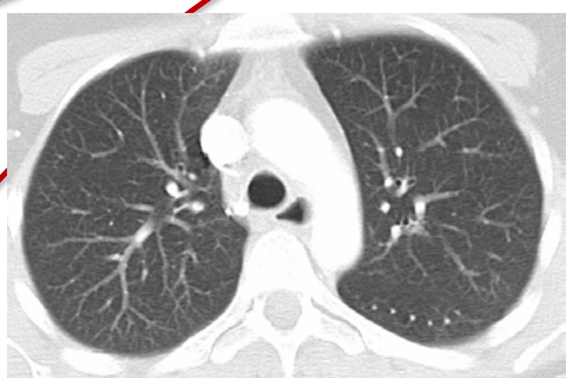
Case courtesy of Radswiki, Radiopaedia.org, rID: 11384

Training data not representative


Data?


Representative
& annotated data

Learning curve



@drveronikach

http://www.veronikach.com

Janelle Shane
@JanelleCShane

Follow

Does anyone have a picture of sheep in a really unusual place? It's for pranking a neural net.

4:55 PM - 1 Mar 2018

1,329 Retweets **3,698** Likes

466 1.3K 3.7K

Dr Veronika CH @DrVeronikaCH · 2 Mar 2018

I've got one of those as well

3 45

Dr Veronika CH @DrVeronikaCH · 2 Mar 2018

I've got one of those as well

3 45

Janelle Shane @JanelleCShane · 2 Mar 2018

I love how it took special care to also get the animal's color wrong: "a little boy holding a brown teddy bear"

1 3 73

This talk: three “solutions”

- Multiple instance learning
- Transfer learning
- Crowdsourcing

Bonus:

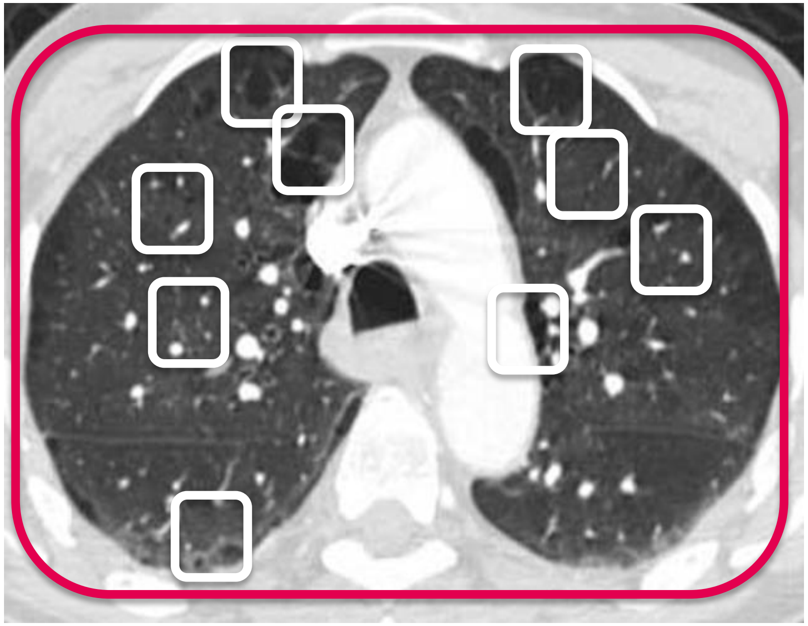
- Not-so-supervised career path

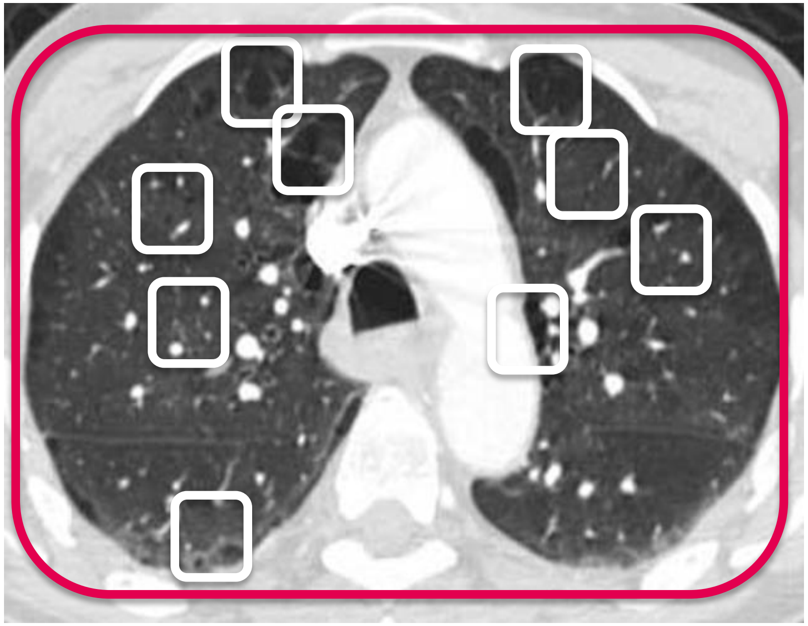
Solution 1: Multiple instance learning

Learn with global information – Carolyn is in both pictures

Solution 1: Multiple instance learning

Learn with global information





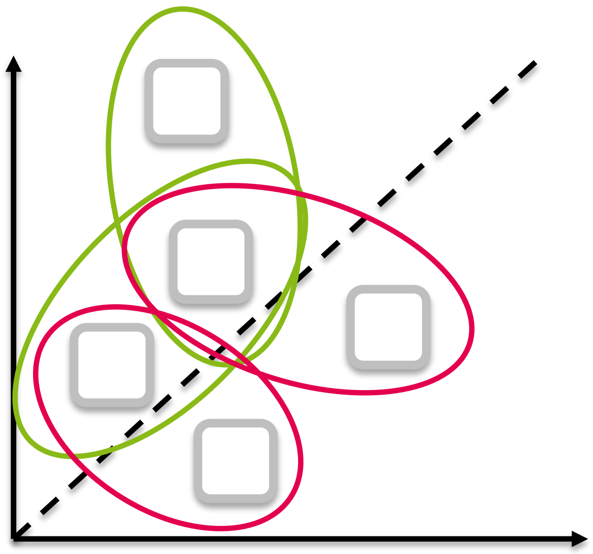
Solution 1: Multiple instance learning

Search for positive instances

vs

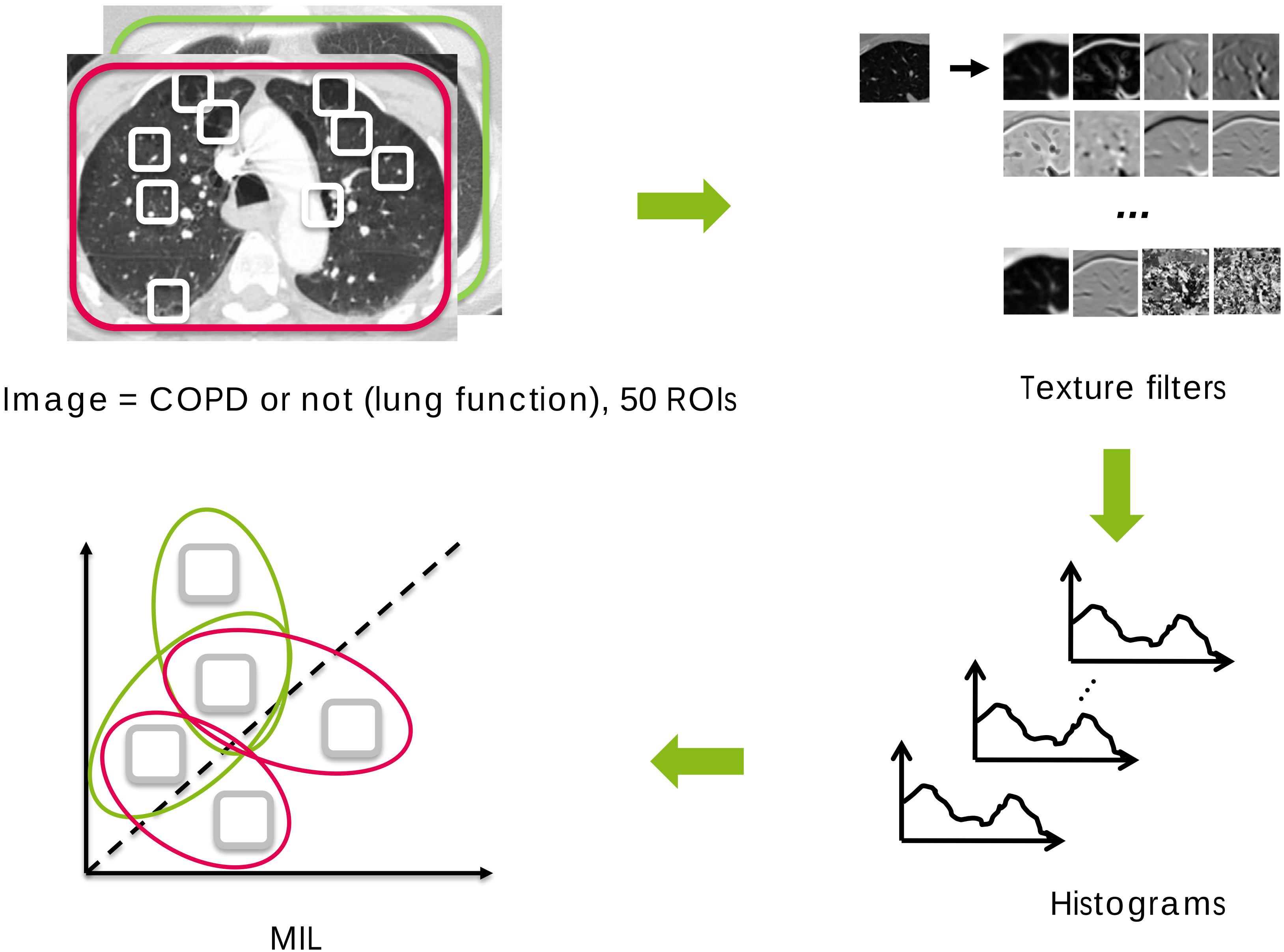
Classify whole bag

MIL



@drveronikach

http://www.veronikach.com



Images Marleen de Bruijne

Performance metric: subject-level AUC

Search for abnormalities

vs

Classify whole subject

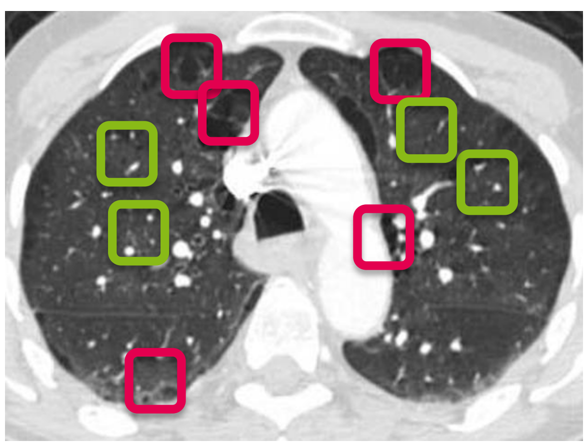
Classifier	AUC $\mathcal{X}_{val}$	AUC $\mathcal{X}_{te}$
Simple logistic noisy	50.0	50.0
Simple logistic avg	**71.9**	**70.5**
Simple k-NN noisy	61.0	65.9
Simple k-NN avg	67.0	67.8
miSVM noisy	**69.7**	65.4
miSVM avg	**74.5**	**71.7**
MILBoost	55.8	61.4
Citation k-NN	65.2	61.5
mean-inst SVM	**74.0**	**74.2**
extremes SVM	**70.8**	68.6
BoW SVM	50.0	50.0
MILES	65.8	68.2
meanmin SVM	70.8	**71.3**
meanmin k-NN	65.0	**69.1**
emd SVM	**73.7**	**74.6**
emd k-NN	65.1	67.1

Cheplygina, V., Sorensen, L., Tax, D. M. J., Pedersen, J. H., Loog, M., & de Bruijne, M. (2014). Classification of COPD with multiple instance learning. In International Conference on Pattern Recognition (pp. 1508-1513).

