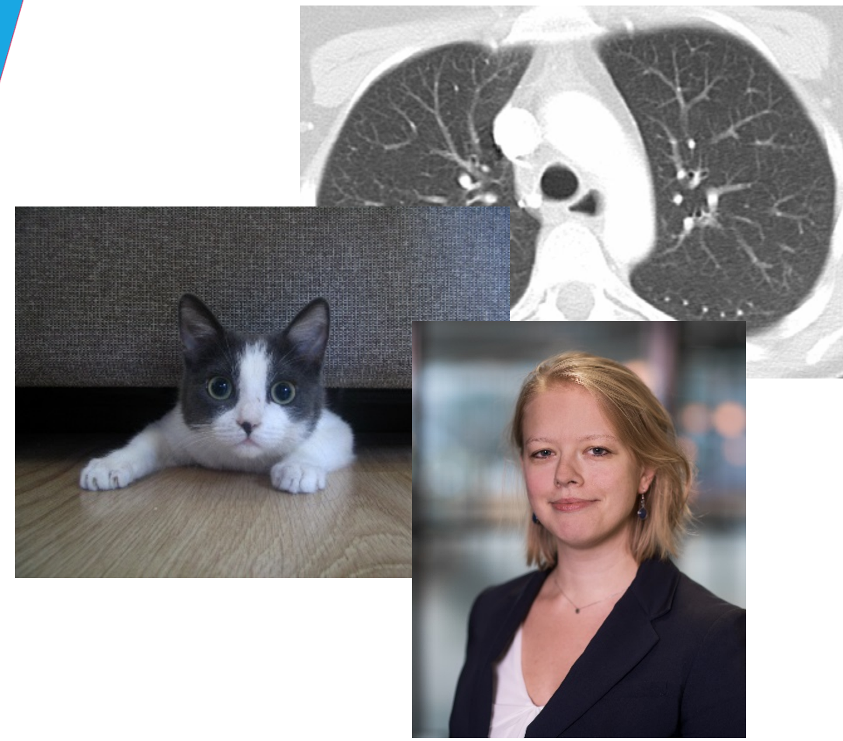
Technische Universiteit **Eindhoven** University of Technology

Multiple instance learning in biomedical applications

Veronika Cheplygina



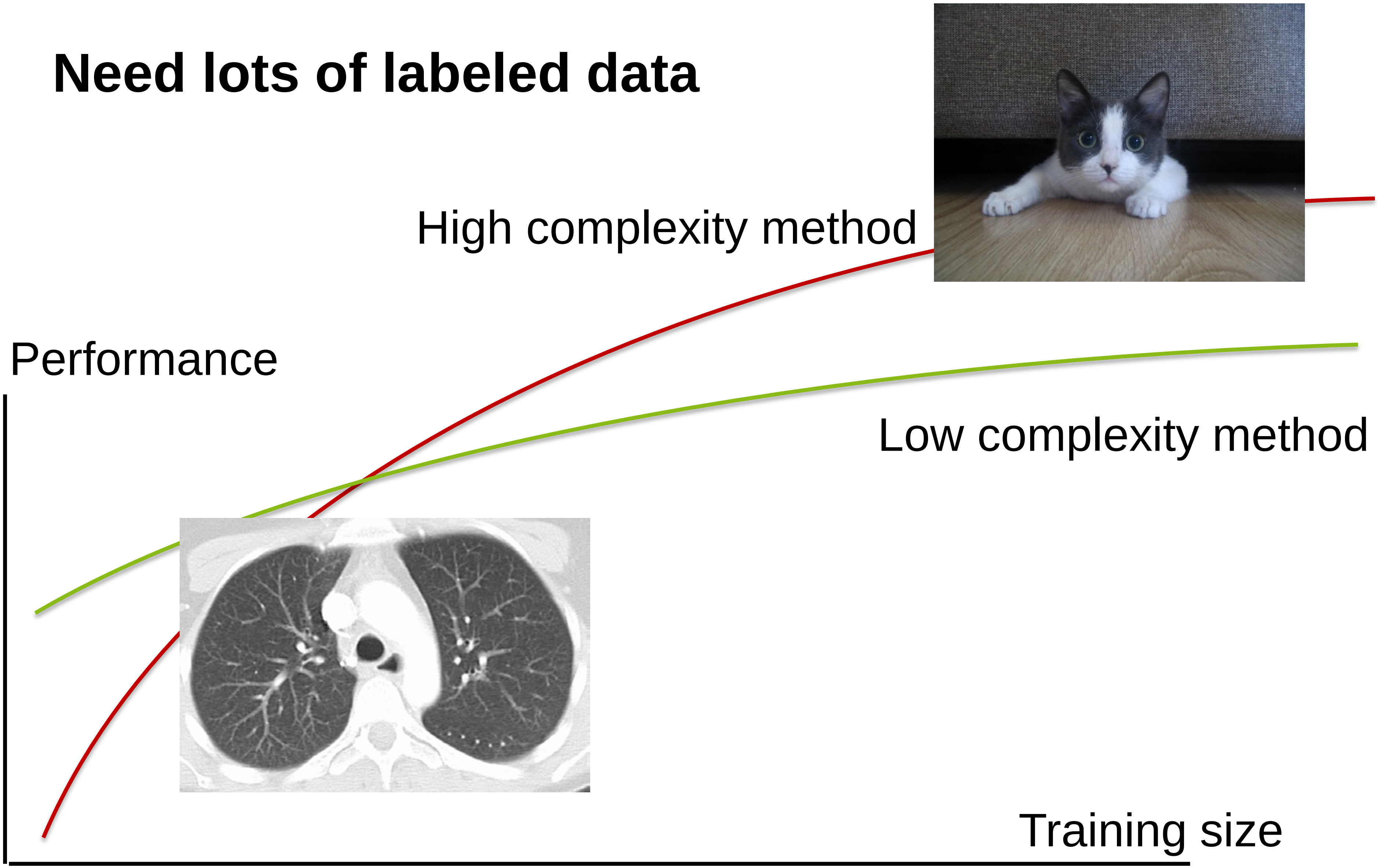
@drveronikach

http://www.veronikach.com

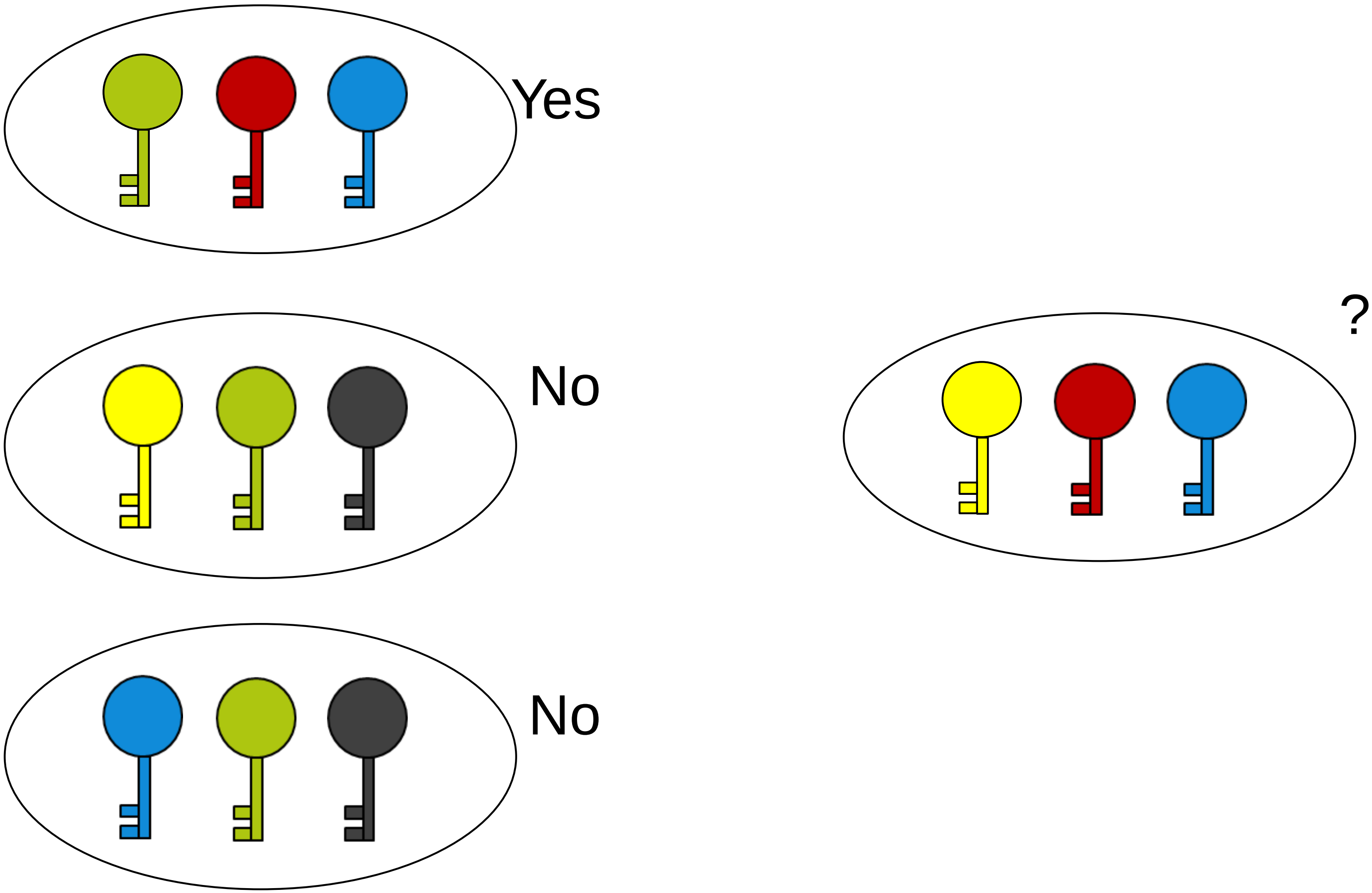
Outline

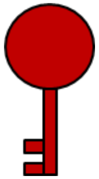
- Multiple instance learning
  - Problem definition
  - Classifiers
- Examples
  - Medical images
  - Bioinformatics
- Challenges

Need lots of labeled data



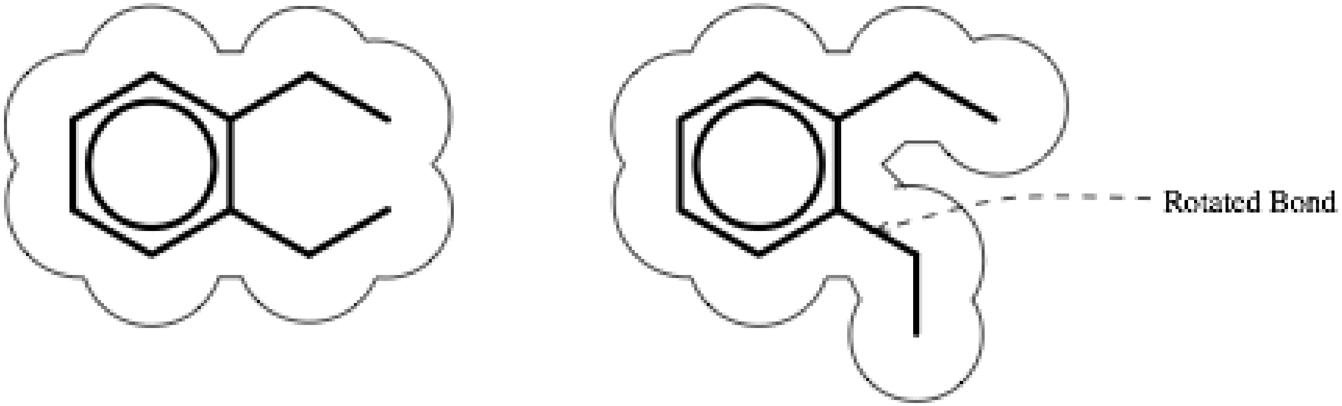
Use labels that are easier to obtain





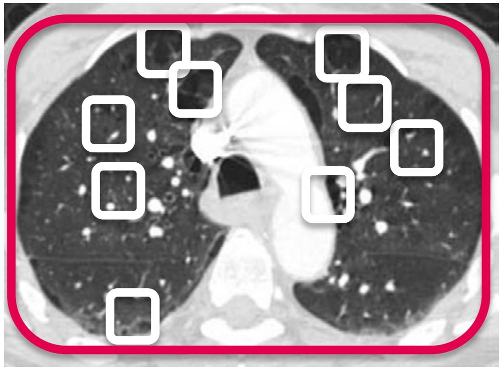
Yes

MIL Applications



- Predict molecule activity
- Different number of conformations per molecule
- Molecule is active *iff* 1 or more conformations are active

MIL Applications

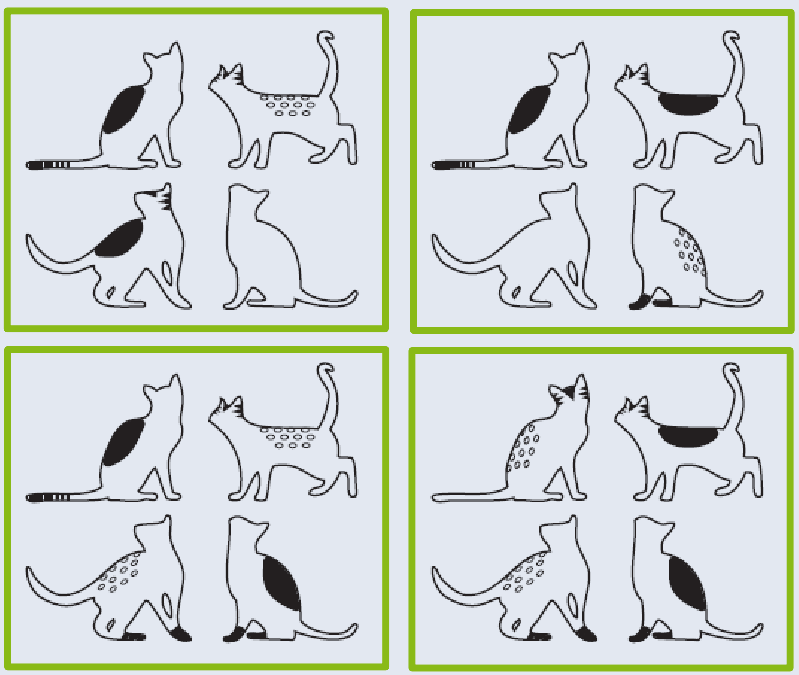


Weak annotations

- At least 1 abnormal patch = abnormal scan
- For a new scan:
  - What is the diagnosis?
  - Where are the abnormalities?