Training

MIL classifier

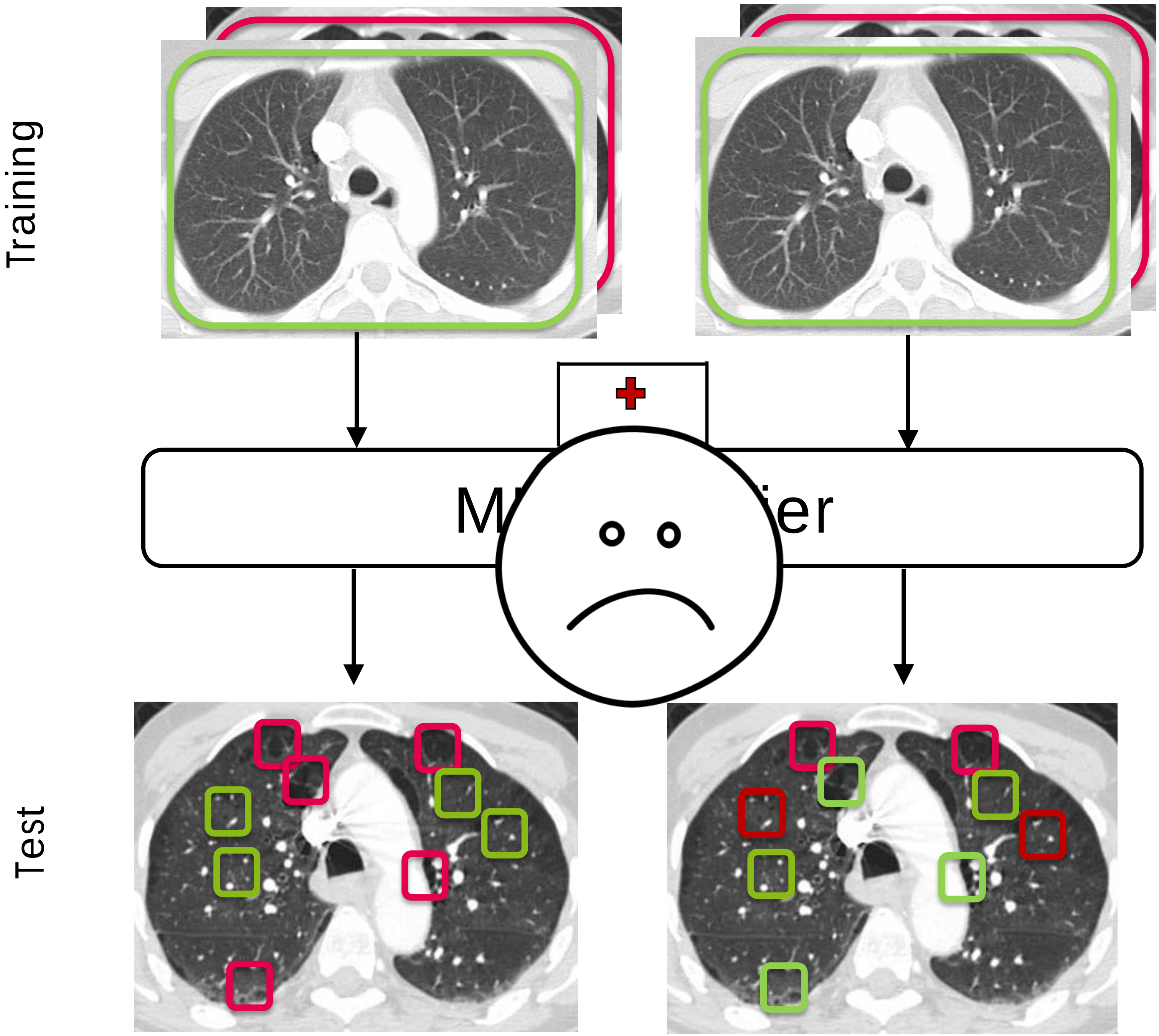
Test



Ground truth?

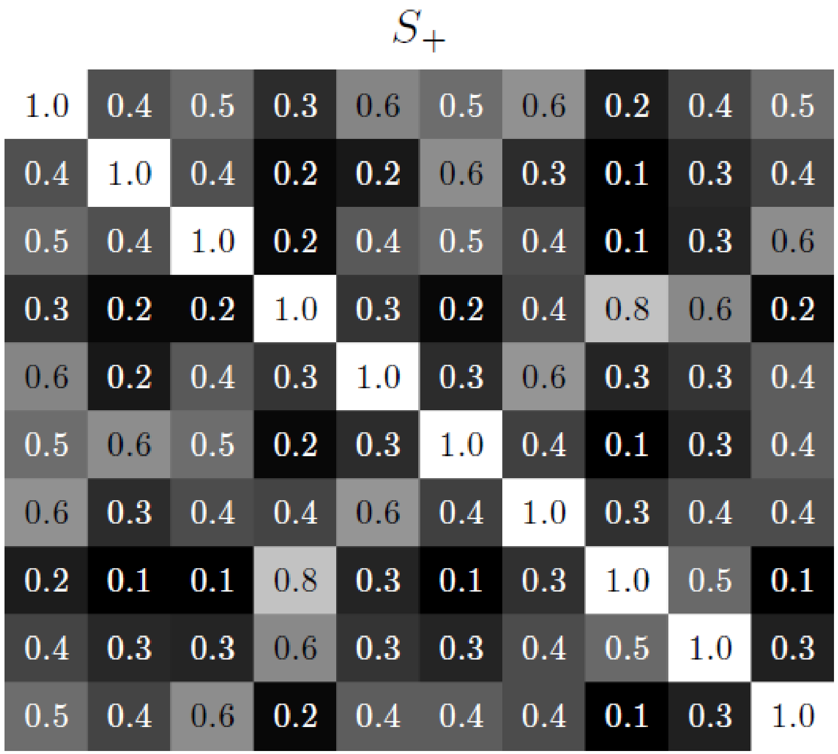
@drveronikach

http://www.veronikach.com

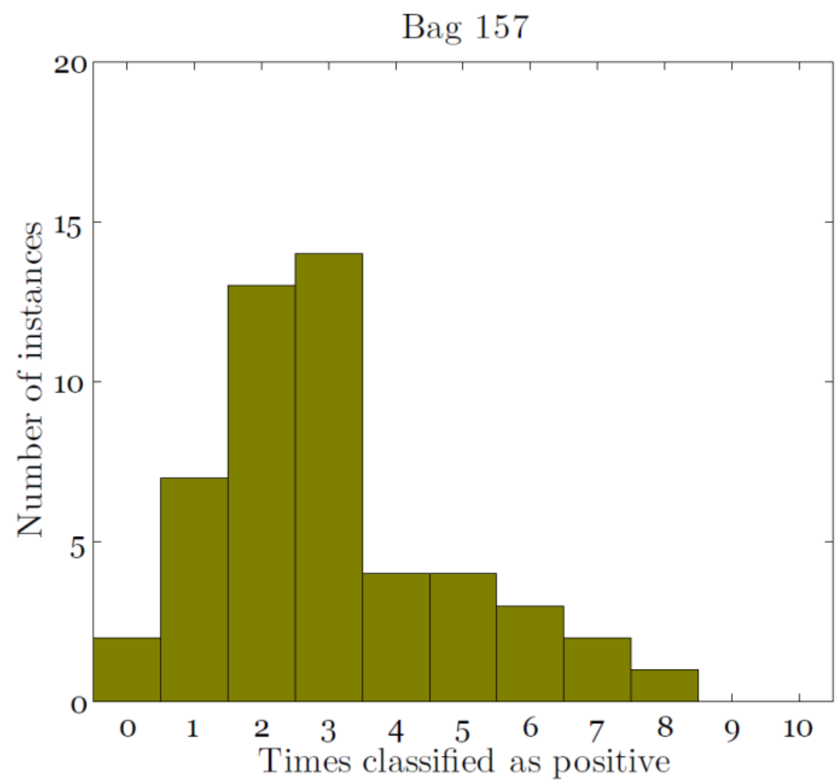

Training
Test
M
ier

Evaluate stability

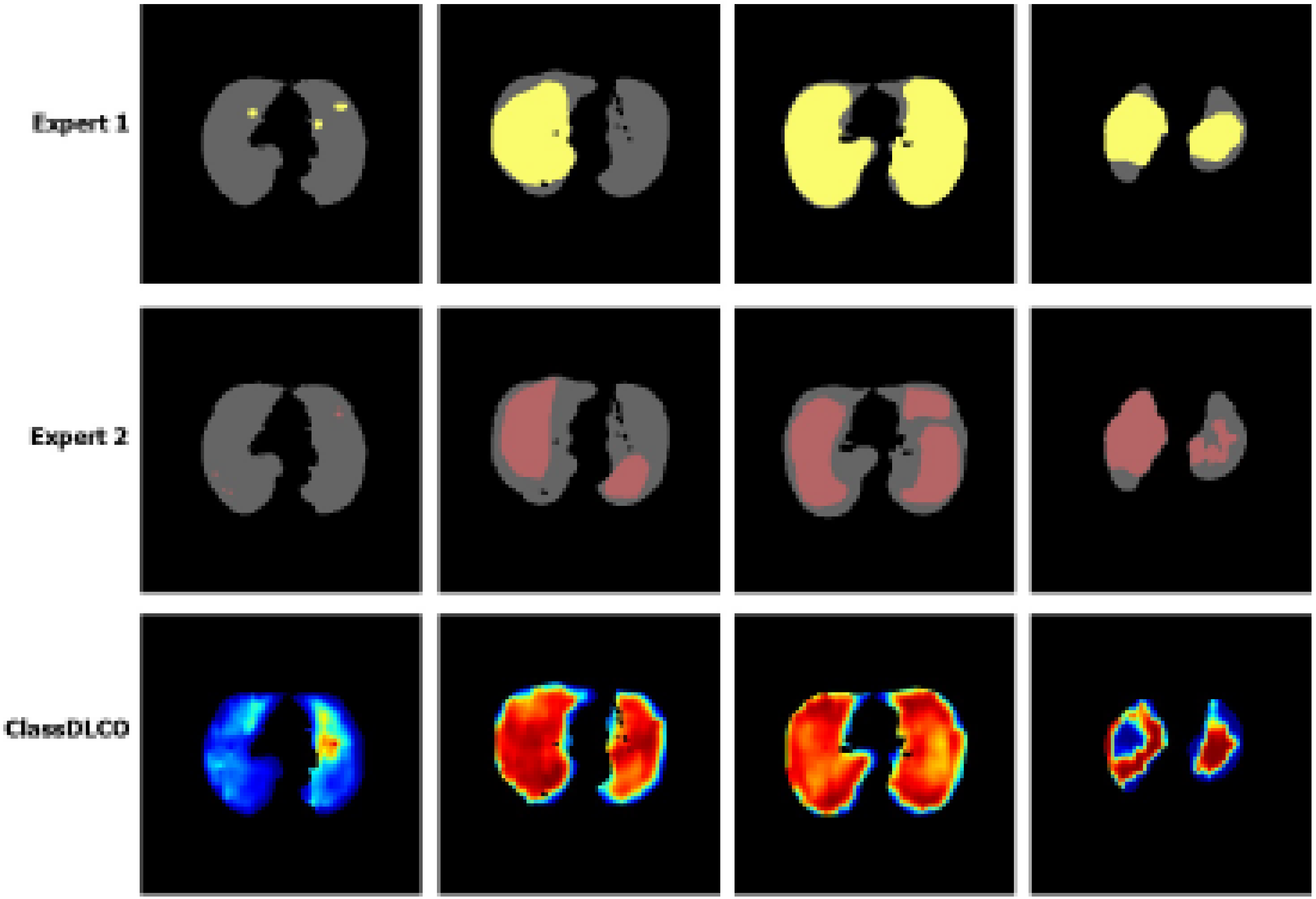
Fraction of agreement on positives



Any patches always positive?



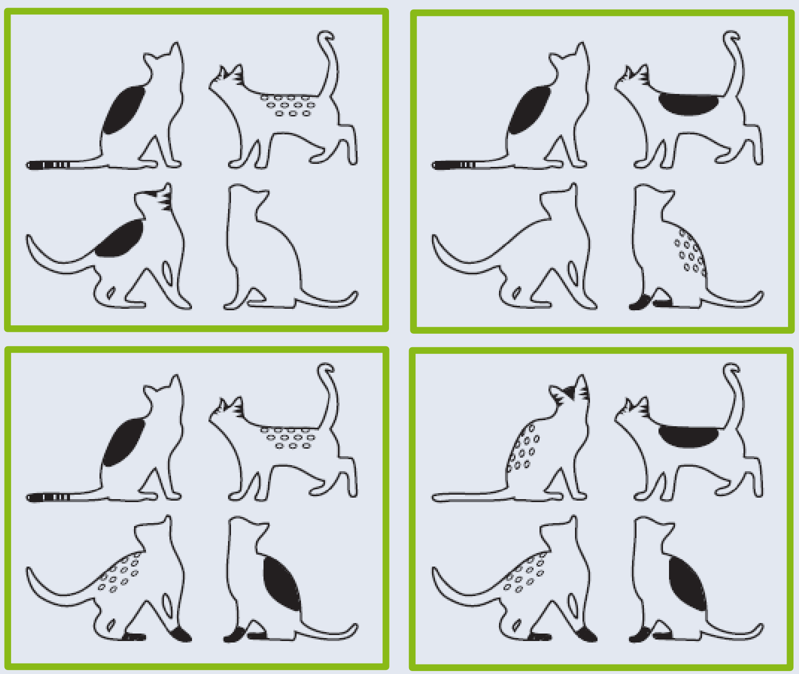
Cheplygina, V., Sørensen, L., Tax, D. M. J., de Bruijne, M., & Loog, M. (2015) Label Stability in Multiple Instance Learning. *In Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 539-546*



Pena, I. P., Cheplygina, V., Paschaloudi, S., Vuust, M., Carl, J., Weinreich, U. M., ... & de Bruijne, M. (2018). Automatic emphysema detection using weakly labeled HRCT lung images. *PloS one*, *13*(10), e0205397.

Dissimilarity-Based Multiple Instance Learning

What is different about the groups of cats on the front cover, from the groups on the back cover? If you can answer this question, you are probably also able to categorize another, previously unseen to you, group of cats. This thesis is about different applications where similar puzzles may occur, and how some machine learning algorithms approach such problems.



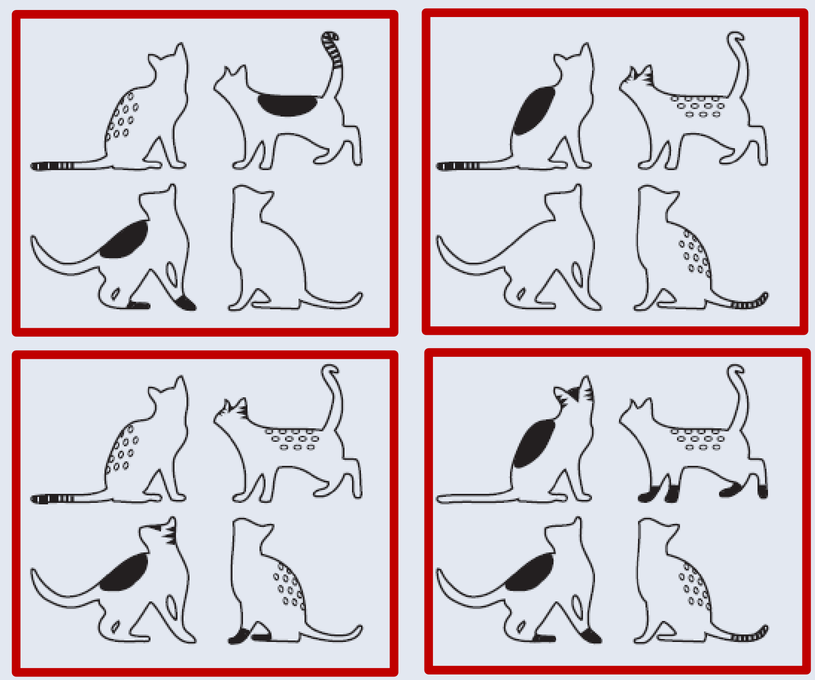
ISBN 978-94-6295-192-1

Dissimilarity-Based Multiple Instance Learning

Veronika Cheplygina

Dissimilarity-Based Multiple Instance Learning

Veronika Cheplygina



Solution 2: Transfer learning

Not learning “from scratch”

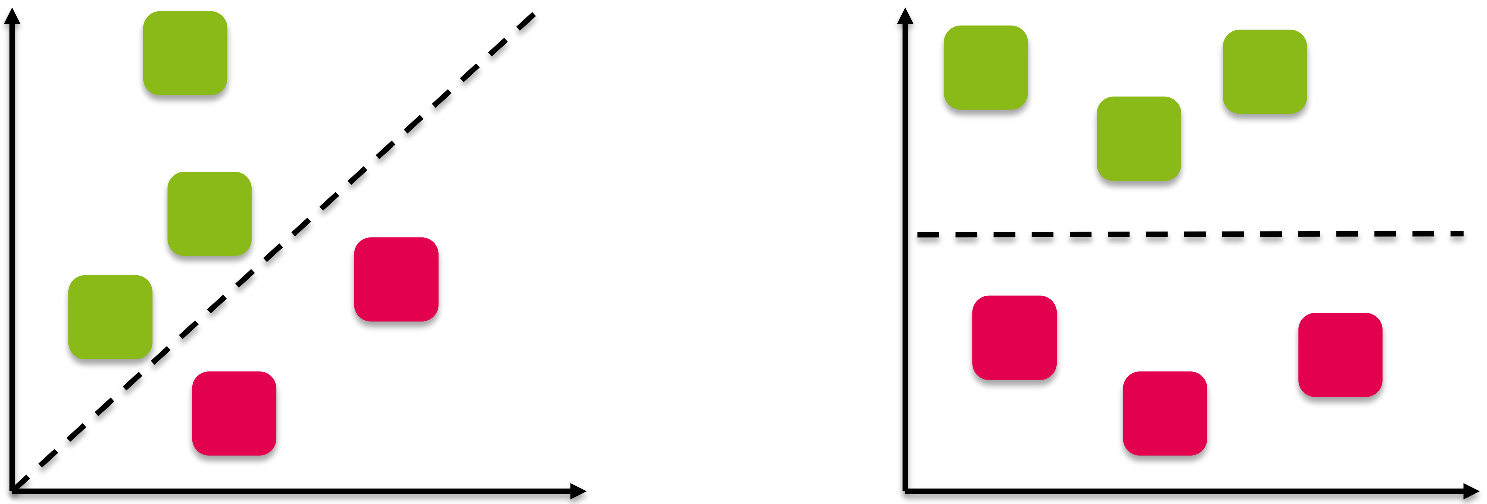
Use other similar datasets

Dataset	Subjects	Age	GOLD (1/2/3/4)	Smoking (c/f/n)	Scanner	Resolution (mm)
DLCST	300 +	59 [50, 71]	69/28/2/0	77/23/0	Philips	$0.72\times0.72\times1$ to
	300 -	57 [49, 69]		74/26/0	16 rows Mx 8000	$0.78\times0.78\times1$
COPDGene1	74 +	64 [45, 80]	21/18/19/16	17/57/0	Siemens	$0.65\times0.65\times0.75$
	46 -	59 [45, 78]		23/20/3	Definition	
COPDGene2	42 +	65 [45, 78]	9/13/7/13	12/30/0	Siemens	$0.65\times0.65\times0.75$
	25 -	60 [47, 78]		9/11/5	Definition AS+	
Frederikshavn	8 +	66 [48, 77]	1/3/3/1	1/7/0	Siemens	$0.58\times0.58\times0.6$
	8 -	56 [25, 73]		1/2/5	Definition Flash	

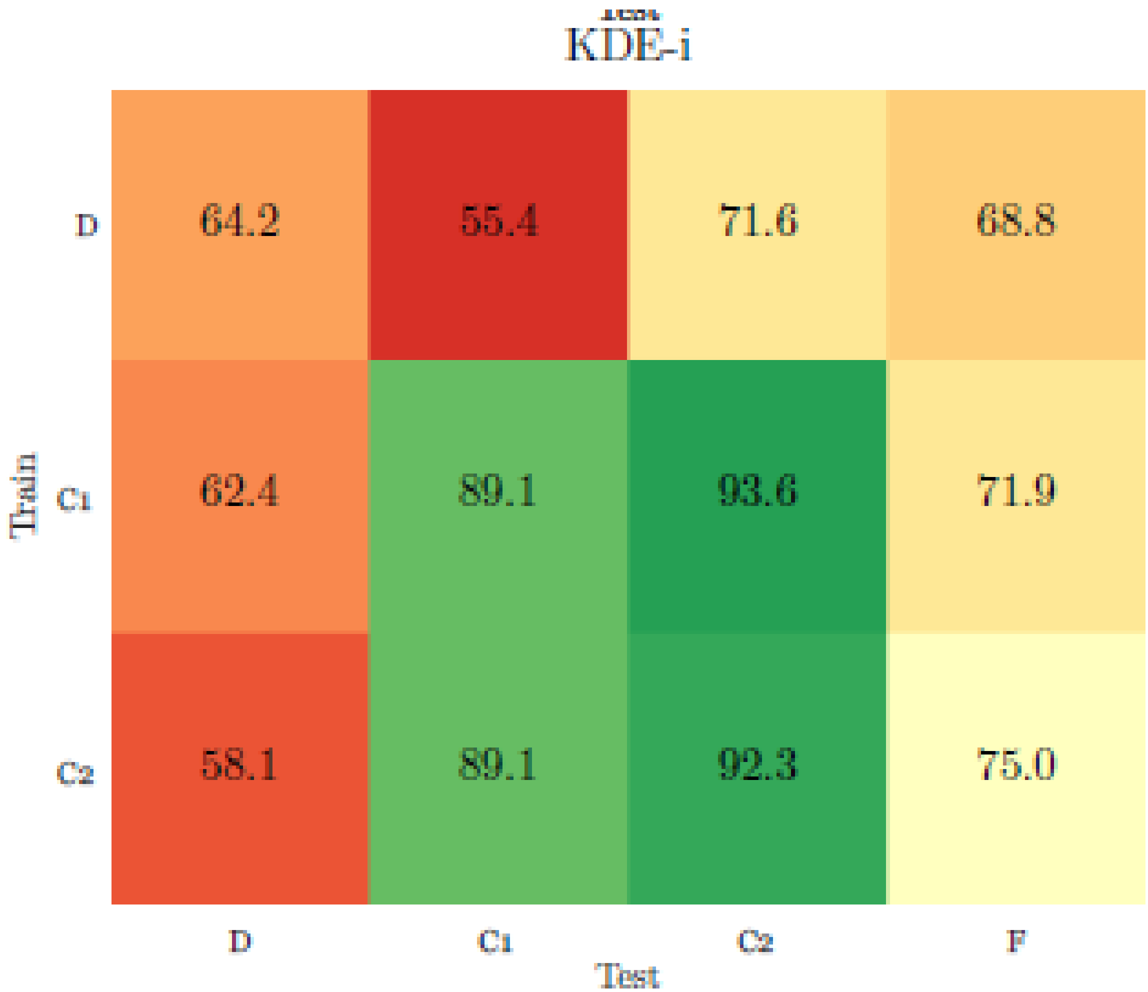
Use data from similar datasets

Source

Target

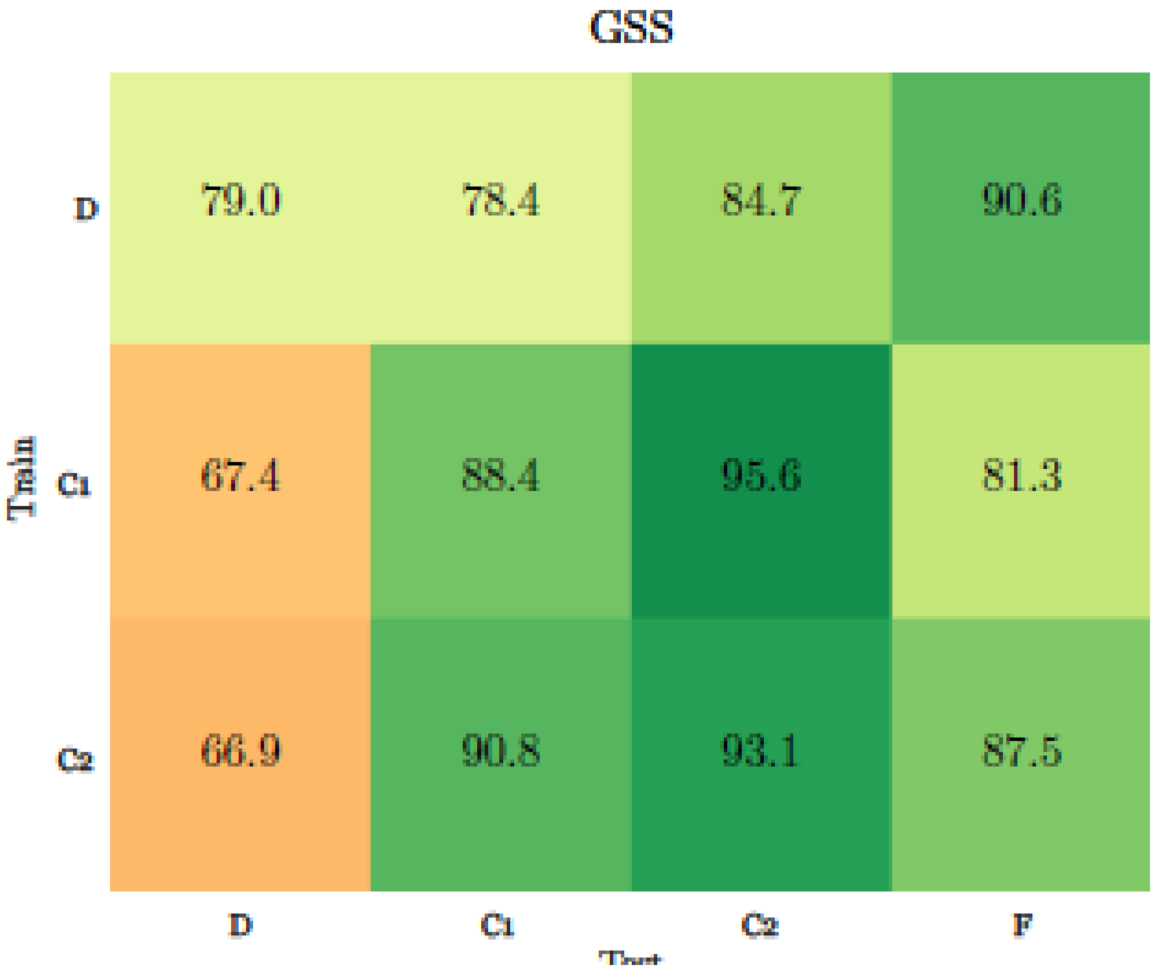


Use other similar datasets – Intensity features



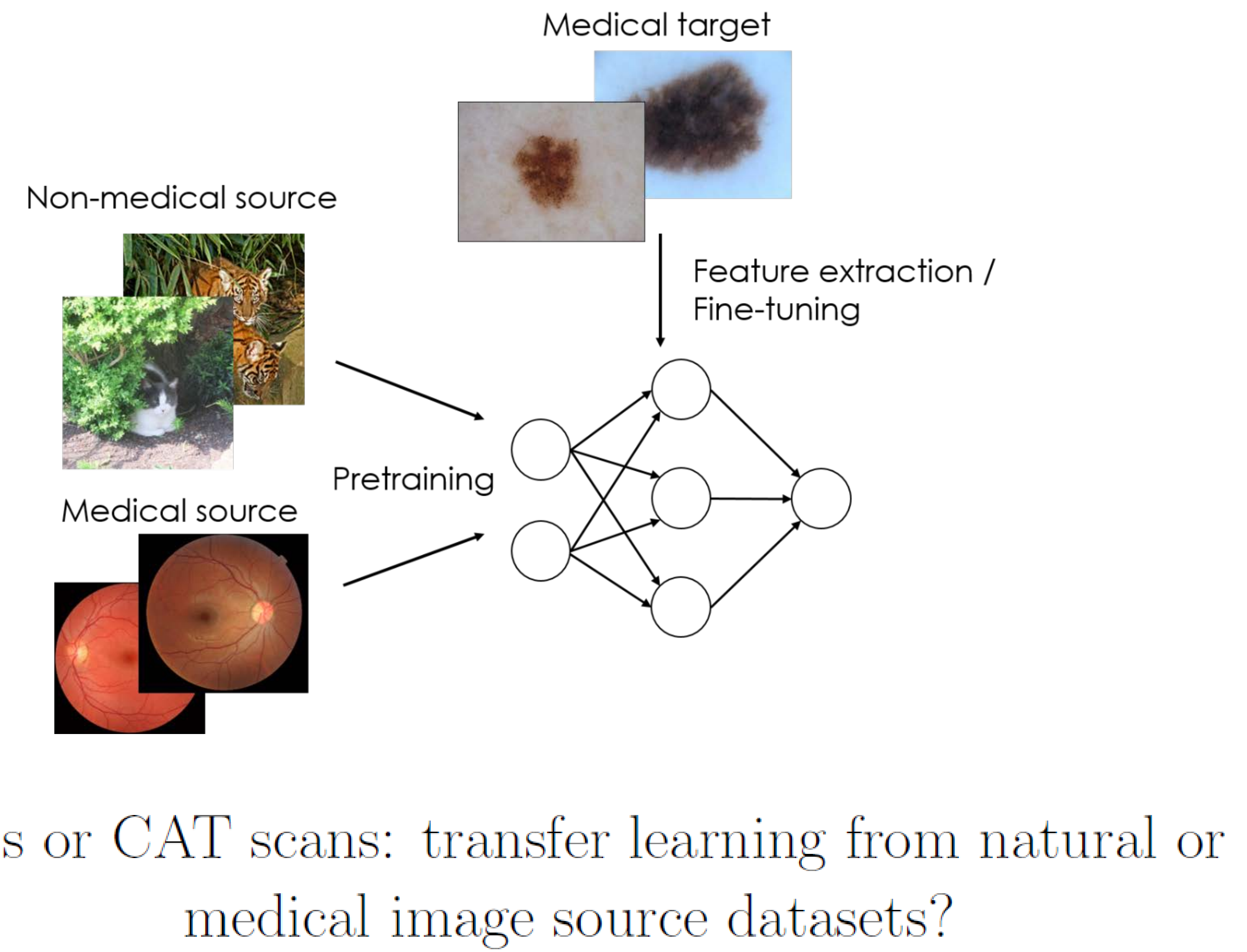
Cheplygina, V., Pena, I. P., Pedersen, J. H., Lynch, D. A., Sørensen, L., & de Bruijne, M. (2018). Transfer learning for multicenter classification of chronic obstructive pulmonary disease. *IEEE journal of biomedical and health informatics*, *22*(5), 1486-1496.