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.



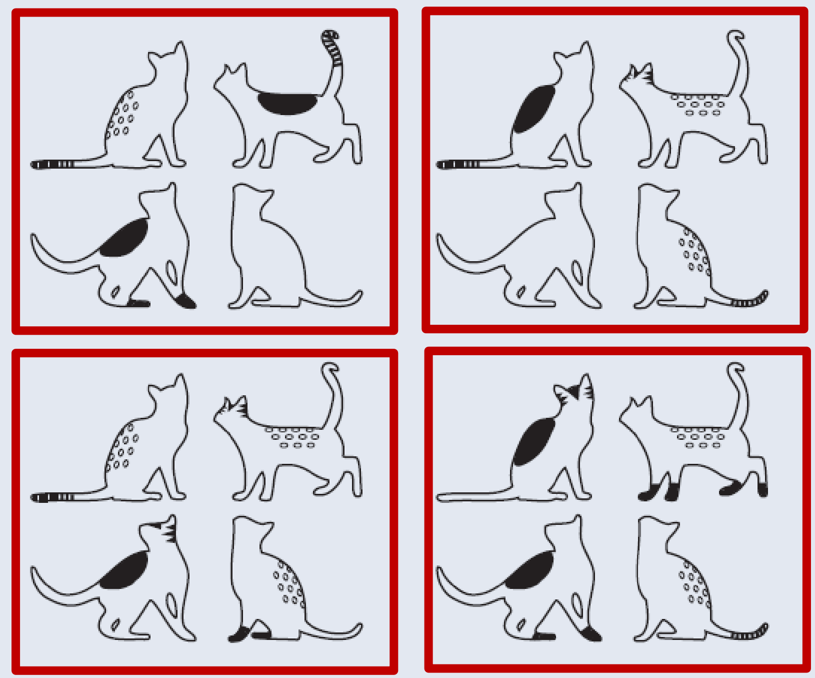
ISBN 978-94-6295-192-1

Dissimilarity-Based Multiple Instance Learning

Veronika Cheplygina

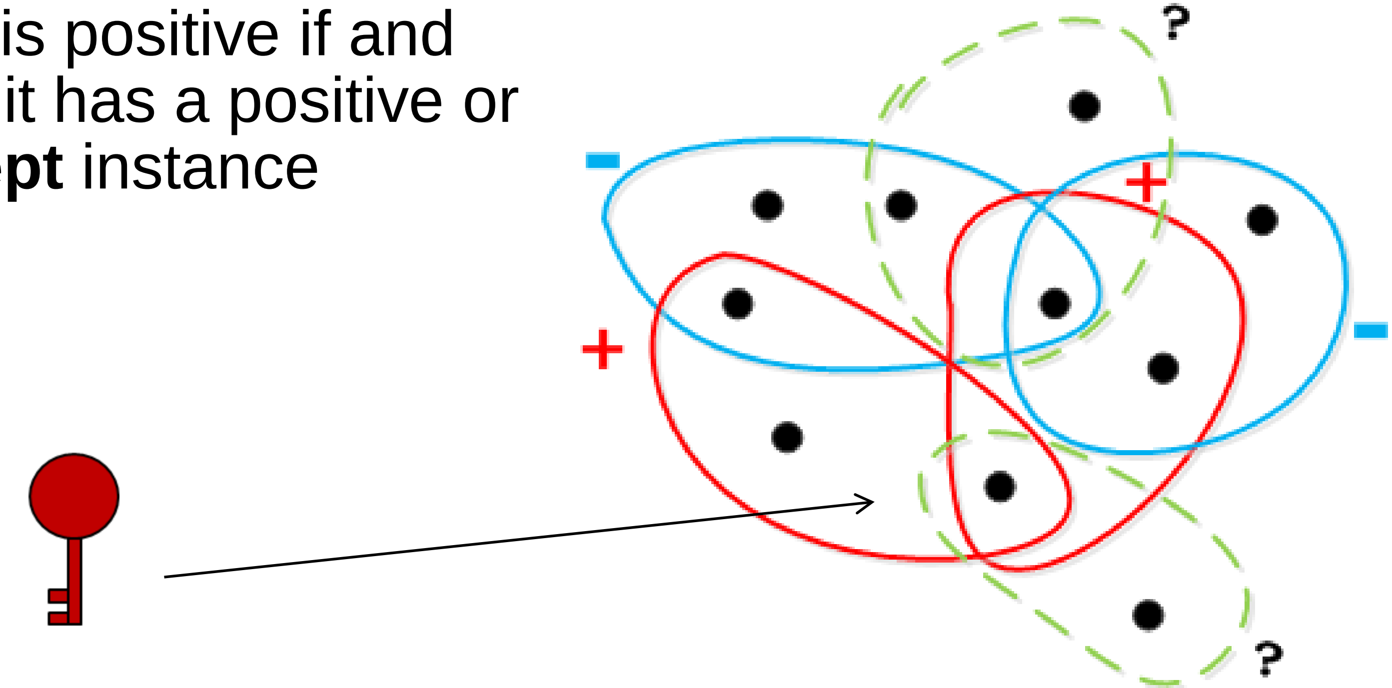
Dissimilarity-Based Multiple Instance Learning

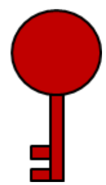
Veronika Cheplygina



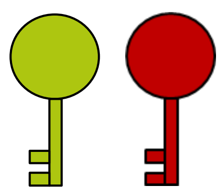
Multiple Instance Learning

- **Bag** of **instances** (feature vectors)
- A bag is positive if and only if it has a positive or **concept** instance





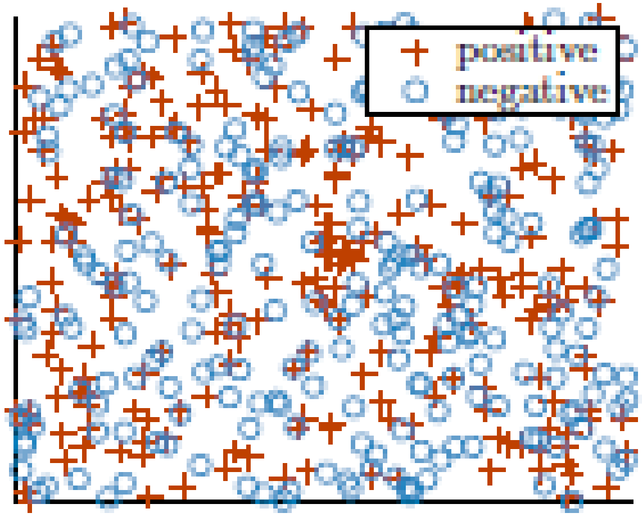
Yes



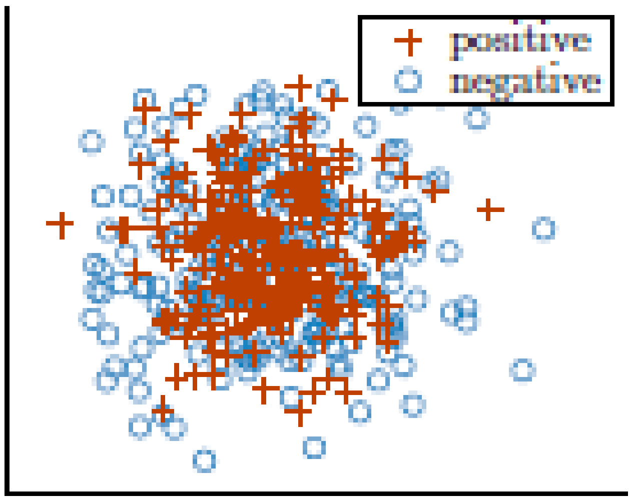
No

Assumptions

Many possible assumptions that fit in this paradigm



(a) Concept



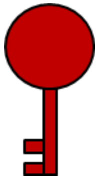
(b) Distribution

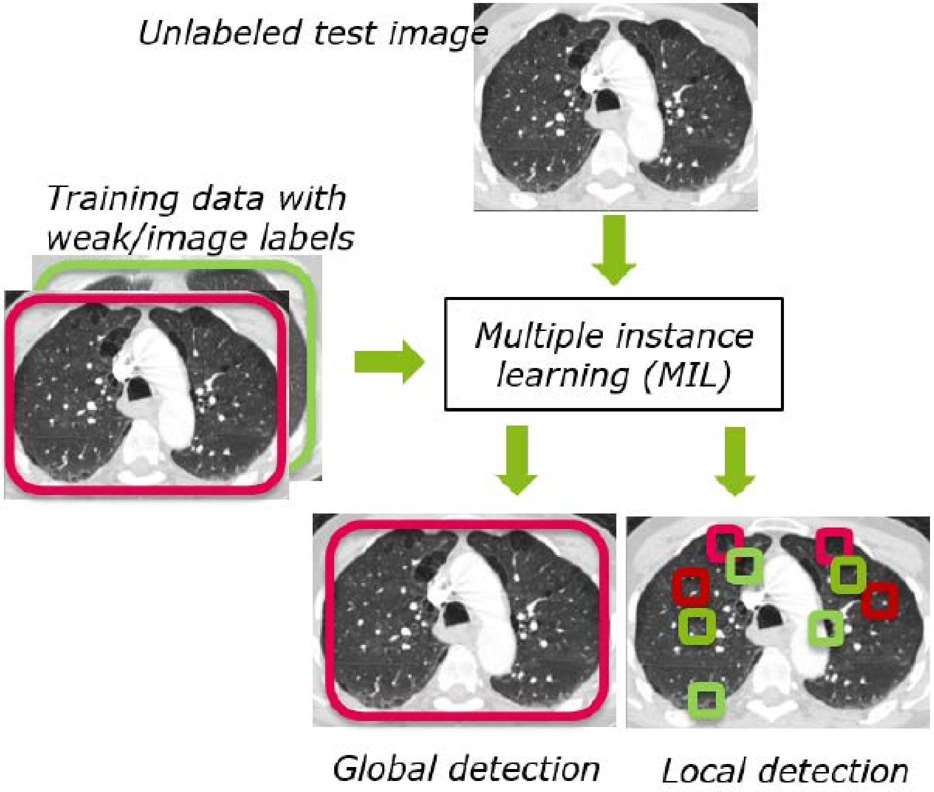
Cheplygina, V., Tax, D. M. J., & Loog, M. (2015). Multiple instance learning with bag dissimilarities. Pattern Recognition, 48(1), 264-275.