Use other similar datasets – Texture features



Cheplygina, V., Pena, I. P., Pedersen, J. H., Lynch, D. A., Sørensen, L., & de Bruijne, M. (2018). Transfer learning for multicenter classification of chronic obstructive pulmonary disease. *IEEE journal of biomedical and health informatics*, *22*(5), 1486-1496.

… or data that's entirely different



Cheplygina, V. (2019). Cats or CAT scans: transfer learning from natural or medical image source datasets?. *Current Opinion in Biomedical Engineering*. URL

Non-medical or medical data for pretraining?

3 papers: non-medical is better

5 papers: medical is better

2 papers: no differences

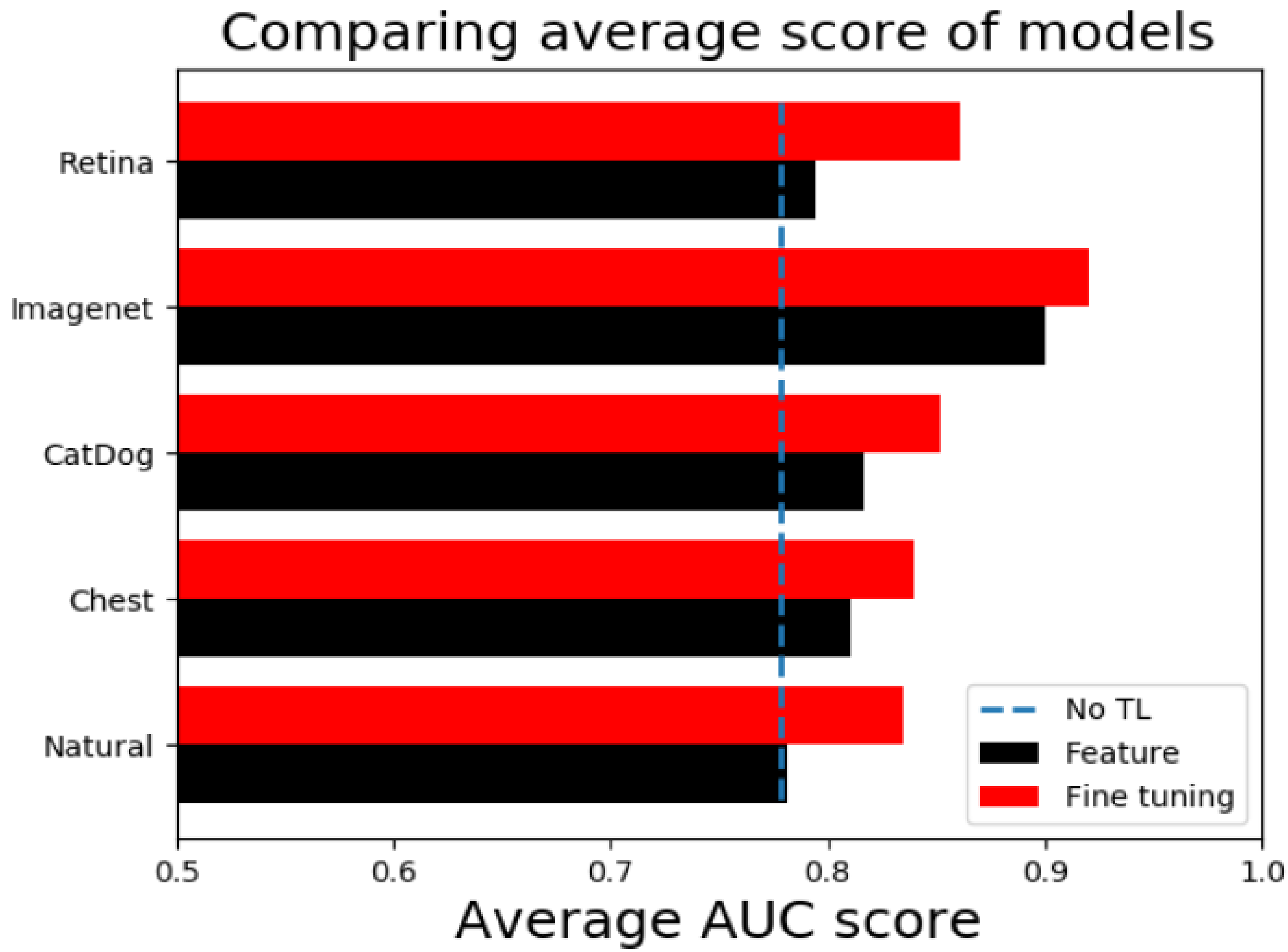
3 papers: inconclusive, BUT more data not better

Systematic comparison?

ImageNet best as source data

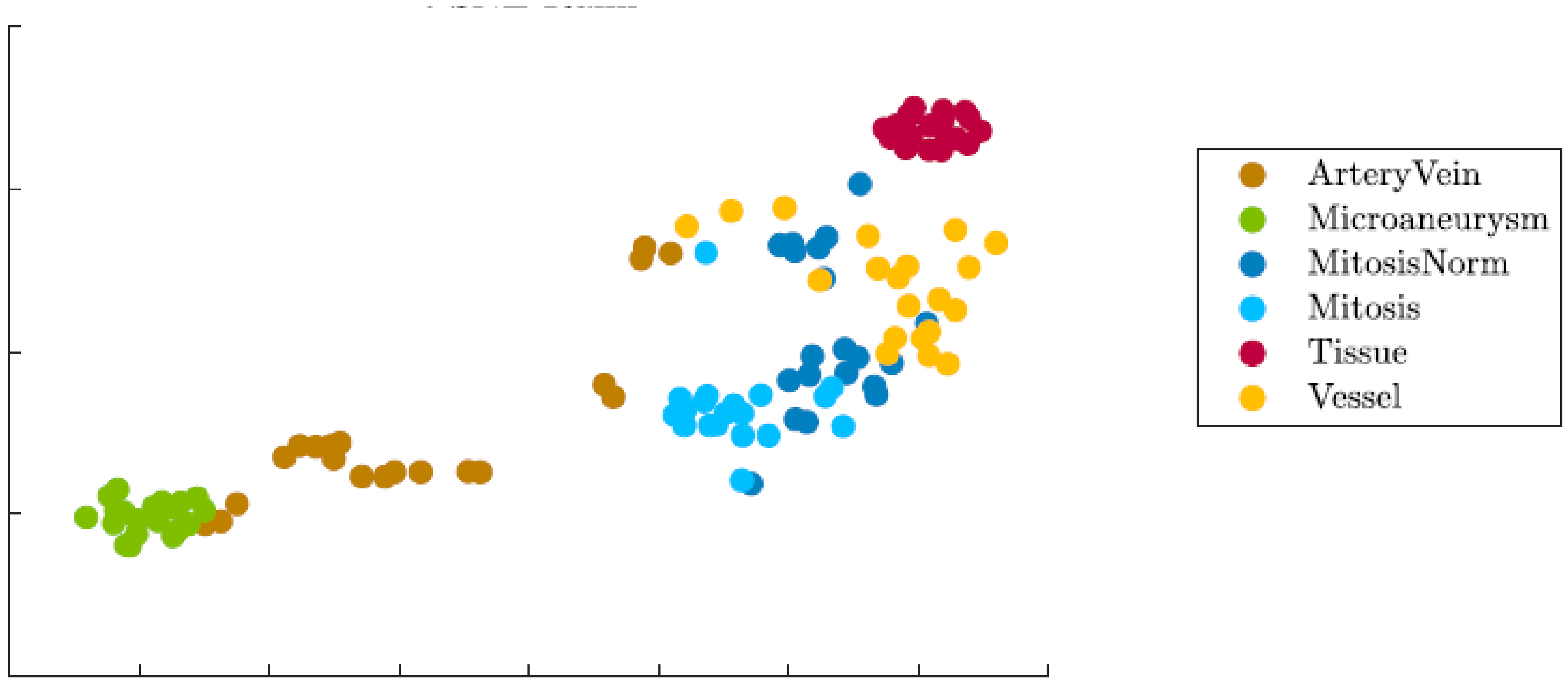
BUT

is Imagenet is a much bigger datas



Work by Floris Fok

Meta-learning: how to quantify similarity of data?



Cheplygina, V., Moeskops, P., Veta, M., Bozorg, B. D., & Pluim, J. (2017). Exploring the similarity of medical imaging classification problems. In *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis (MICCAI LABELS)* (pp. 59-66)

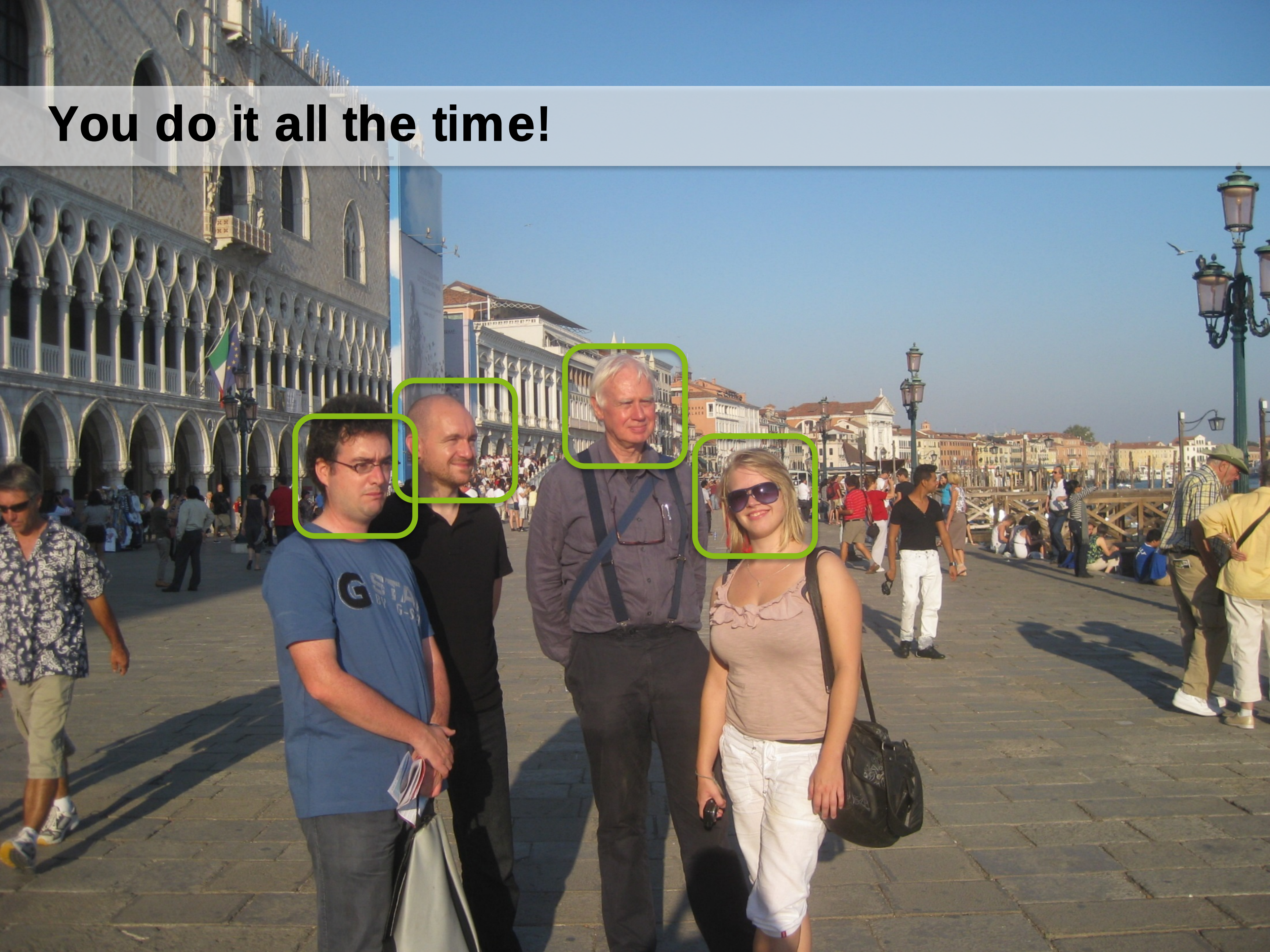
Solution 2: Transfer learning

Not learning “from scratch”

Cheplygina, V., de Bruijne, M., & Pluim, J. P. W. (2019). Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis*. URL

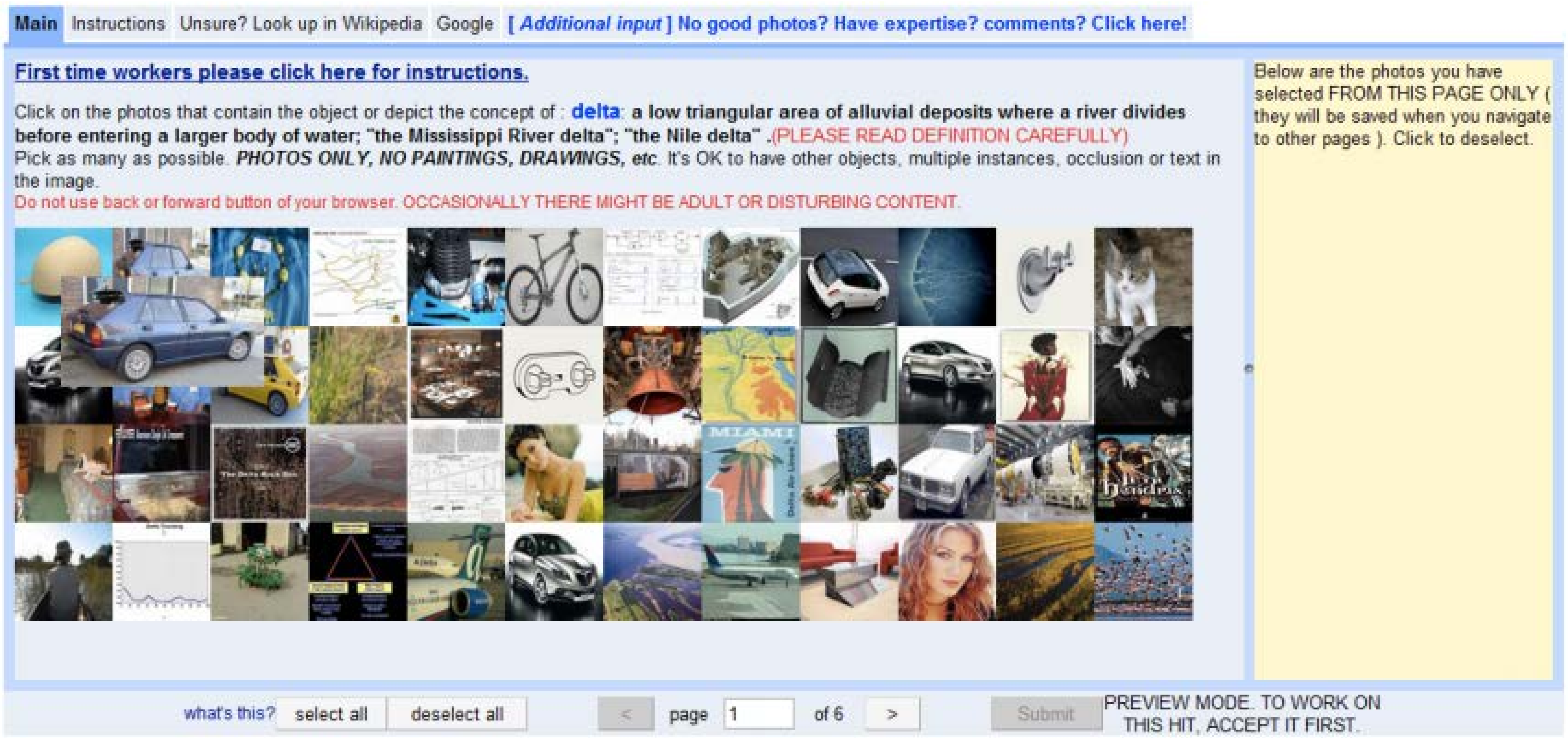
Solution 3: Crowdsourcing

You do it all the time!



2009: ImageNet

amazon mechanical turk beta

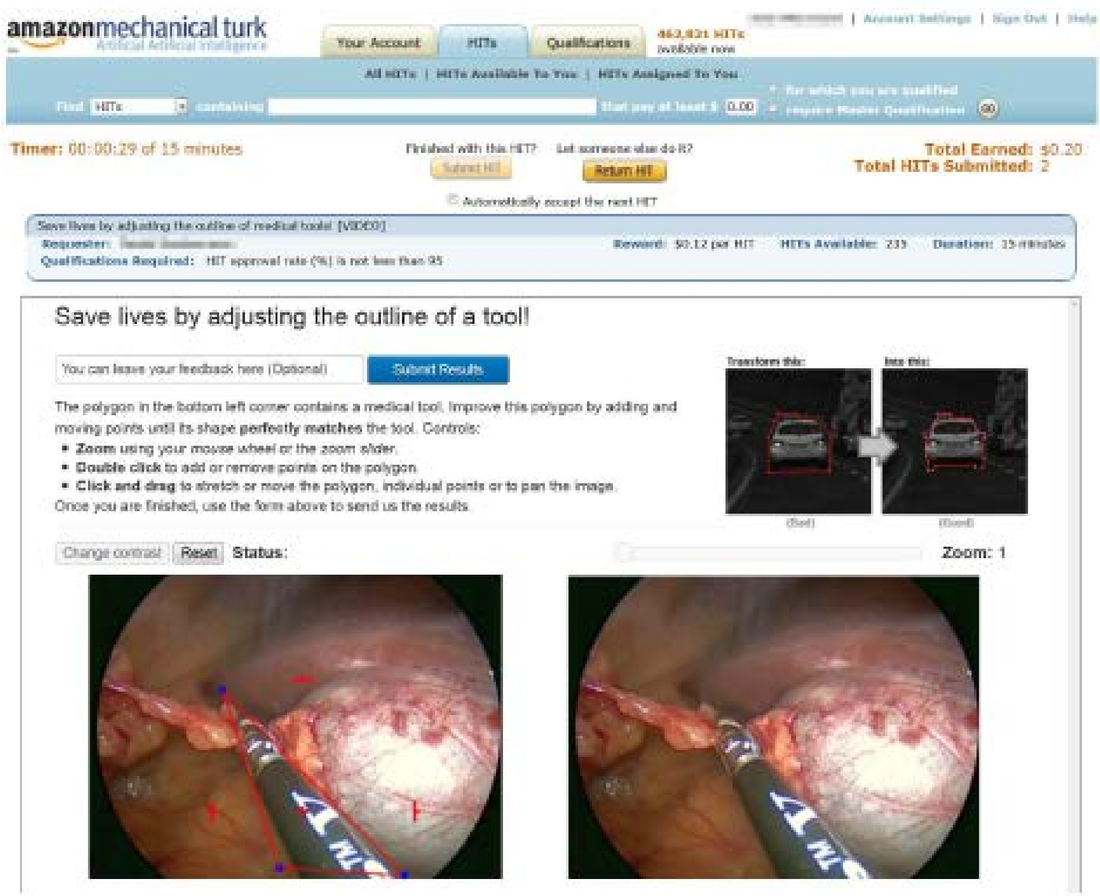


Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 248-255). IEEE.

Can Masses of Non-Experts Train Highly Accurate Image Classifiers?

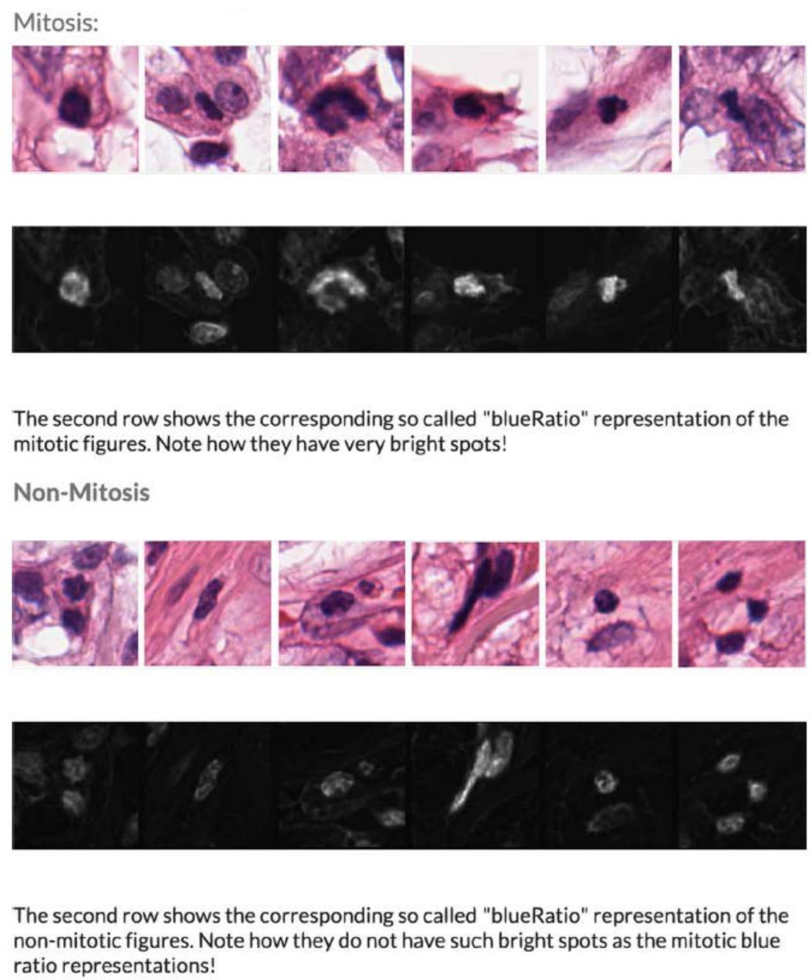
A Crowdsourcing Approach to Instrument Segmentation in Laparoscopic Images

Lena Maier-Hein[1,⋆,⋆⋆], Sven Mersmann[1], Daniel Kondermann[2], Sebastian Bodenstedt[3], Alexandro Sanchez[2], Christian Stock[4], Hannes Gotz Kenngott[5], Mathias Eisenmann[3], and Stefanie Speidel[3]



AggNet: Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images

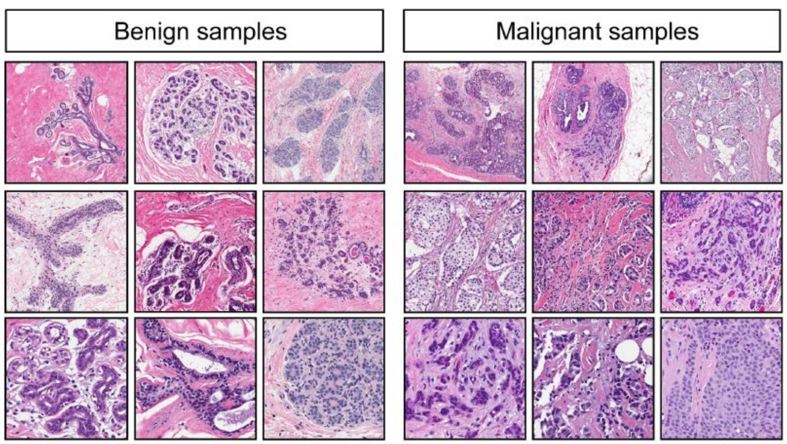
Shadi Albarqouni*, *Student Member, IEEE*, Christoph Baur, Felix Achilles, *Student Member, IEEE*, Vasileios Belagiannis, *Student Member, IEEE*, Stefanie Demirci, and Nassir Navab, *Member, IEEE*

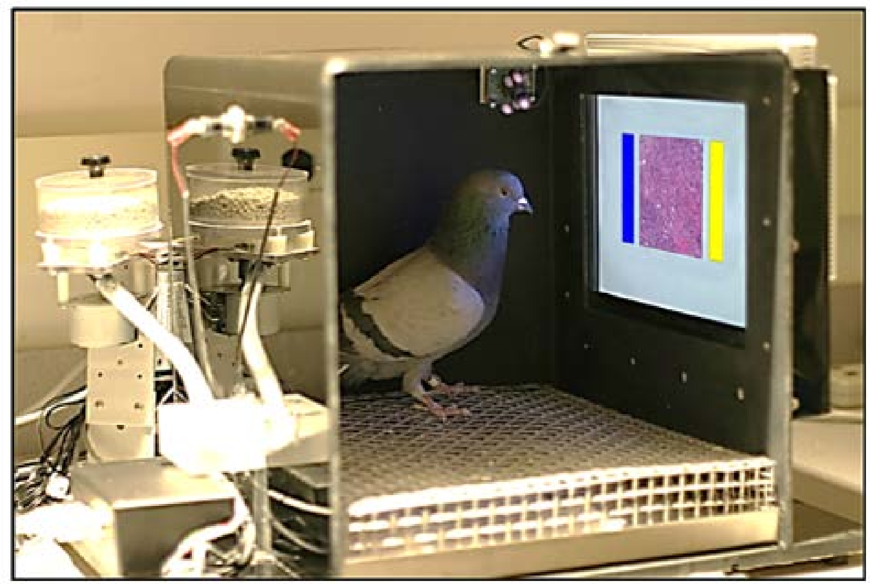


RESEARCH ARTICLE

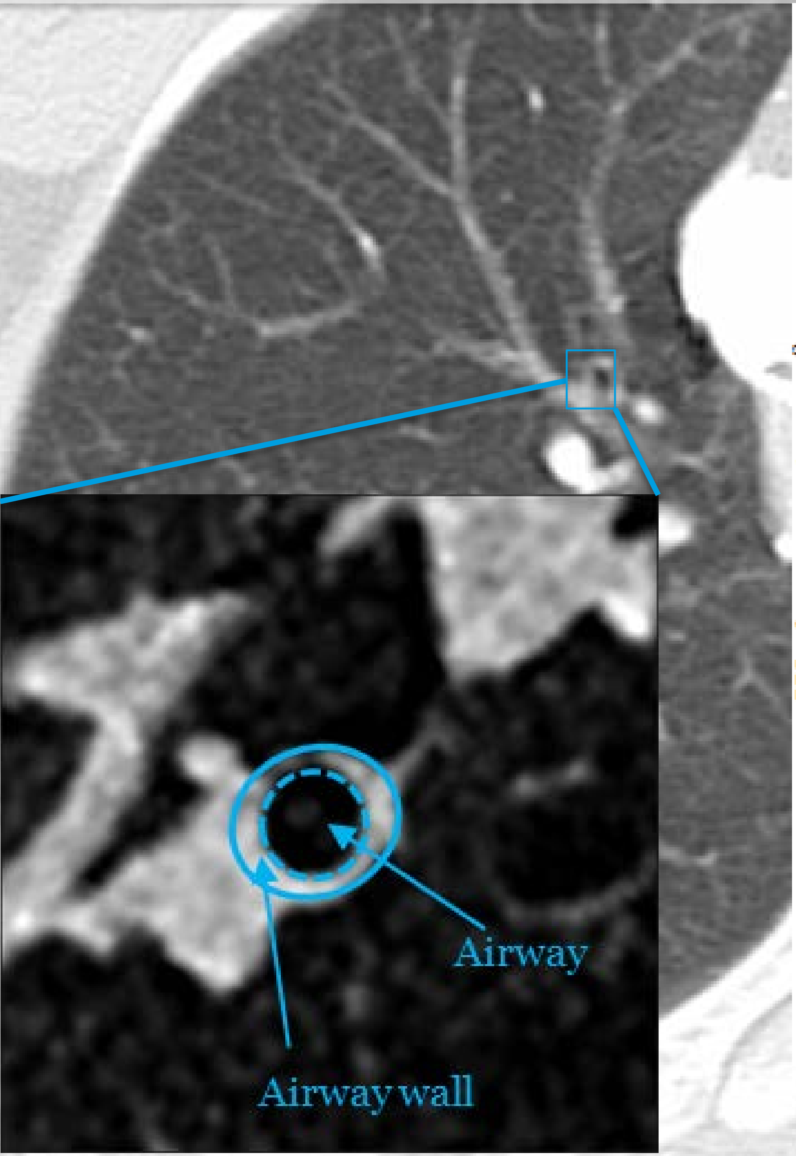
Pigeons (*Columba livia*) as Trainable Observers of Pathology and Radiology Breast Cancer Images