Goal of classifier

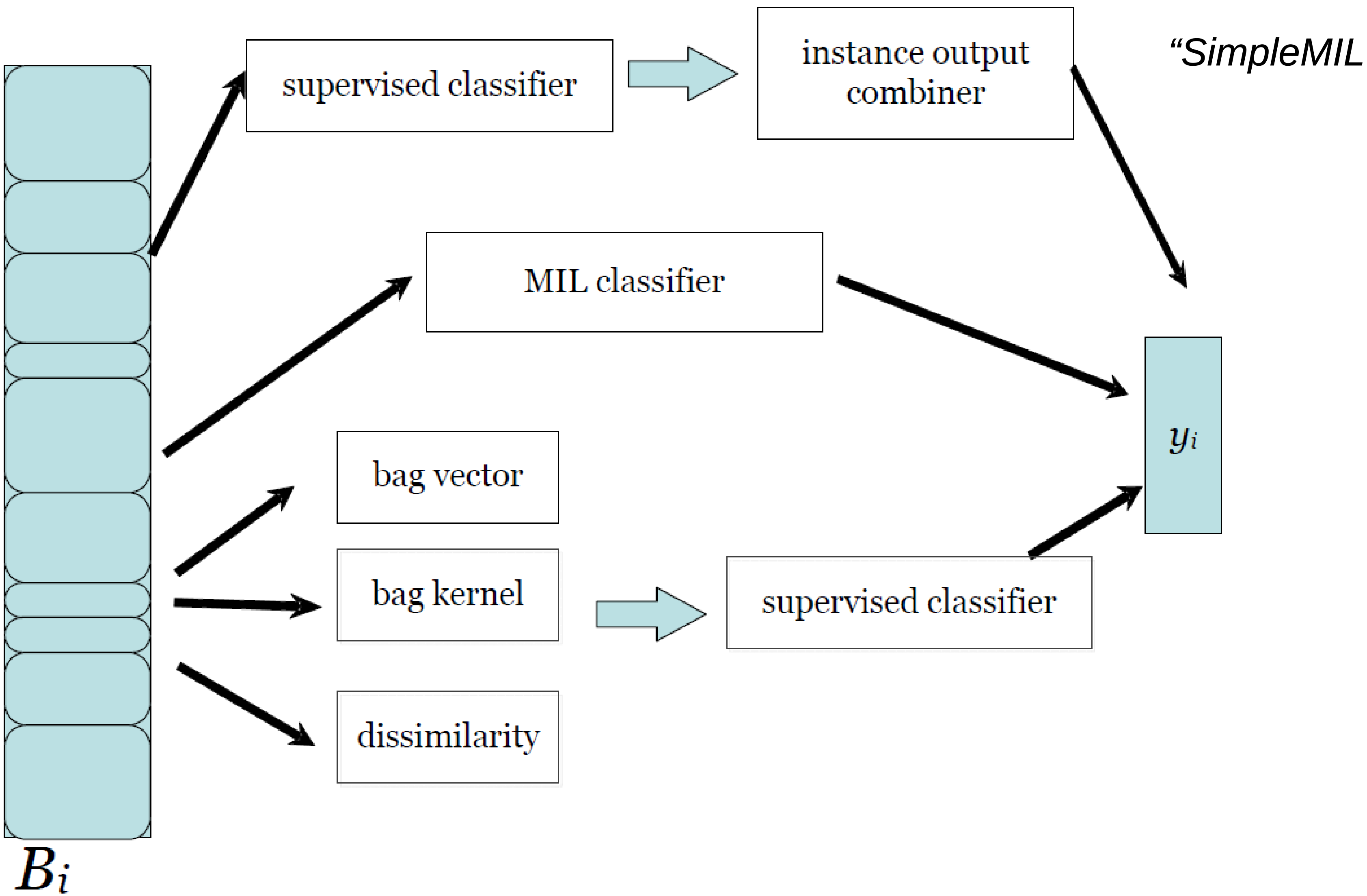
- Predict labels for previously unseen bags

- Predict labels for previously unseen instances





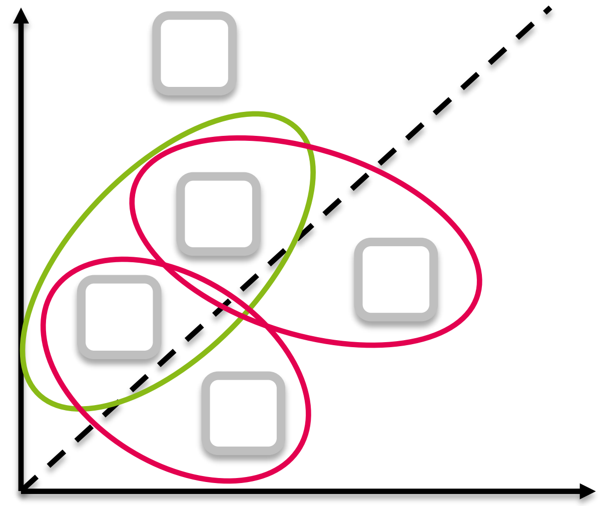
MIL classifiers



Learning: instance-level

- Find concept or instance labels, s.t. bag constraints are satisfied

- Explicit combining of instances, e.g. max or mean



- Instance labels are provided
- Not always robust

Learning: instance-level

- Supervised SVM

Classifier

Labeled instances / pixels

- "Find a plane which separates the instances well, such that the instance (pixel) labels are more or less correct"

Learning: instance-level

- miSVM

$$\min_{\{y_i\}} \min_{\mathbf{w},b,\xi} \; \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i$$

Instance labels

$$\text{s.t.} \quad \forall i: \; y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i \;, \; \xi_i \geq 0 \;, \; y_i \in \{-1, 1\} \;, \text{and}$$

$$\sum_{i \in I} \frac{y_i + 1}{2} \geq 1, \;\; \forall I \text{ s.t. } Y_I = 1, \;\; \text{and} \;\; y_i = -1, \;\; \forall I \text{ s.t. } Y_I = -1 \,.$$

Bag labels

- Find the <u>instance labels</u> and a plane …, such that …, <u>and the bag label assumption holds</u>

Learning: instance-level

- Classifier is in instance space → instance posteriors can be visualized

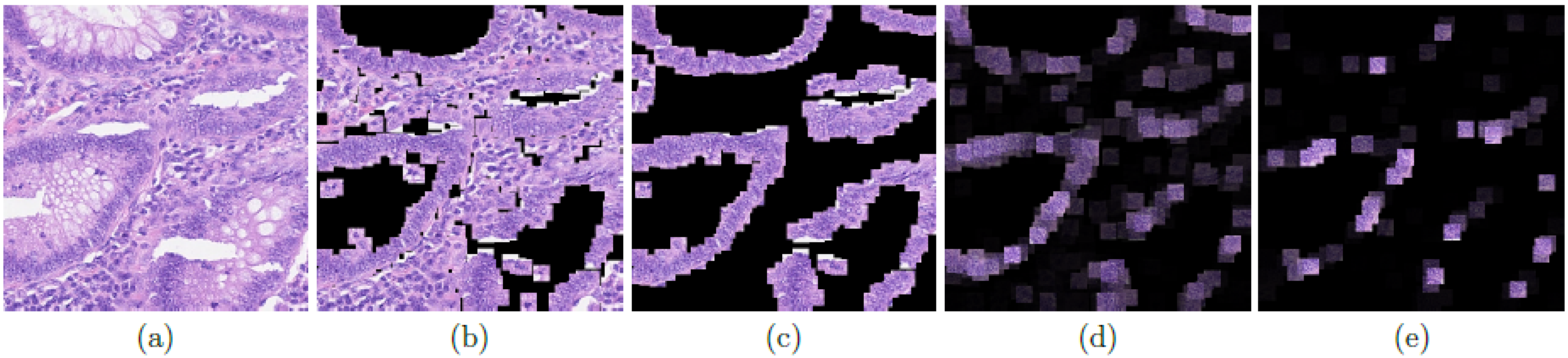
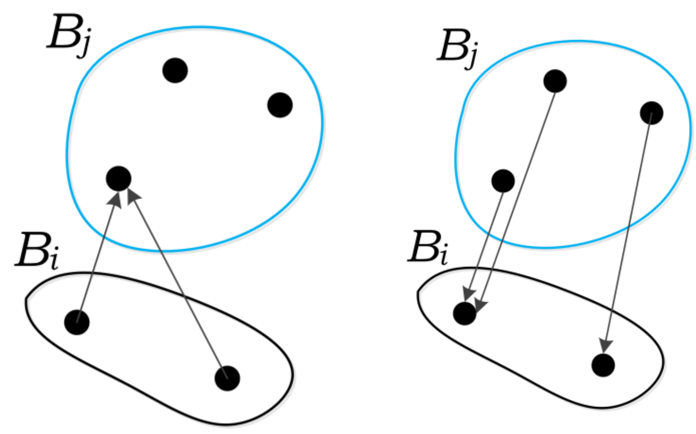


Figure 10. Colon cancer example 1: (a) H&E stained histology image. (b) 27×27 patches centered around all marked nuclei. (c) Ground truth: Patches that belong to the class epithelial. (d) Attention heatmap: Every patch from (b) multiplied by its attention weight. (e) Instance+max heatmap: Every patch from (b) multiplied by its score from the INSTANCE+max model.

Ilse, M., Tomczak, J. M., & Welling, M. (2018). Attention-based deep multiple instance learning. *arXiv preprint arXiv:1802.04712*.

Learning: bag-level



- Bag distance or kernel

- Bag “summary” + supervised classifier

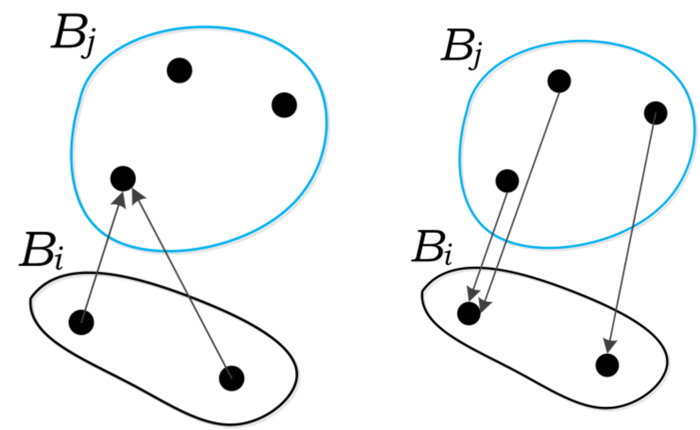
- Implicit mean combining

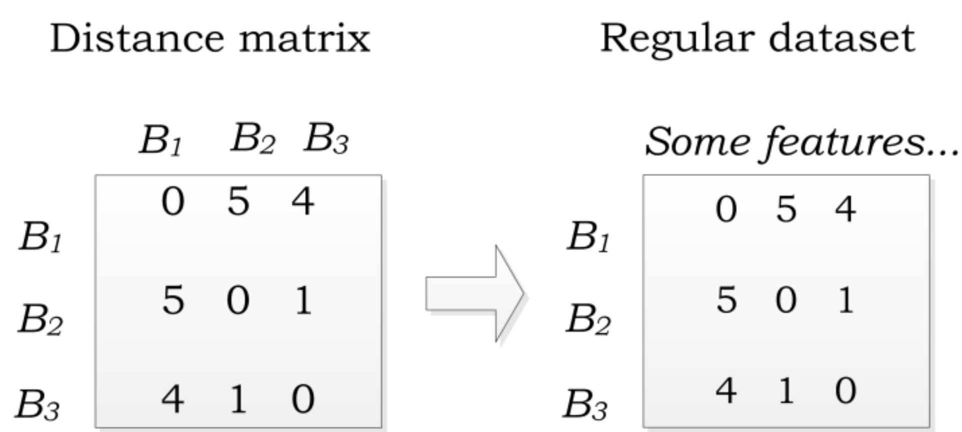
- Robust in practice
- Usually no instance labels

	B_1	B_2	B_3
B_1	0	5	4
B_2	5	0	1
B_3	4	1	0

Learning: bag-level

- Example: dissimilarity-space
- Use ALL information in distance matrix
- Good for e.g. multi-concept data

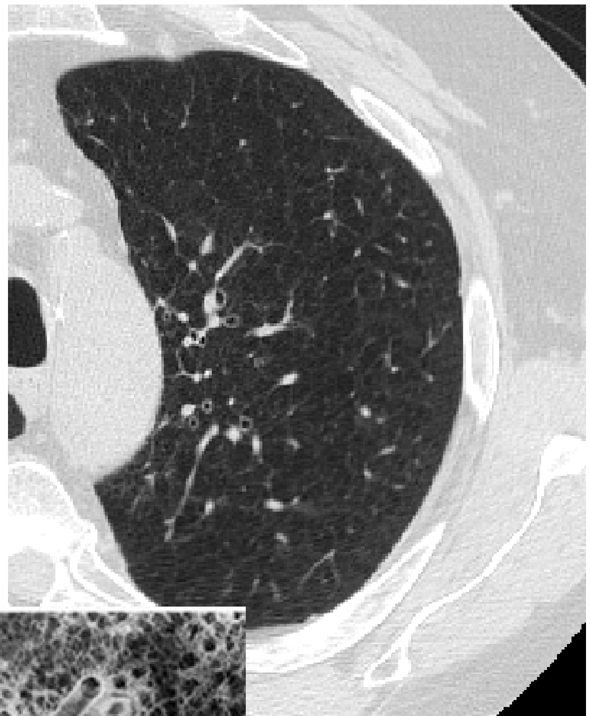


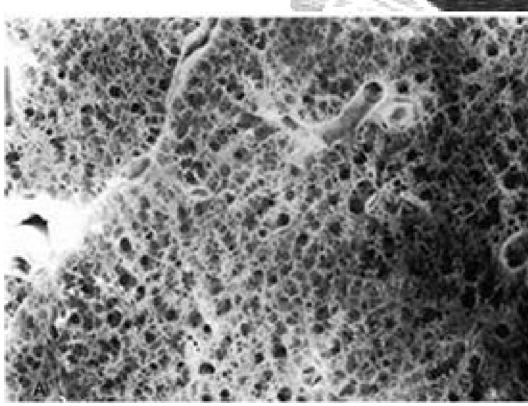


Examples

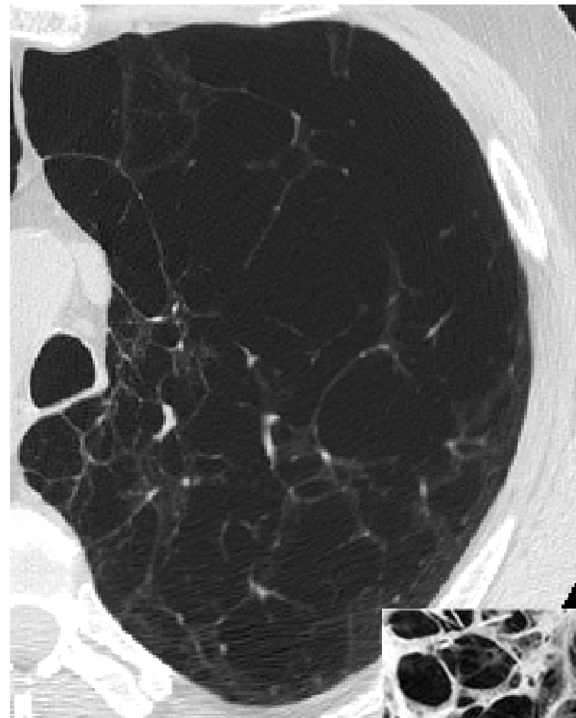
https://pixabay.com/photos/thought-idea-innovation-imagination-2123970/

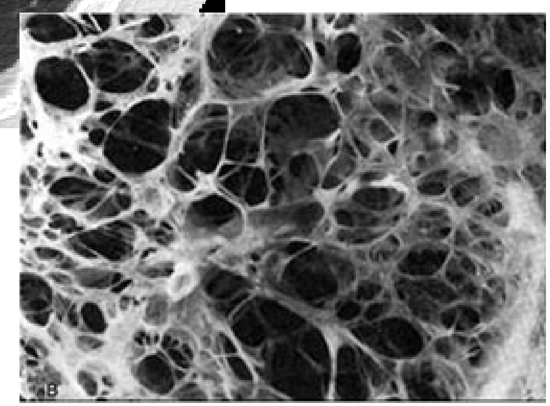
COPD in CT: Destruction of lung tissue (emphysema)





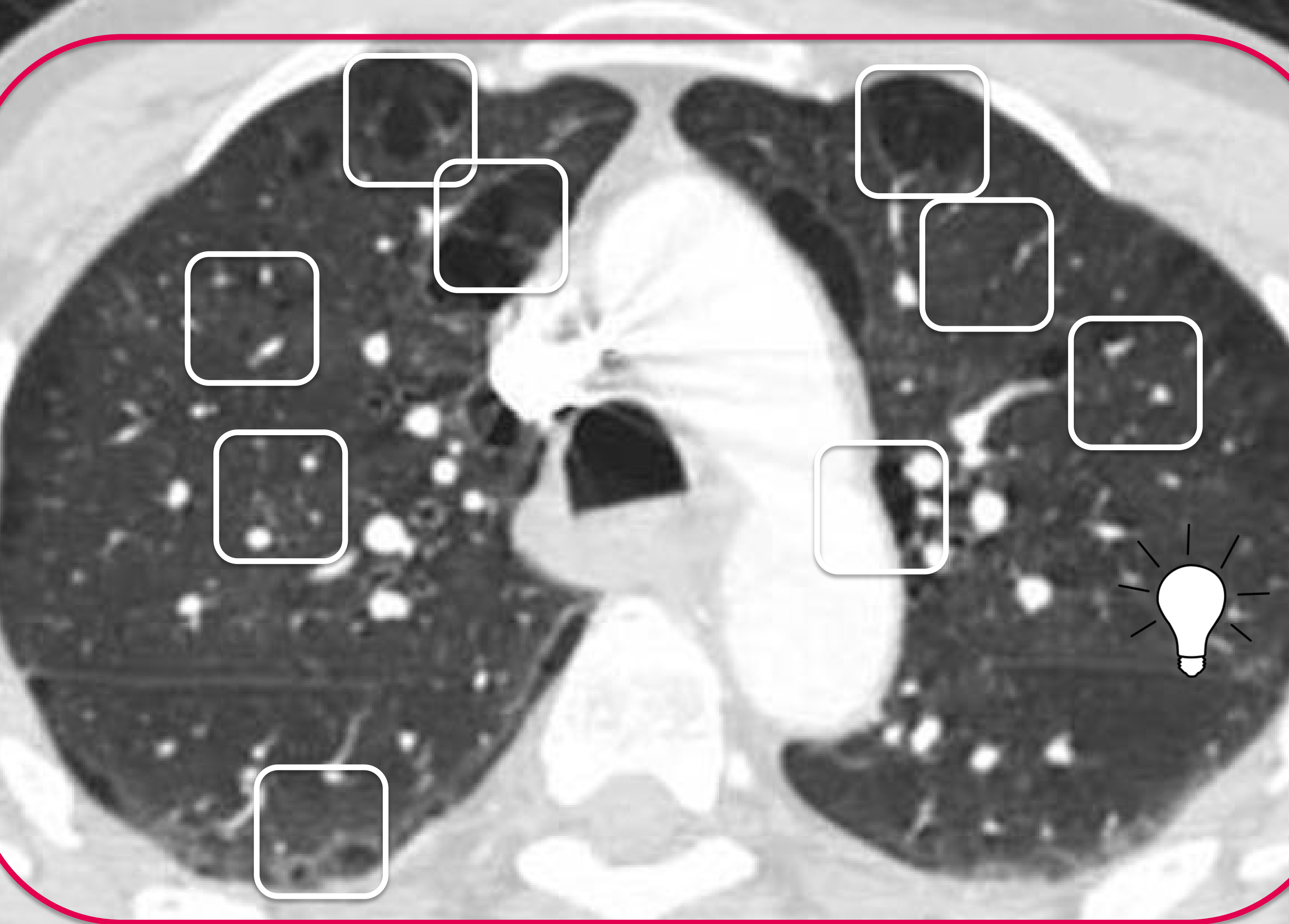
Normal





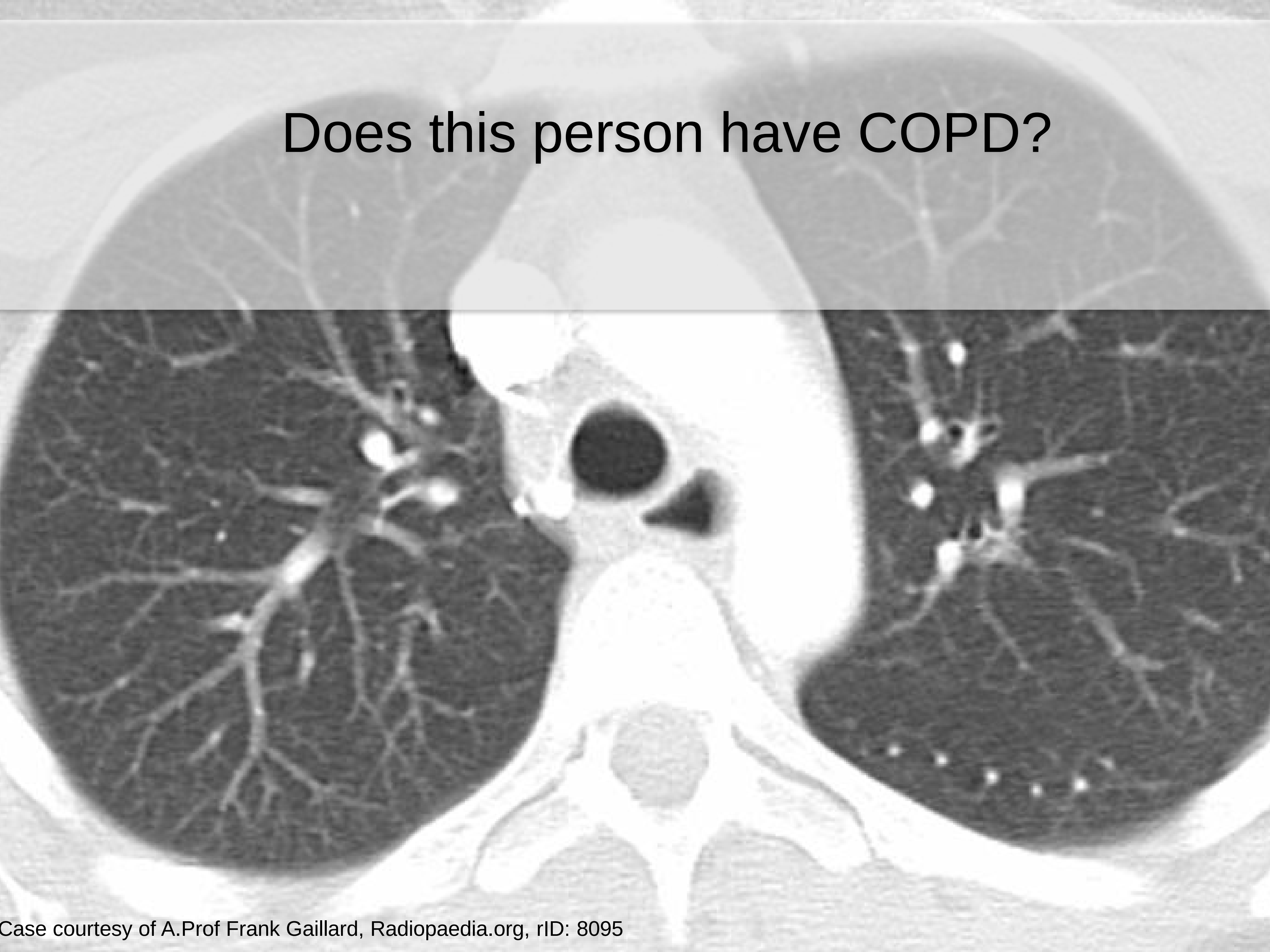
COPD

Images Marleen de Bruijne



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

Does this person have COPD?



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

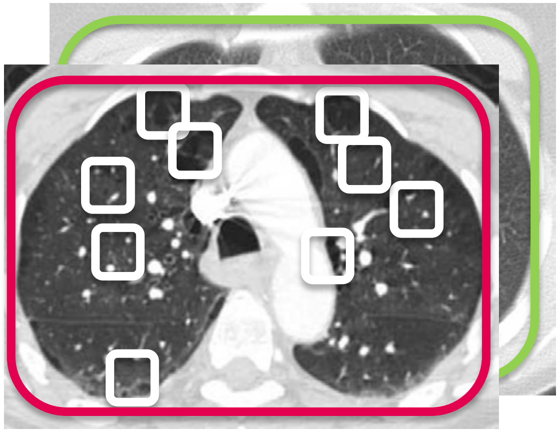
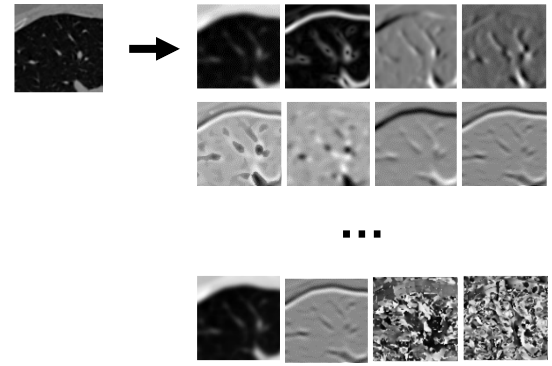
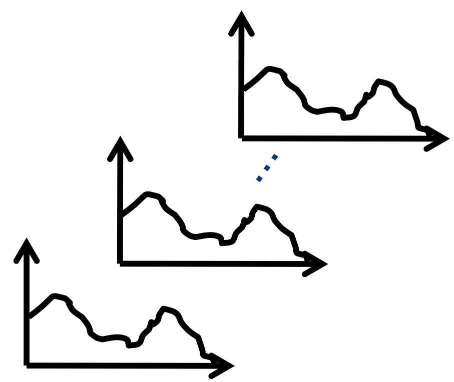


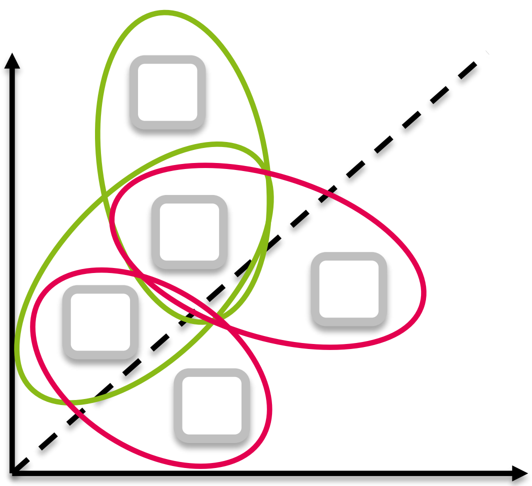
Image = COPD or not (lung function), 50 ROIs