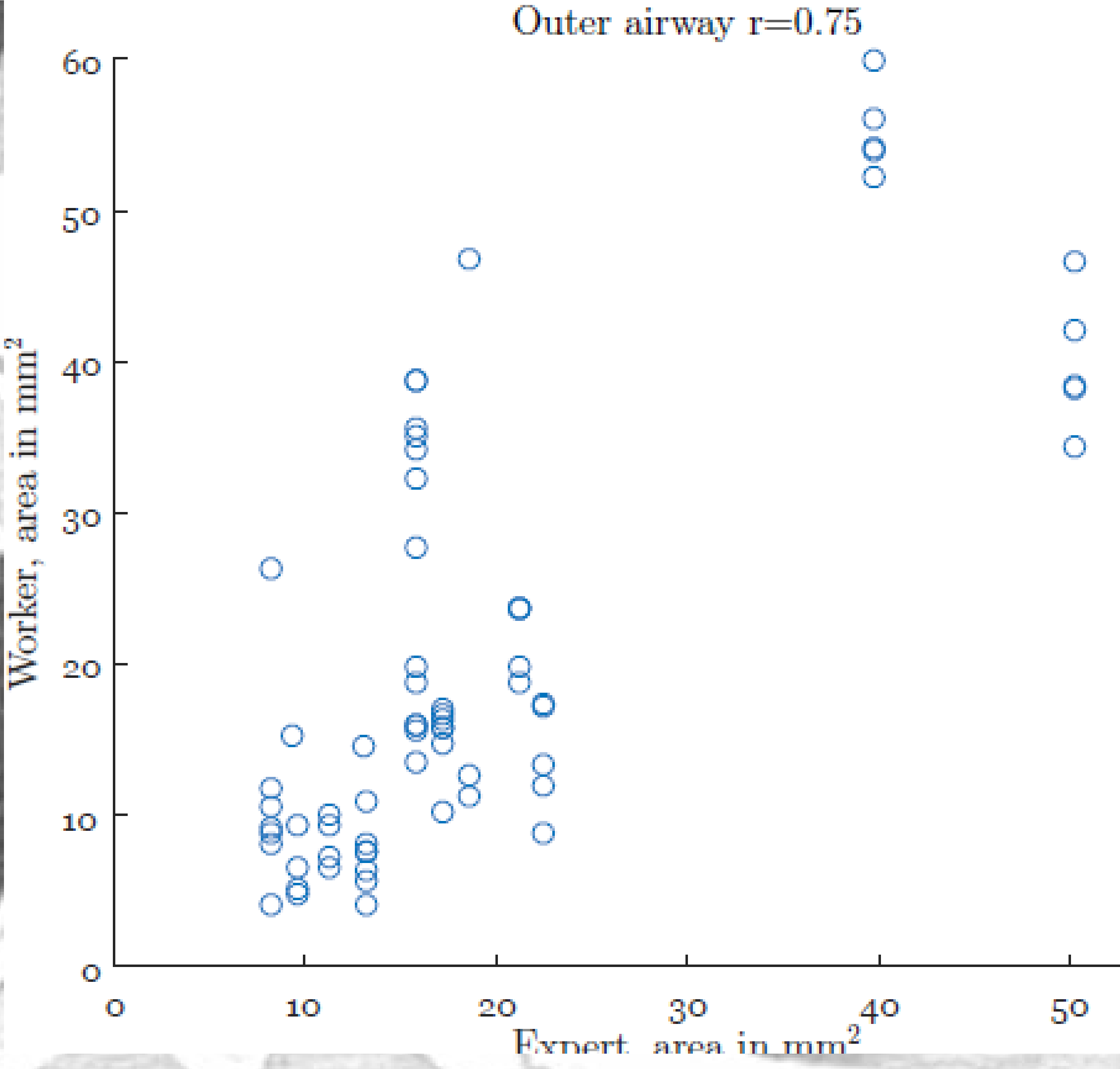
Richard M. Levenson[1]*, Elizabeth A. Krupinski[3], Victor M. Navarro[2], Edward A. Wasserman[2]*



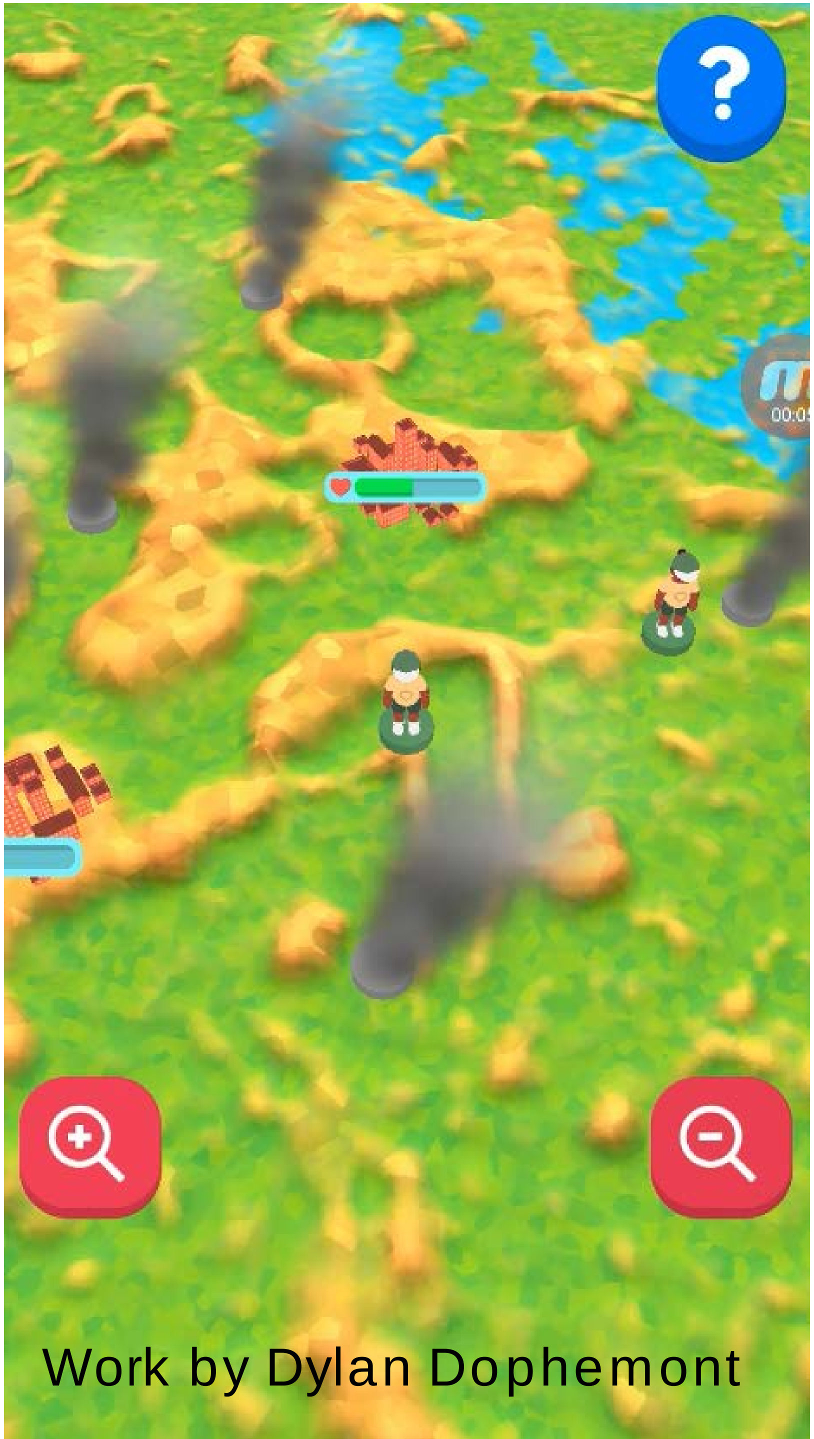


Crowdsourcing airway annotations



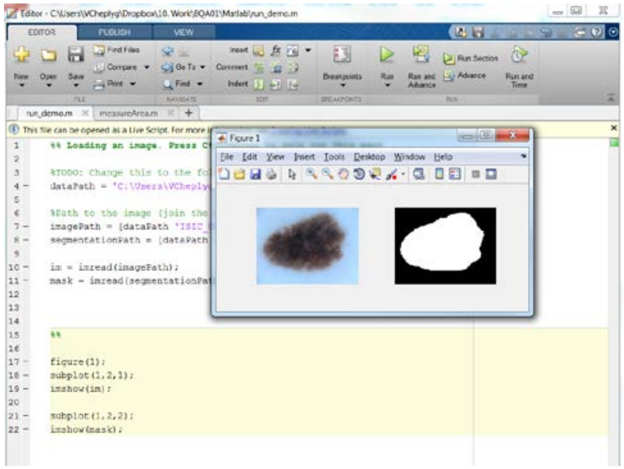


Cheplygina, V et al. (2016). Early Experiences with Crowdsourcing Airway Annotations MICCAI LABELS 2016


00:0
Work by Dylan Dophemont

03:08
1
2
OK

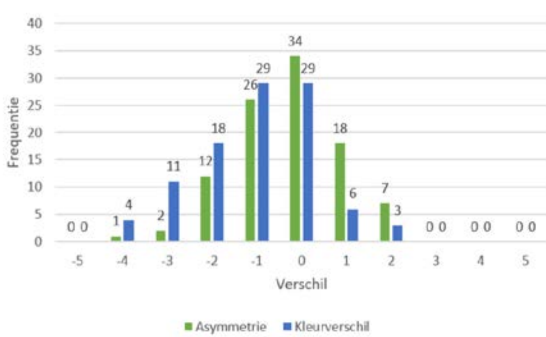
1. Measure features with algorithms



2. Measure features yourself

	A	B	C	D
1	ID	Asymmetry_7_1	Color_7_1	Border_7_1
2	ISIC_0000549	2	4	1
3	ISIC_0000550	1	3	1
4	ISIC_0000551	2	2	1
5	ISIC_0000552	1	4	1
6	ISIC_0000554	2	3	1
7	ISIC_0000555	2	3	1
8	ISIC_0001100	2	5	1
9	ISIC_0001102	2	5	1
10	ISIC_0001103	1	5	1
11	ISIC_0001105	0	2	1
12	ISIC_0001118	2	5	1
13	ISIC_0001119	2	3	1
14	ISIC_0001126	2	2	1
15	ISIC_0001128	1	3	1
16	ISIC_0001131	1	5	1
17	ISIC_0001133	1	5	1
18	ISIC_0001134	2	3	1
19	ISIC_0001140	2	2	1
20	ISIC_0009923	1	2	1
21	ISIC_0009925	2	2	1
22	ISIC_0009929	1	2	1
23	ISIC_0009930	1	2	1
24	ISIC_0009931	1	3	1
25	ISIC_0009932	2	3	1
26	ISIC_0009933	1	2	1
27	ISIC_0009935	1	3	1
28	ISIC_0009936	1	2	0

3. Evaluate



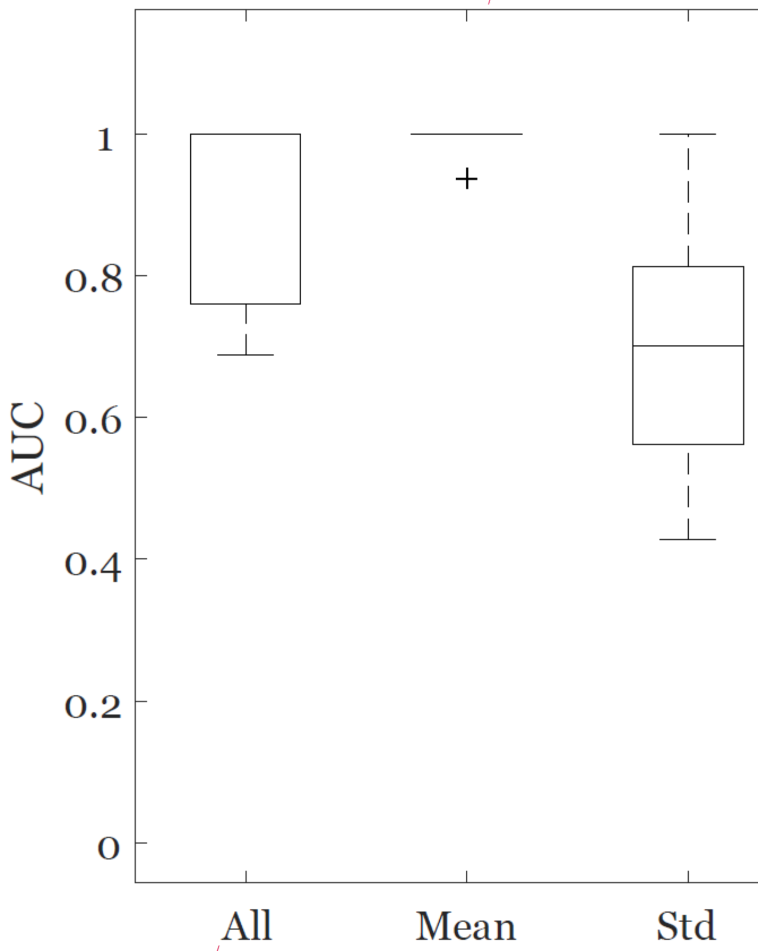
Grafiek 1: De frequenties van de verschilwaardes tussen de metingen in Matlab en de metingen op het oog

Crowdsourcing!