Texture filters



Histograms



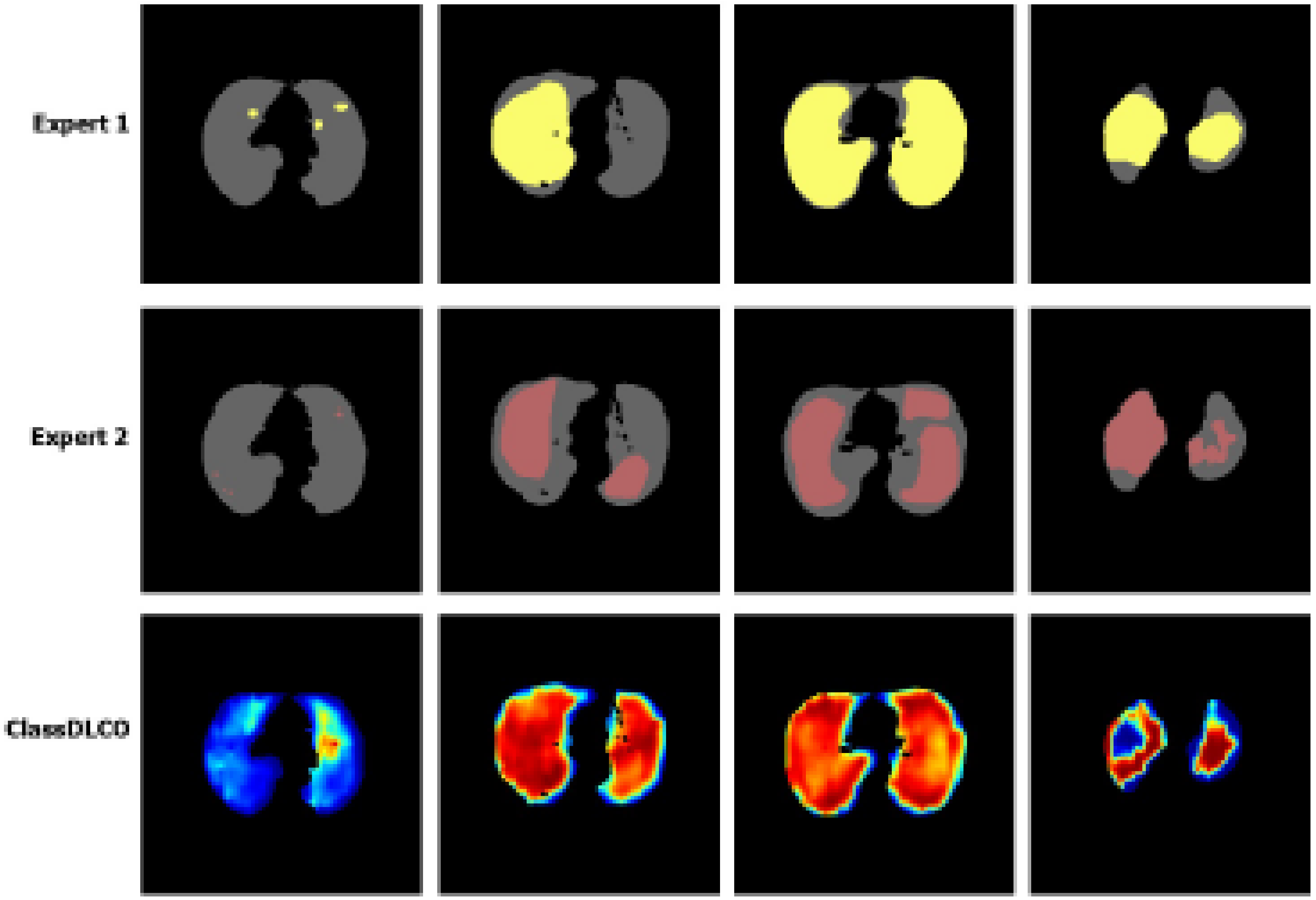
MIL

Images Marleen de Bruijne

- 600 chest CT images (train/validation/test)
- Instance-level vs bag-level classifiers
- Naïve approach already good
- Good to take all instances into account

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).



Work by Isabel Pino Pena

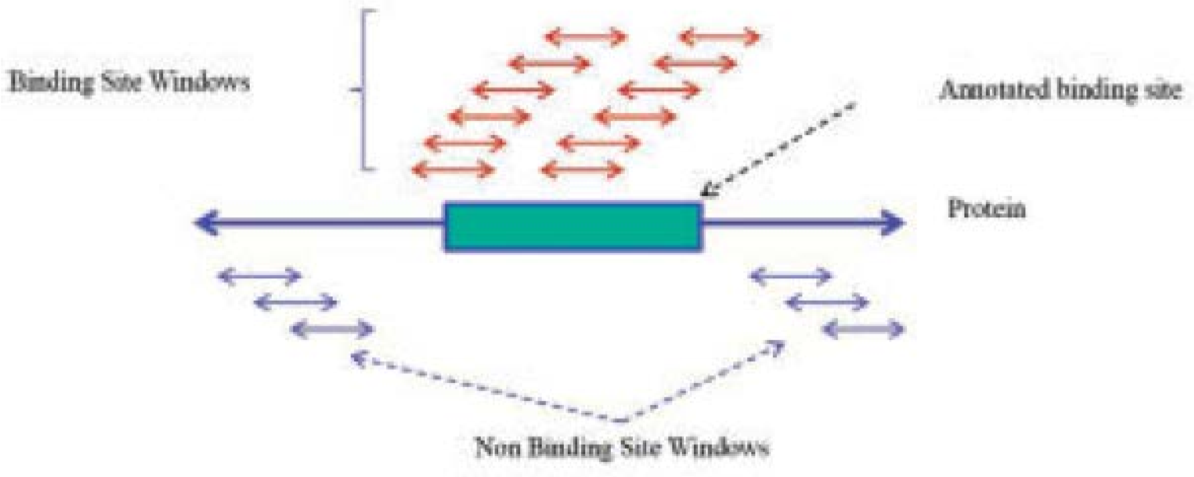
MIL in medical imaging

- Classification of cancerous tissue in histology most popular
- Brain, eye, lung, heart, breast and others
- About half classify instances

Reference	Application	MIL category	Method
	Brain		
Tong et al. (2014)	AD classification	global	excl bag
Chen et al. (2015b)	cerebral small vessel disease detection	global	instance
Dubost et al. (2017)	enlarged perivascular space detection	local	instance
	Eye		
Venkatesan et al. (2015)	diabetic retinopathy classification	global	excl bag
Quellec et al. (2012)	diabetic retinopathy classification	global, local	instance
Schlegl et al. (2015)	fluid segmentation	local	instance
Manivannan et al. (2016)	retinal nerve fiber layer visibility classification	global, local	instance
Lu et al. (2017)	fluid detection	global	instance
	Breast		
Maken et al. (2014)	breast cancer detection	global	multiple
Sanchez de la Rosa et al. (2015)	breast cancer detection	global, local	excl bag
Shin et al. (2017)	mass localization, classification	global, local	instance
	Lung		
Dundar et al. (2007)	pulmonary embolism detection	false positive	instance
Bi and Liang (2007)	pulmonary embolism detection	false positive	instance
Liang and Bi (2007)	pulmonary embolism detection	false positive	instance
Cheplygina et al. (2014)	COPD classification	global	multiple
Melendez et al. (2014)	tuberculosis detection	global, local	instance
Stainvas et al. (2014)	lung cancer lesion classification	false positive	instance
Melendez et al. (2016)	tuberculosis detection	global, local	instance
Kim and Hwang (2016)	tuberculosis detection	global, local	instance
Shen et al. (2016)	lung cancer malignancy prediction	global, local	instance
Cheplygina et al. (2017)	COPD classification	global	instance
Li et al. (2017b)	abnormality detection (14 classes)	global, local	instance
	Abdomen		
Dundar et al. (2007)	polyp detection	false positive	instance
Wu et al. (2009)	polyp detection	false positive	instance
Lu et al. (2011)	polyp detection, size estimation	false positive	instance
Wang et al. (2012)	polyp detection	false positive	instance
Wang et al. (2015a)	lesion detection	global	prim bag
Wang et al. (2015b)	lesion detection	global	prim bag
	Histology/Microscopy		
Dundar et al. (2010)	breast lesion detection	global	instance
Samsudin and Bradley (2010)	pap smear classification	global	multiple
McCann et al. (2012)	colitis detection	global	instance
Zhang et al. (2013)	skin biopsy annotation	global	multiple
Kandemir et al. (2014)	breast cancer detection	global	excl bag
Xu et al. (2014)	colon cancer detection	global, local	instance
Hou et al. (2015)	glioblastoma, low-grade glioma detection	global	instance
Li et al. (2015)	breast cancer detection	global	prim bag
Mercan et al. (2016)	breast cancer detection	global	instance
Kraus et al. (2016)	cell type classification	global, local	instance
Jia et al. (2017)	cancerous region segmentation (colon)	global, local	instance
Tomczak et al. (2017)	breast cancer detection	global	instance
	Multiple		
Vural et al. (2006)	abnormality detection in three applications	false positive	instance
Kandemir and Hamprecht (2015)	abnormality detection in two applications	global, local	multiple
Hwang and Kim (2016)	lesion detection in two applications	global, local	instance
	Other		
Situ et al. (2010)	dermoscopic feature annotation	global	prim bag
Liu et al. (2010)	cardiac event detection	global	instance
Yan et al. (2016)	bodypart recognition	global	instance

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

Bioinformatics



- Proteins that bind or not to Calmodulin
- Instances = sequences of length 21, substring co-occurrence features
- MI SVM
- “*The weights for the 1-spectrum features closely follow the amino acid propensities in CaM binding sites*”

Minhas, F. U. A. A., & Ben-Hur, A. (2012). Multiple instance learning of Calmodulin binding sites. *Bioinformatics*, *28*(18), i416-i422.