100 annotated images

- 5 features annotated by 6 people = 30 features
- Predict healthy vs melanoma **without** images

- Mean of annotators best, but “disagreement” (standard deviation) also informative

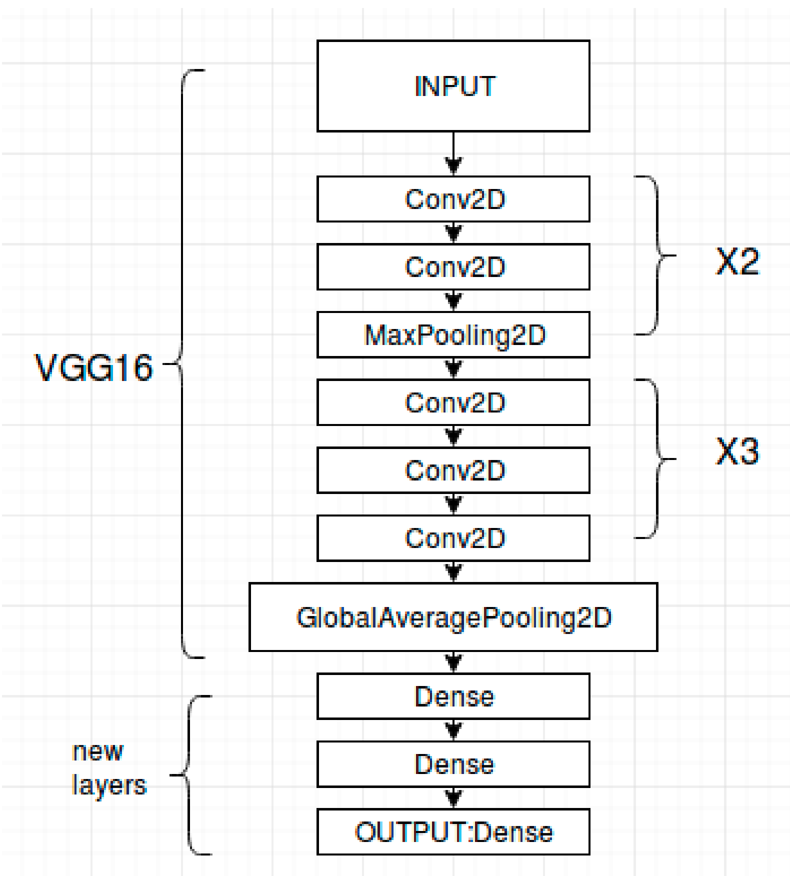


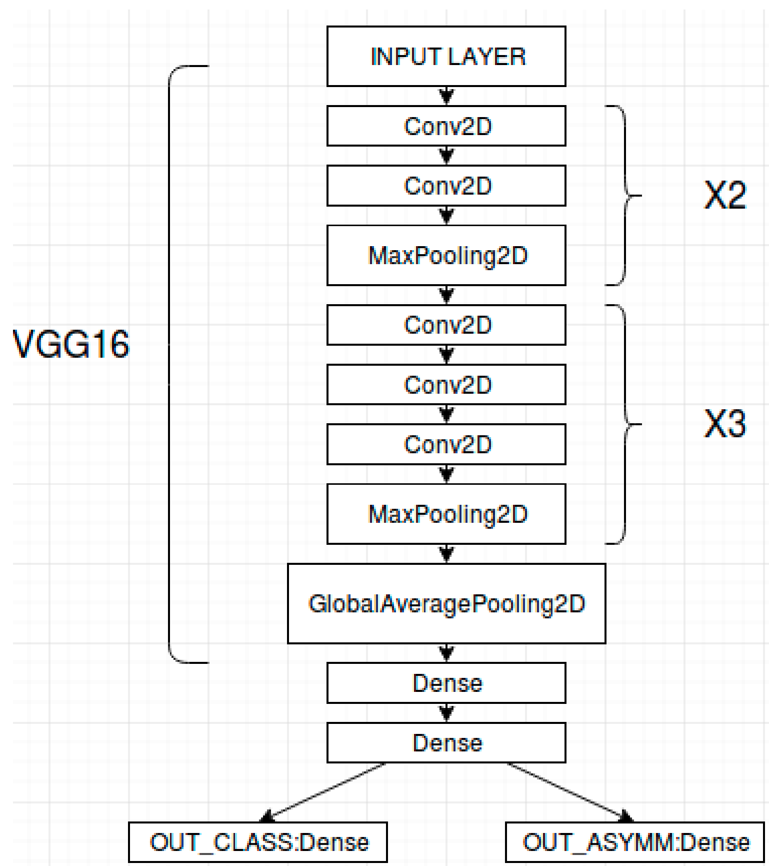
Cheplygina, V., & Pluim, J. P. W. (2018). Crowd disagreement about medical images is info
In *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Bi Data and Expert Label Synthesis* (pp. 105-111).

Work by Elif Kubra Contar

Same network

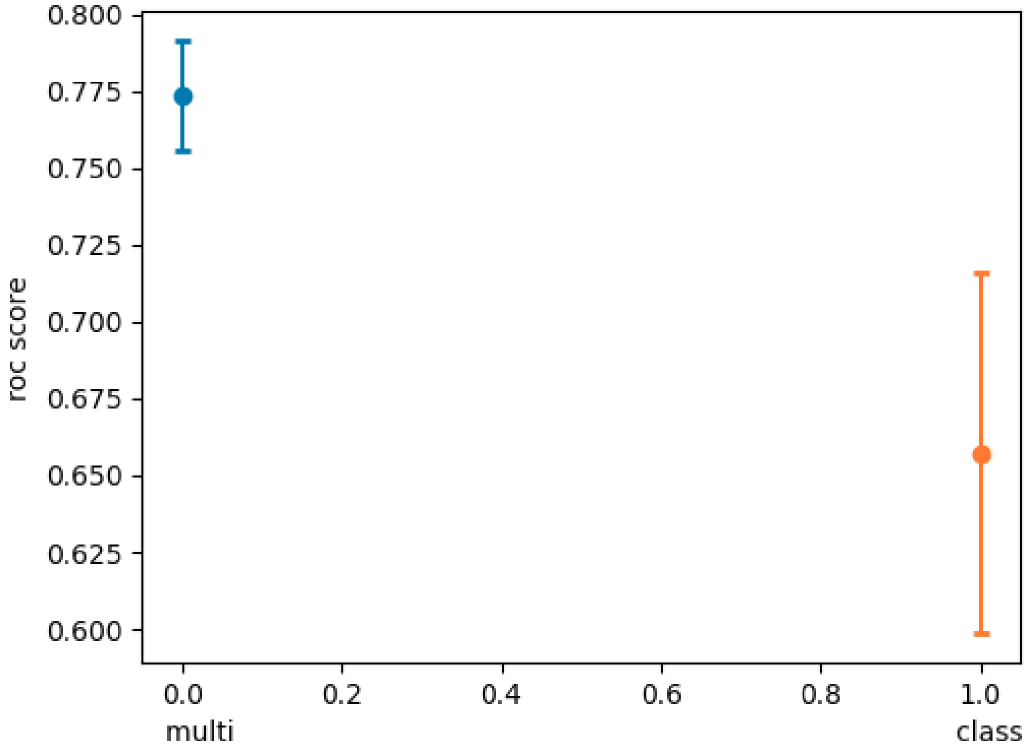
- Single-task with class label
- Multi-task with class label and asymmetry





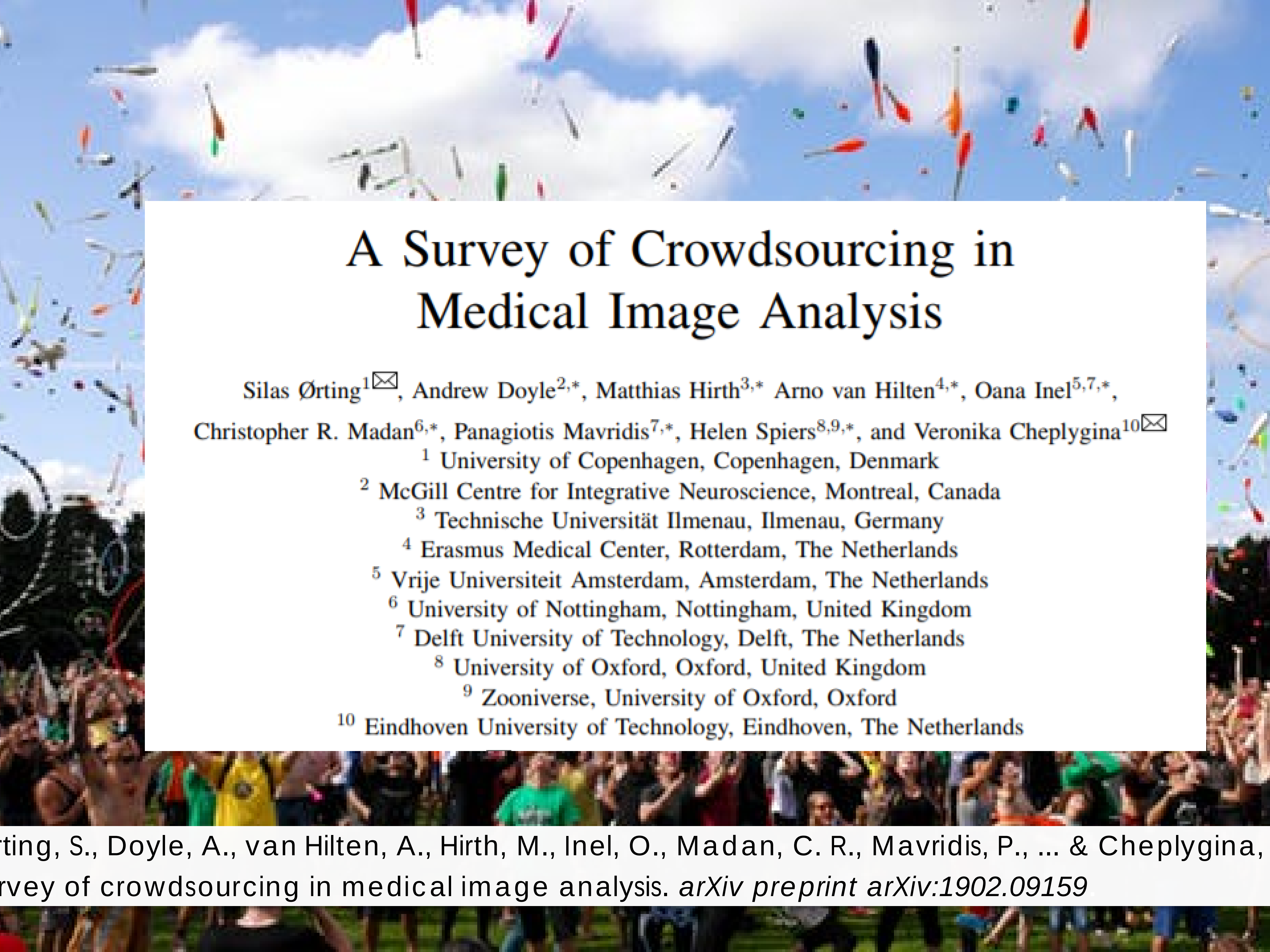
Work by Elif Kubra Contar

Multi-task network outperforms single-task network



@drveronikach

http://www.veronikach.com



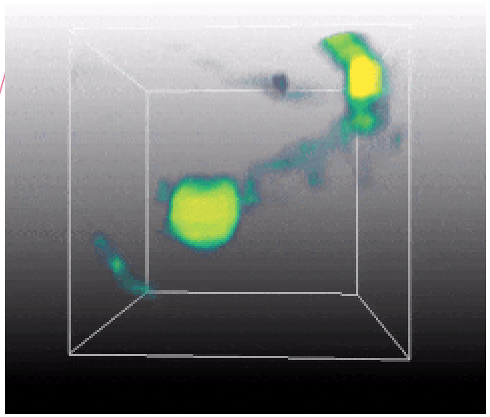
A Survey of Crowdsourcing in Medical Image Analysis

Silas Ørting[1]✉, Andrew Doyle[2,*], Matthias Hirth[3,*] Arno van Hilten[4,*], Oana Inel[5,7,*],
Christopher R. Madan[6,*], Panagiotis Mavridis[7,*], Helen Spiers[8,9,*], and Veronika Cheplygina[10]✉
[1] University of Copenhagen, Copenhagen, Denmark
[2] McGill Centre for Integrative Neuroscience, Montreal, Canada
[3] Technische Universität Ilmenau, Ilmenau, Germany
[4] Erasmus Medical Center, Rotterdam, The Netherlands
[5] Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[6] University of Nottingham, Nottingham, United Kingdom
[7] Delft University of Technology, Delft, The Netherlands
[8] University of Oxford, Oxford, United Kingdom
[9] Zooniverse, University of Oxford, Oxford
[10] Eindhoven University of Technology, Eindhoven, The Netherlands

rting, S., Doyle, A., van Hilten, A., Hirth, M., Inel, O., Madan, C. R., Mavridis, P., ... & Cheplygina,
rvey of crowdsourcing in medical image analysis. *arXiv preprint arXiv:1902.09159*

Next

- Other features / annotators (1400+ images)
- Crowdsourcing for lung nodules (3D)
- Evaluate the evaluation



Not-so-supervised academics

2011-2014 PhD

"Publish papers"

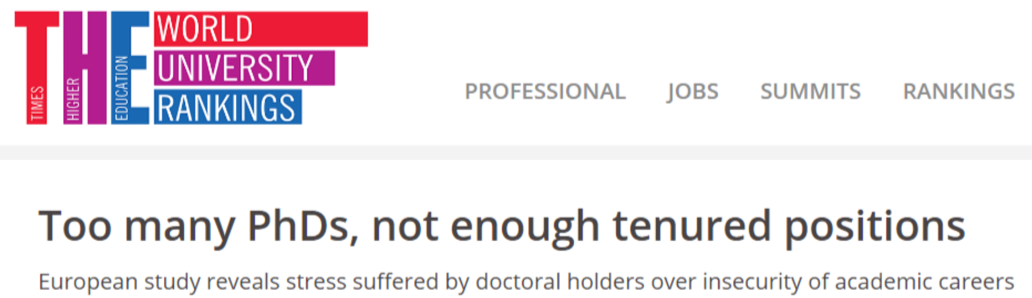
- Science vs prestige

"Good for your CV"

- Experience vs time

Maybe academic career?

- “You have to go abroad”
- Not good enough?
- Mentors!



2015

TU Delft Delft University of Technology

Erasmus MC Universitair Medisch Centrum Rotterdam

2015 – 2016

- “Publish, develop own research + get funding for it”

- Social media
  - Impact, community

https://pixabay.com/en/mountain-climbing-mountaineer-802099/

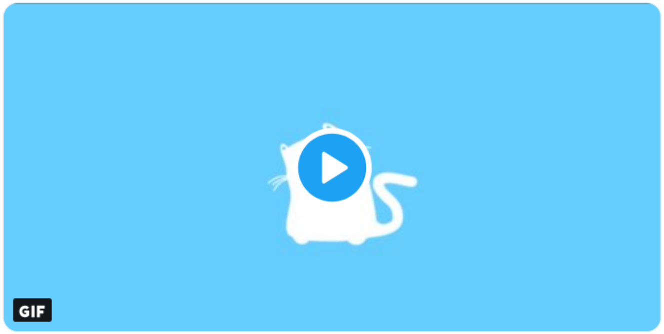
2017

https://veronikach.com/progress-reports/getting-tenure-track-

2017

Dr Veronika Cheplygina
@vcheplygina

Excited to announce that I'll join the Medical Image Analysis group @TUeindhoven as assistant professor in Feb 2017! veronikach.com/news/



5:23 PM - 26 Nov 2016

2 Retweets 32 Likes

13 2 32

1. Excitement

I get to do research and teach and learn from others for the next 5 years! How amazing have so many ideas, I can't wait!

2. Relief

I get to have a job for 5 years and don't have to apply for positions for like, a very long t started looking for my next position halfway through my postdoc, which was a job in its not reflect well on my postdoc project. A few things were not really going well for me in the news about the position couldn't have come at a better time.

3. Fear

I worry they will discover I'm an impostor and they should have hired somebody else. I reassure myself by thinking that if I'm an impostor and they are the the real deal, they figured out that I was one already. But I also worry about just being able to handle it all.

4. Guilt

As many other researchers are forced out of academia, I feel guilty for "surviving" whil "good, but not excellent CV" (citing reviews on some of my rejected grant applications have to deal with hundreds of rejections – I applied to four jobs, interviewed for three, offered one. Sure, I worked hard, but I think luck and privilege played a big role.

5. Hope

I get to be one step closer to maybe one day being able to change things, just a little bi

https://veronikach.com/progress-reports/getting-tenure-track-

2017+ Not without challenges

https://pixabay.com/en/mountain-climbing-mountaineer-802099/

2017+ Find support

https://unsplash.com/photos/iuqmGmst5Po

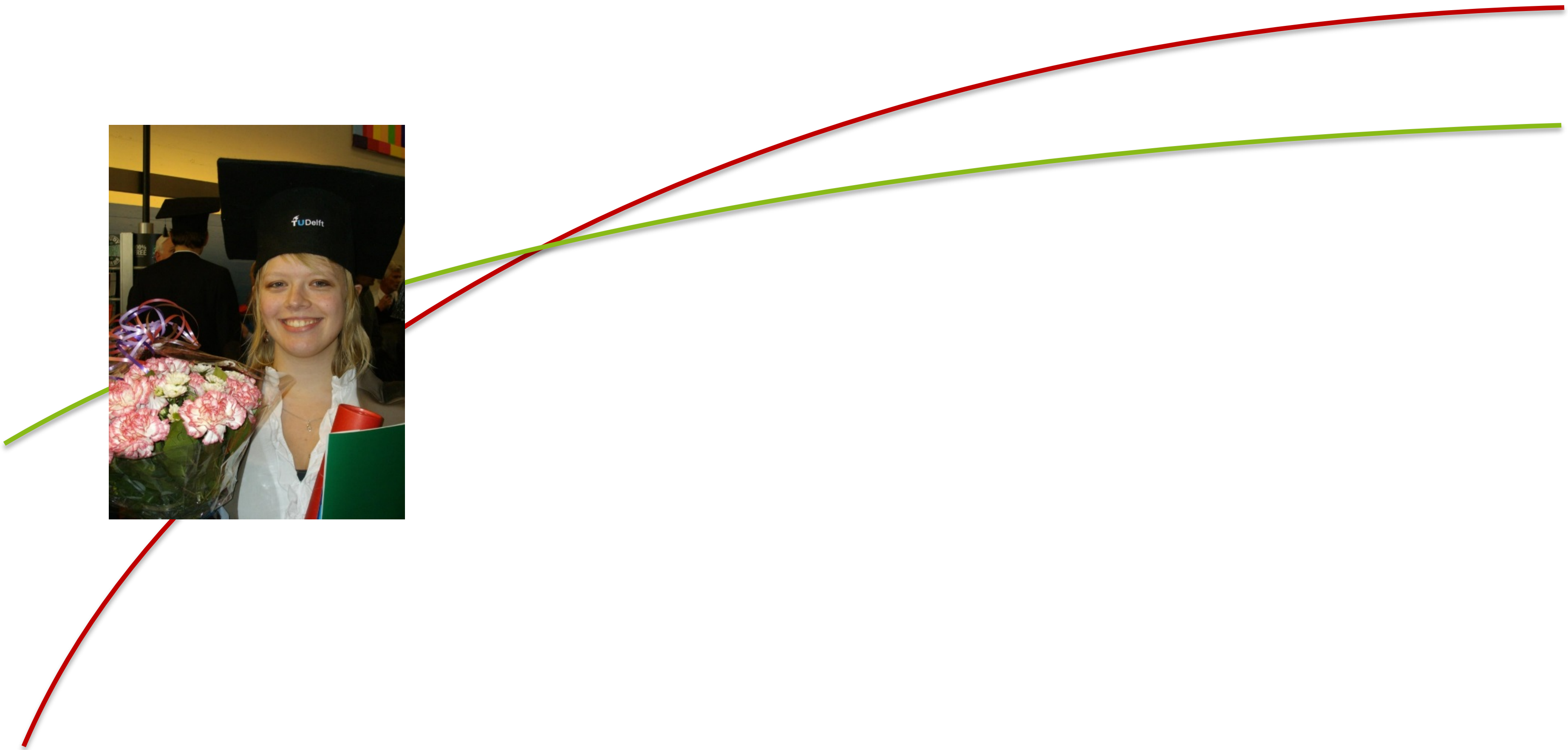
Academia as supervised learning?

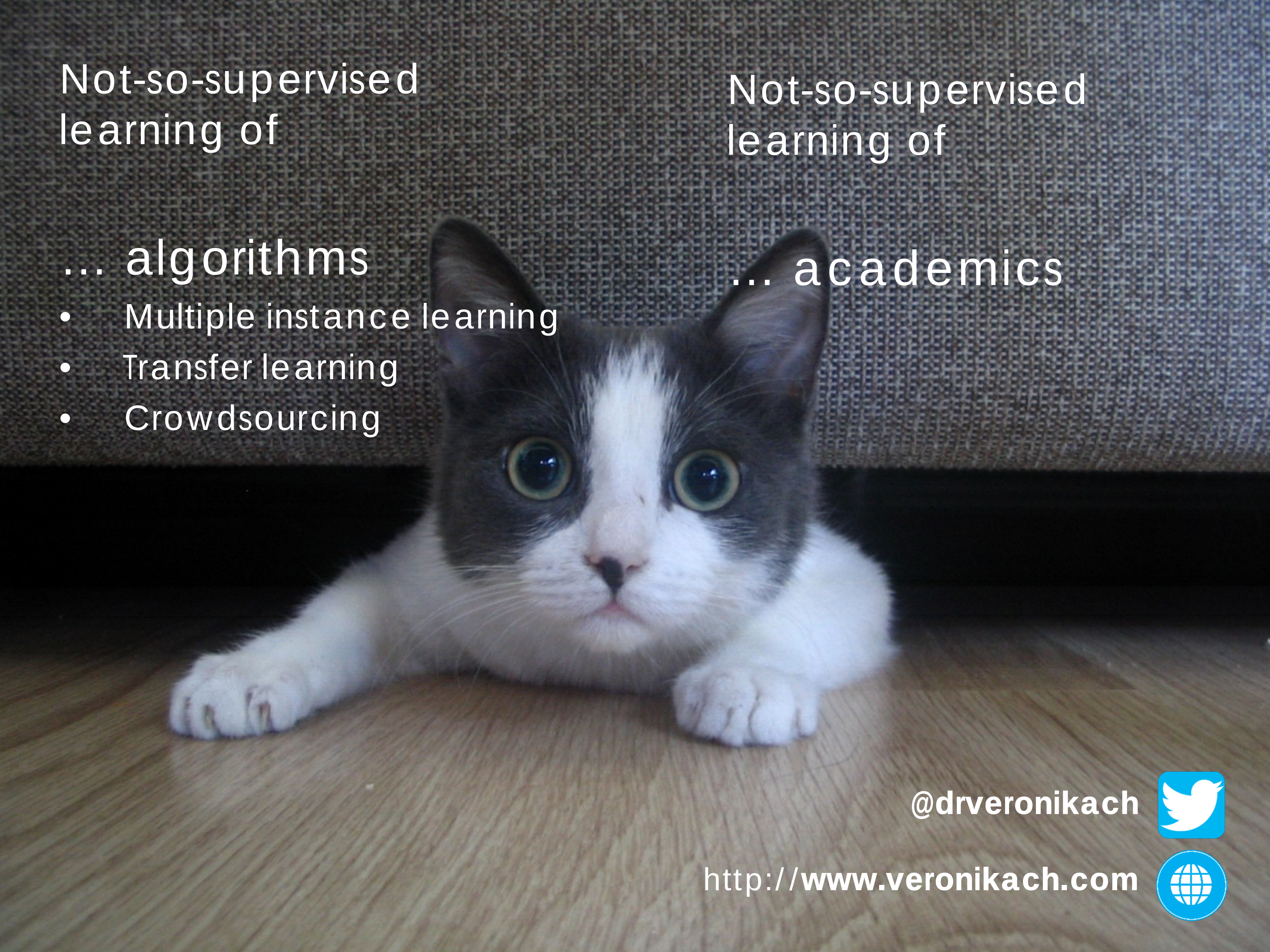
- Input = CV at time *t*
- Output = Success / failure at *t+1*
- Successes at *t+1* define “decision tree”

But CVs ≠ true data distribution

- Input space is much larger (Shadow CV)
- Output space is much larger (Impact, being happy)
- Noisy labels, many unlabelled inputs
- Overfitting!

Learning curve





Not-so-supervised learning of

... algorithms

- Multiple instance learning
- Transfer learning
- Crowdsourcing

Not-so-supervised learning of

... academics

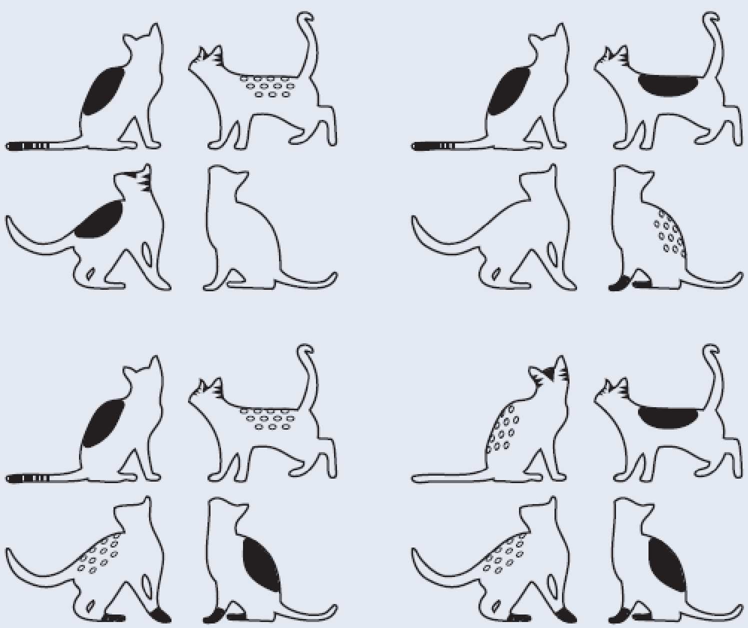
@drveronikach

http://www.veronikach.com


THANKS
@drveronikach
http://www.veronikach.com

Dissimilarity-Based Multiple Instance Learning

What is different about the groups of cats on the front cover, from the groups on the back cover? If you can answer this question, you are probably also able to categorize another, previously unseen to you, group of cats. This thesis is about different applications where similar puzzles may occur, and how some machine learning algorithms approach such problems.



ISBN 978-94-6295-192-1

Dissimilarity-Based Multiple Instance Learning

Veronika Cheplygina

Dissimilarity-Based Multiple Instance Learning

Veronika Cheplygina