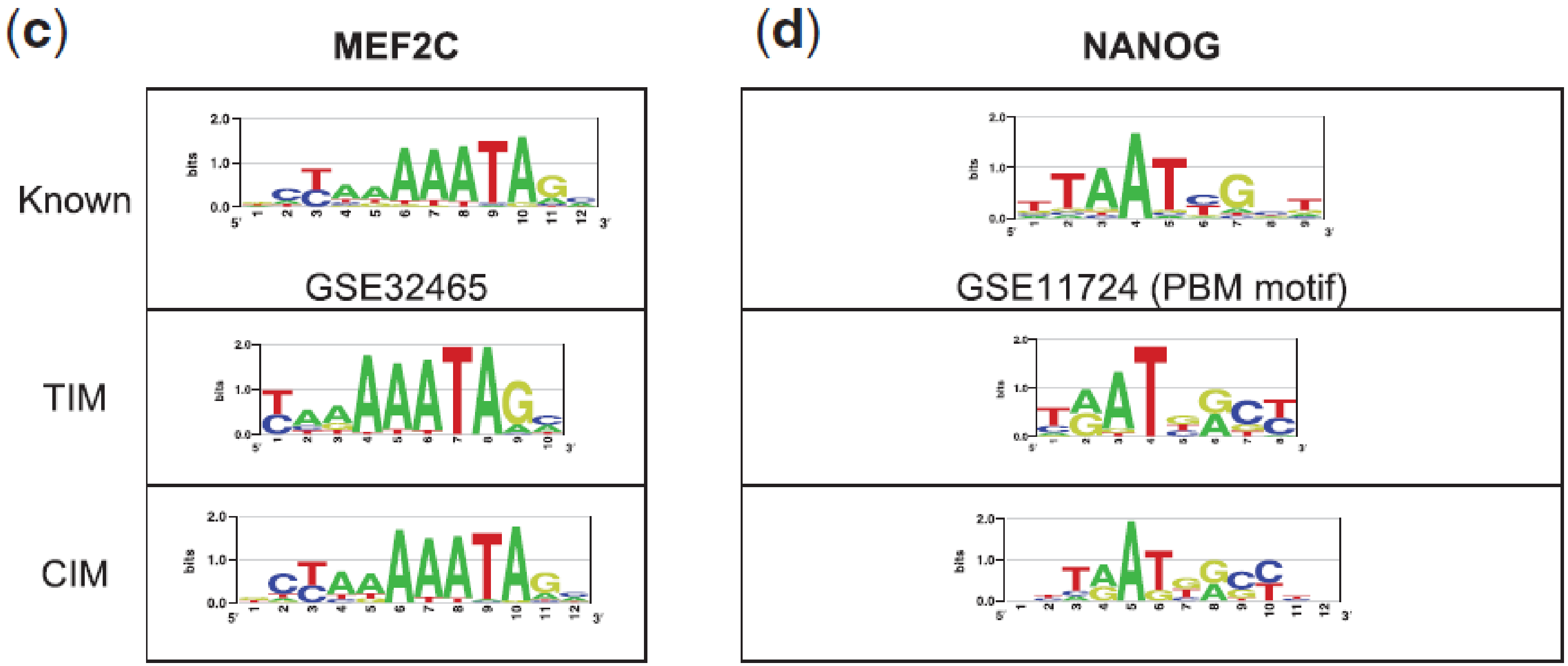
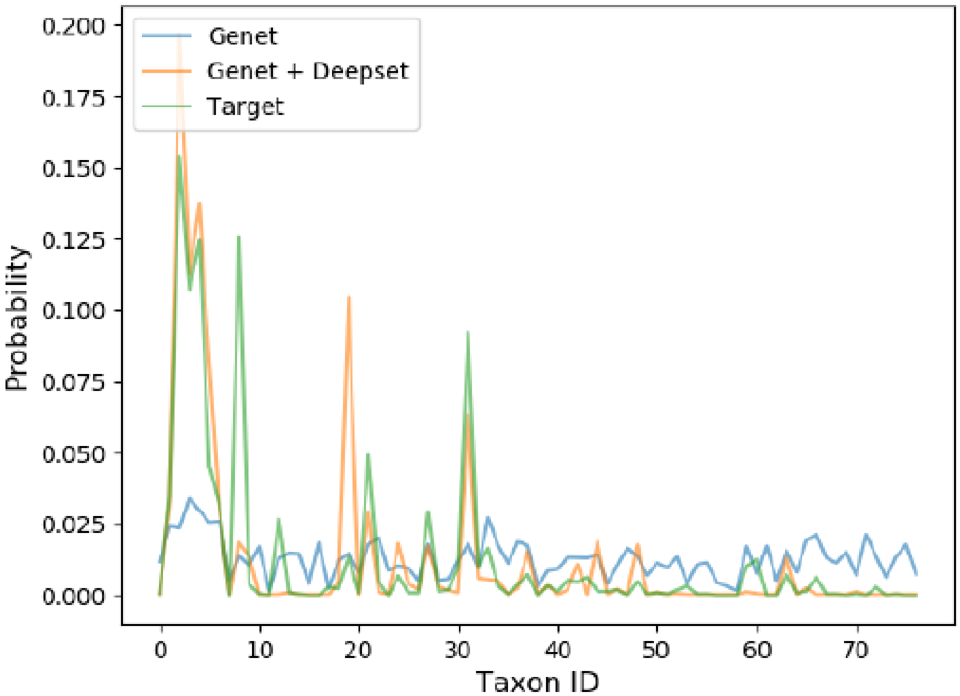
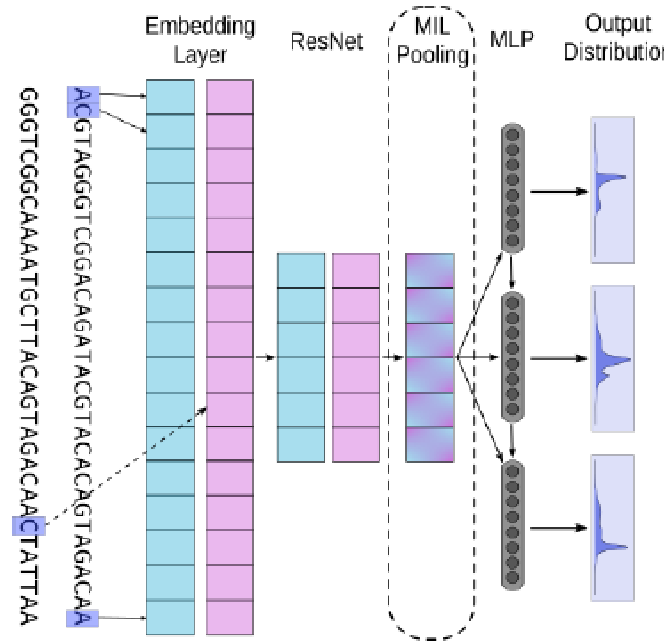


Fig. 5. Comparison between literature motifs and best match top instance motifs ('TIM') and best match center instance motifs ('CIM'). Motifs were identified by HOMER which was separately run on top (TI) and central (CI) sequences

- Bag = DNA sequence, binds or not to a TF
- Instances = all subsequences, represented by co-occurrence features
- SimpleMIL, mean combining

Gao, Z., & Ruan, J. (2017). Computational modeling of in vivo and in vitro protein-DNA interactions by multiple instance learning. *Bioinformatics*, *33*(14), 2097-2105.





(a) GeNet + MIL pooling

- Predict distribution of taxa in a metagenomics DNA read (bag)

Georgiou, A., Fortuin, V., Mustafa, H., & Rätsch, G. (2019). Deep Multiple Instance Learning for Taxonomic Classification of Metagenomic read sets. *arXiv preprint arXiv:1909.13146*.

Similarities & Differences

- Representation
  - Instance = often “part of” bag (not “is a”)
  - “Distribution” problems more common
  - Instance correlations?

- Label type
  - Class label, proportion in medical imaging
  - Label, ranking, distribution

Similarities & Differences

- Methods
  - Existing vs “novel” methods
  - Simple methods are competitive

- Instance labels
  - Source of labels
  - Use of instance labels for training

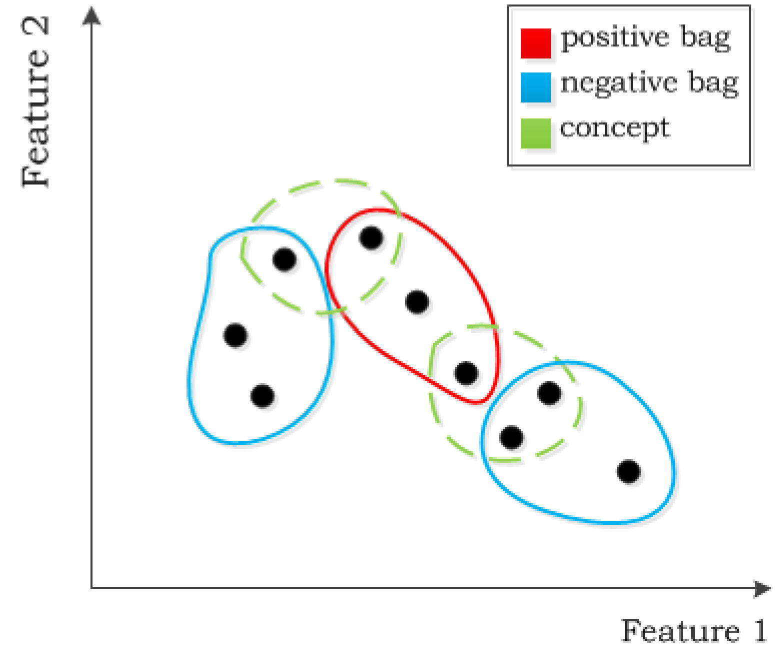
- Open science

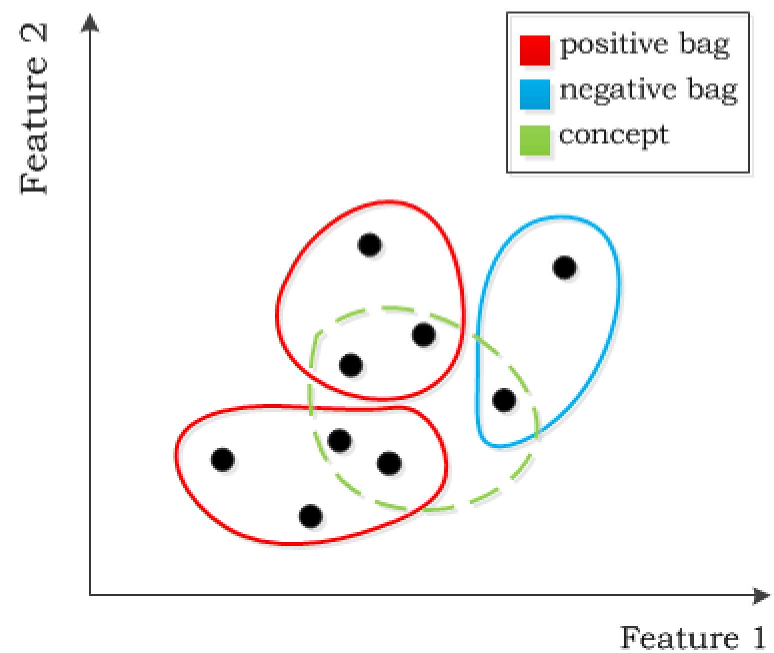
Challenges

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

Many variants

- Relationship bag/instances, “is a” or “is part of”?
- What are the assumptions about the labels?
- What is the goal of the classifier?





Multiple Instance Learning

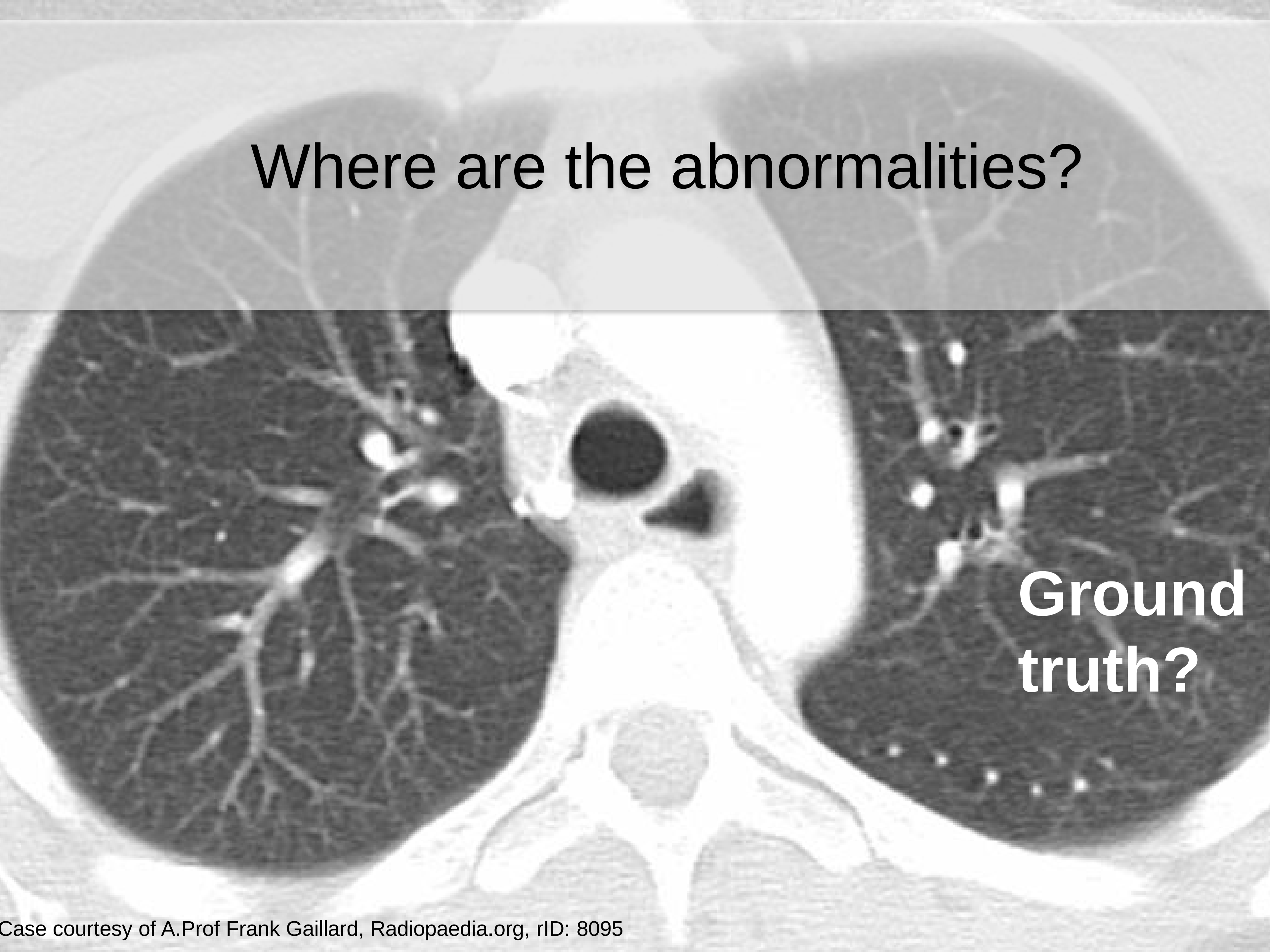
Group-based Classification

Set Classification

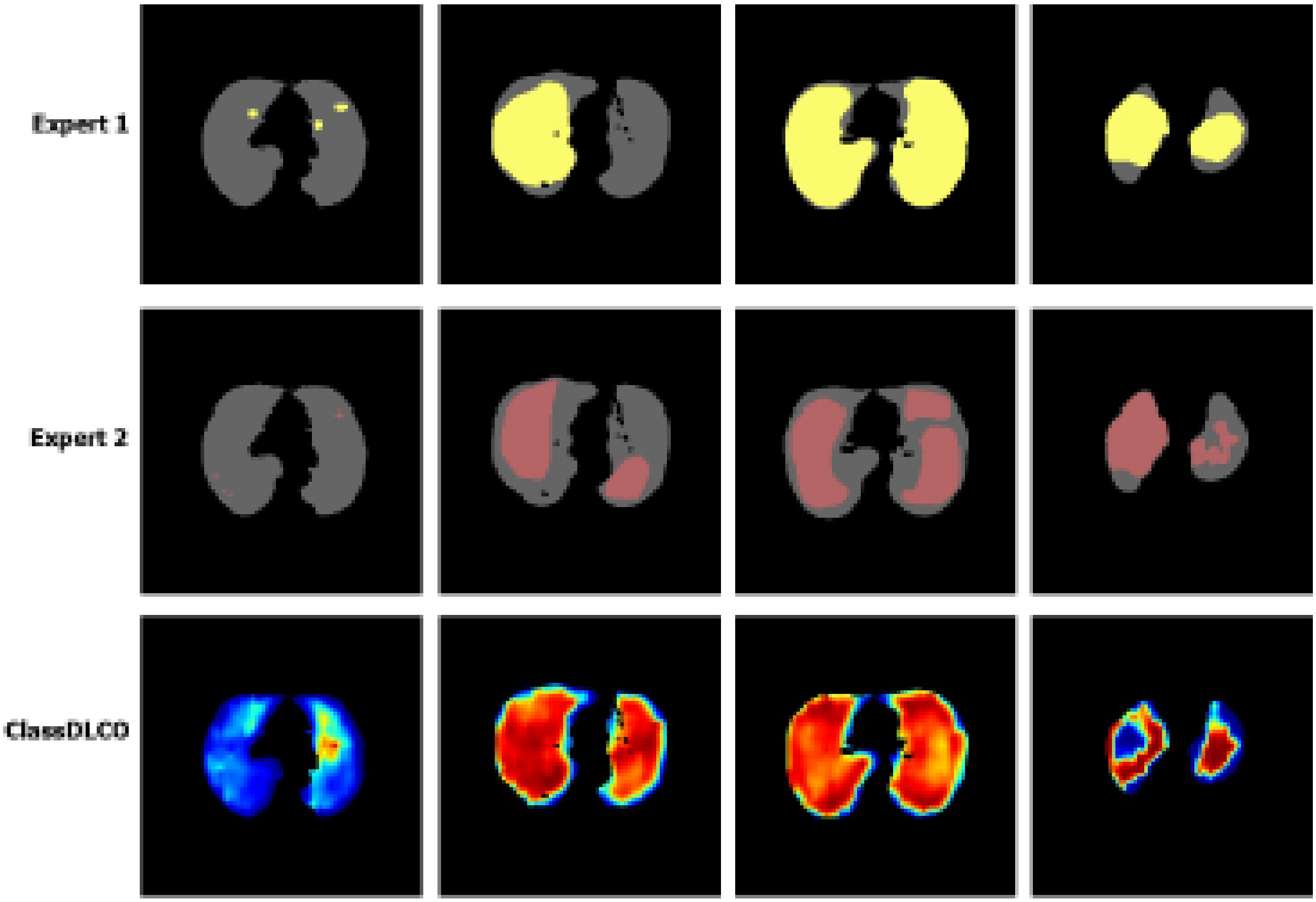
Collective Classification

Cheplygina, V., Tax, D. M. J., & Loog, M. (2015). On classification with bags, groups and sets. *Pattern Recognition Letters*, *59*, 11-17.

Where are the abnormalities?



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

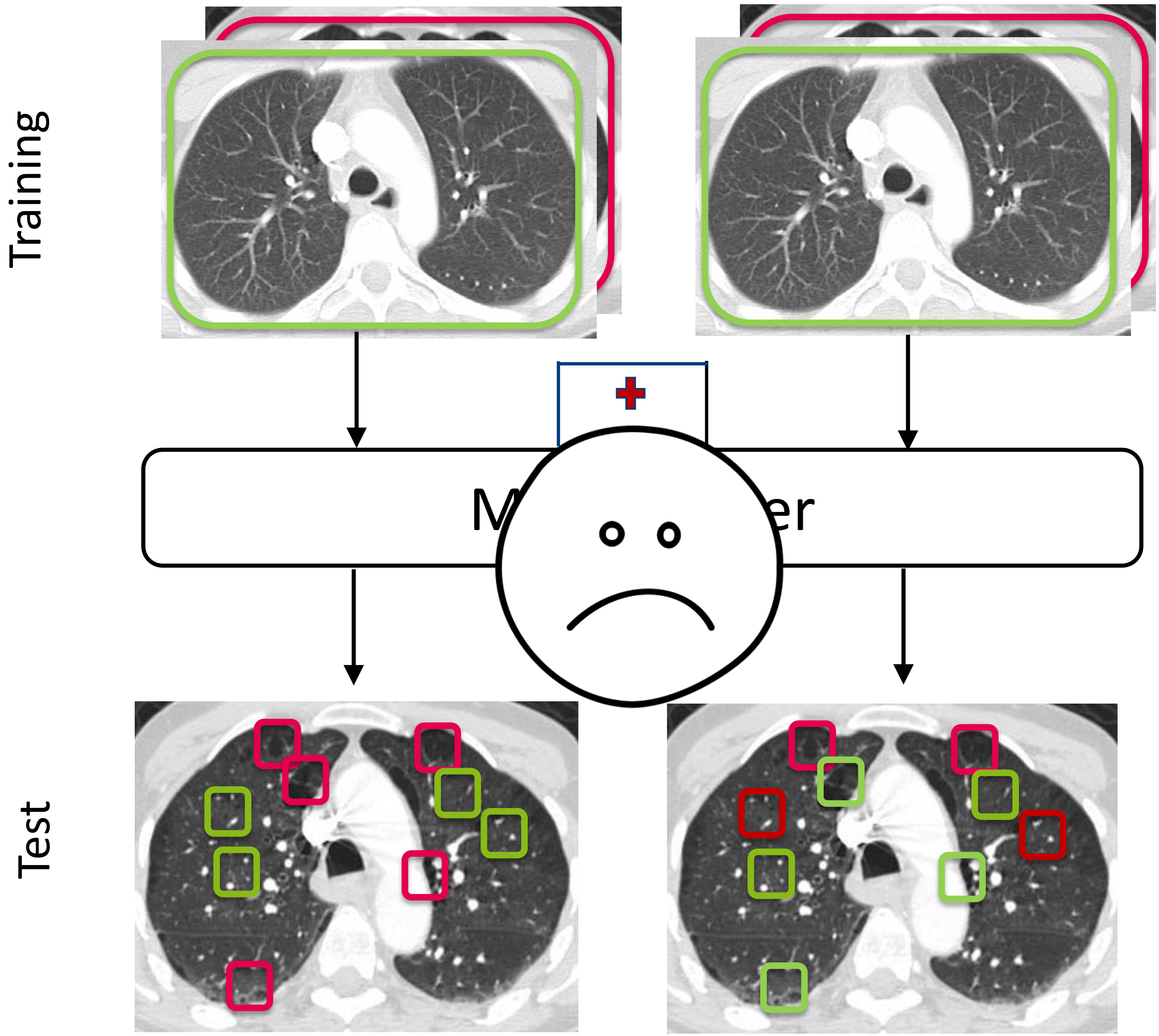


Work by Isabel Pino Pena

Instance labels often not available

Task	B	I	Task	B	I
Cancer histology [9]	+	+	Diabetic retinopathy [9]	+	−
COPD in CT [6]	+	−	Tuberculosis in XR [13]	+	∘
Cancer histopathology [22]	+	+	Osteoarthritis in MRI [12]	+	∘
Diabetic retinopathy [16]	+	+	Pulmonary embolism in CT [11]	+	−
Myocardial infarction in ECG [18]	+	−	COPD in CT [17]	+	−
Colorectal cancer in CT [8]	+	−			

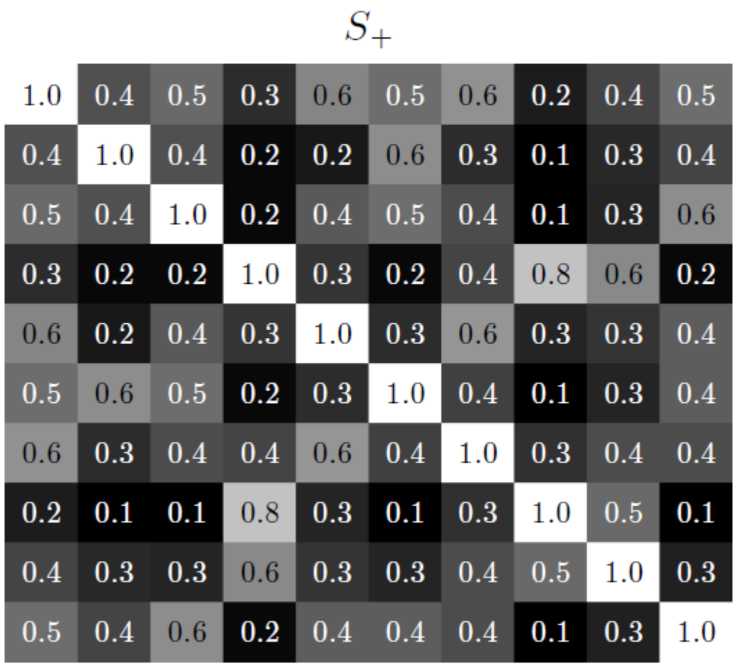
Table 1. Evaluation of MIL in CAD tasks. Columns show bag (B) and instance (I) evaluation: supervised (+), qualitative (∘) or none (−).


Training
Test

Instance labels can be unstable

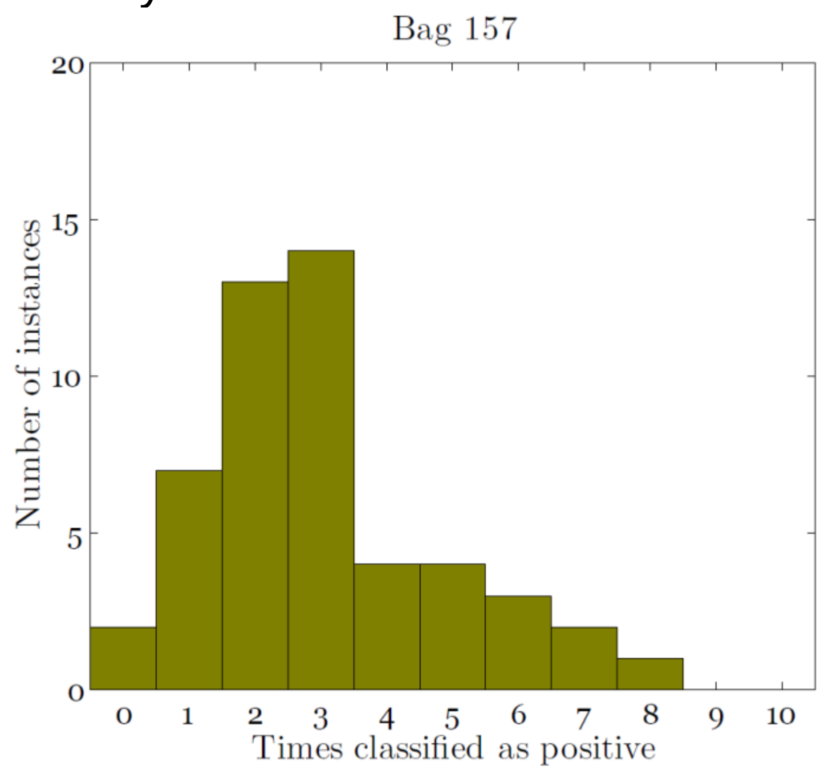
Fraction of agreement

Ideally: always 1.0



How often patches are labeled abnormal

Ideally: either 0x or 10x

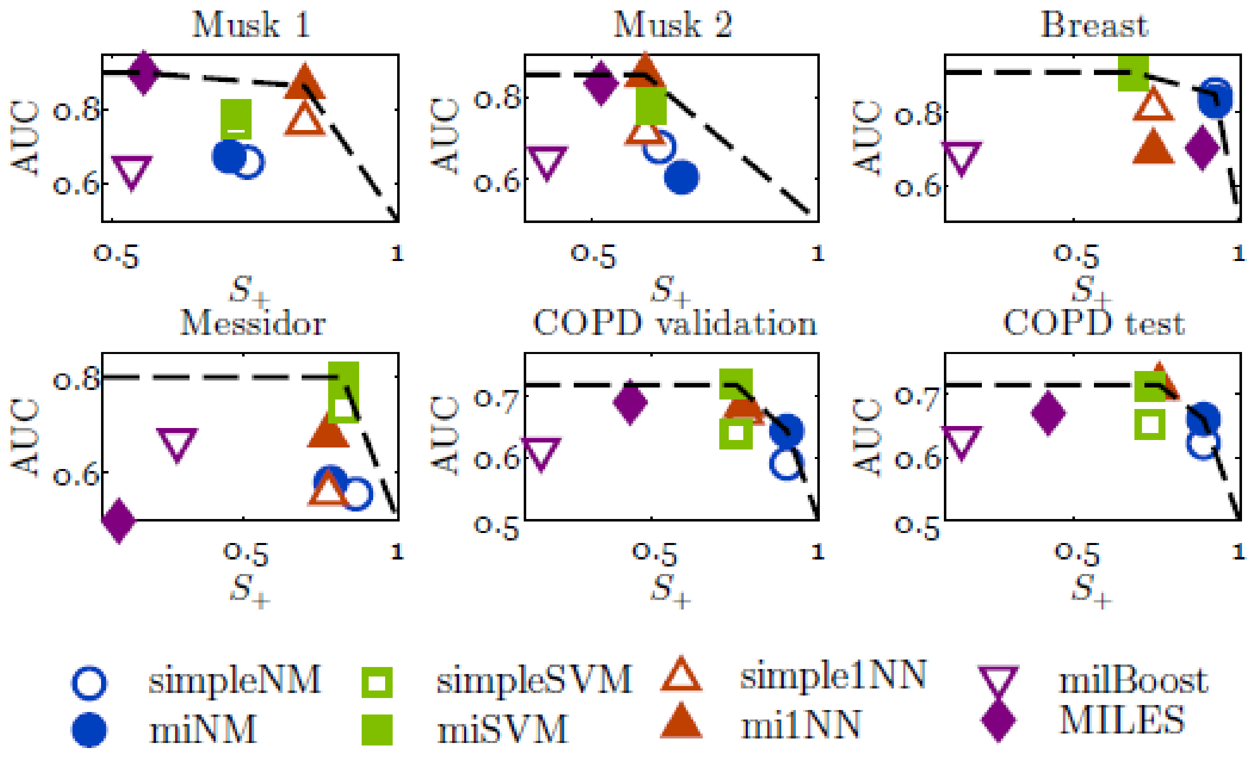


Cheplygina, V., Sørensen, L., Tax, D. M. J., de Bruijne, M., & Loog, M. (2015) Label Stability in Multiple Instance Learning. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI),* pp. 539-546

Similar situation in other datasets

6 datasets, 8 MIL classifiers

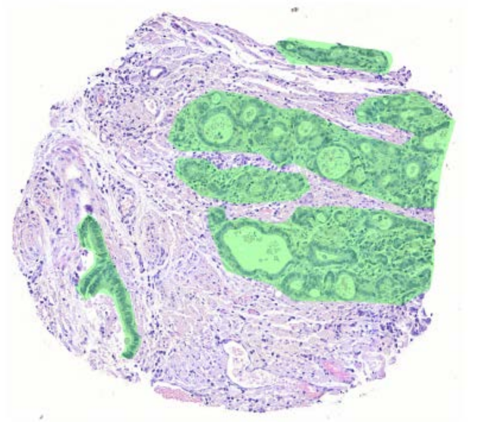
Trade-off bag performance and instance stability



Bag vs instance performance

Bag-level

	Accuracy (%)	F1 score	AUC-ROC	AUC-PR
mi-Graph [33]	**86.4**	**0.90**	**0.93**	**0.97**
MILBoost [24]	83.0	0.88	0.91	0.96
B-KI-SVM [17]	82.6	0.88	0.91	0.95
GPMIL [15]	81.2	0.88	0.90	0.93
I-KI-SVM [17]	80.3	0.86	0.89	0.93
iAPR [7]	79.4	0.87	0.88	0.94
Citation k-NN [26]	74.5	0.83	0.72	0.82
EMDD [31]	72.2	0.83	0.72	0.82
mi-SVM [2]	68.4	0.81	0.86	0.76
MI-SVM [2]	68.1	0.81	0.89	0.94
SIL-SVM [3]	68.1	0.81	0.92	0.95
Fully Supervised SVM	85.0	0.90	0.92	0.96

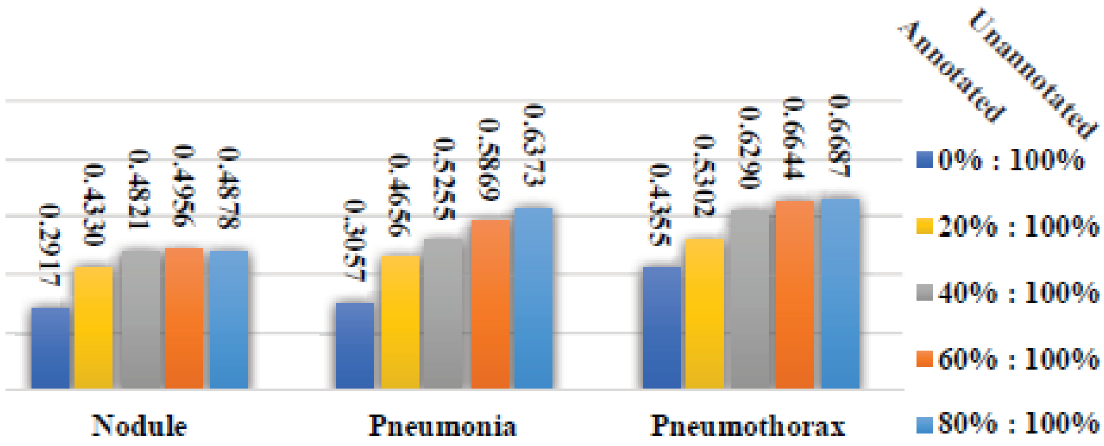


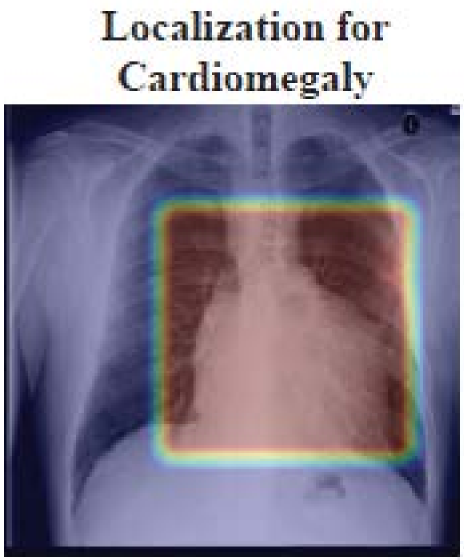
Instance-level

	Accuracy (%)	F1 score	AUC-ROC	AUC-PR
MILBoost [24]	**66.7**	0.70	0.75	0.71
GPMIL [15]	65.8	0.54	0.77	0.69
B-KI-SVM [17]	64.7	0.48	0.67	0.67
I-KI-SVM [17]	63.0	0.37	0.69	0.68
mi-SVM [2]	62.7	**0.71**	**0.84**	**0.82**
iAPR [7]	57.8	0.34	0.50	0.47
Citation k-NN [26]	54.3	0.67	0.69	0.76
EMDD [31]	54.1	0.33	0.56	0.52
MI-SVM [2]	46.9	0.64	0.74	0.71
SIL-SVM [3]	46.9	0.64	0.80	0.75
mi-Graph [33]	-	-	-	-
Fully Supervised SVM	83.5	0.82	0.91	0.90

Kandemir, M., & Hamprecht, F. A. (2015). Computer-aided diagnosis from weak supervision: a benchmarking study. Computerized Medical Imaging and Graphics, 42, 44-50.

Bag vs instance performance



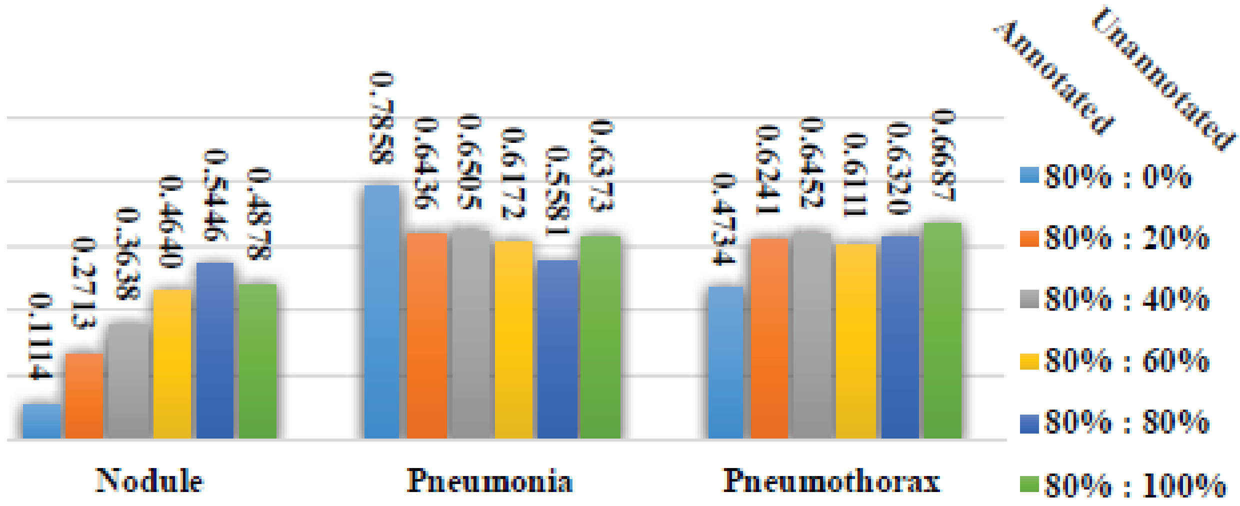


Training on 100% bags and adding 0-80% instances
Testing on 20% instances

Adding instances helps!

Li, Z., Wang, C., Han, M., Xue, Y., Wei, W., Li, L. J., & Li, F. F. (2017). Thoracic Disease Identification and Localization with Limited Supervision. arXiv preprint arXiv:1711.06373.

Bag vs instance performance



Training on 80% instances and 0-100% bags
Testing on 20% instances

Adding bags does **can even hurt performance**

Li, Z., Wang, C., Han, M., Xue, Y., Wei, W., Li, L. J., & Li, F. F. (2017). Thoracic Disease Identification and Localization with Limited Supervision. arXiv preprint arXiv:1711.06373.

Summary

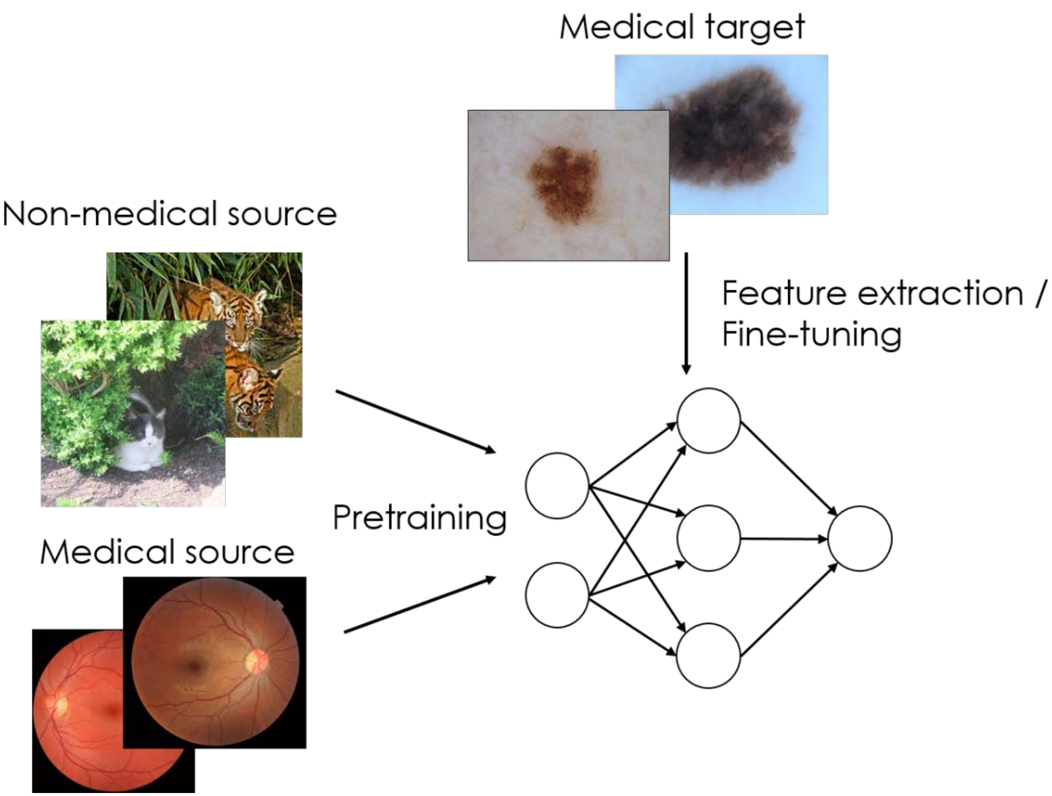
- Multiple instance learning
  - Definition
  - Classifiers

- Examples
  - Medical imaging
  - Bioinfomatics

- Challenges
  - Definitions, what is MIL?
  - Instance label evaluation
  - Bag/instance trade-off


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

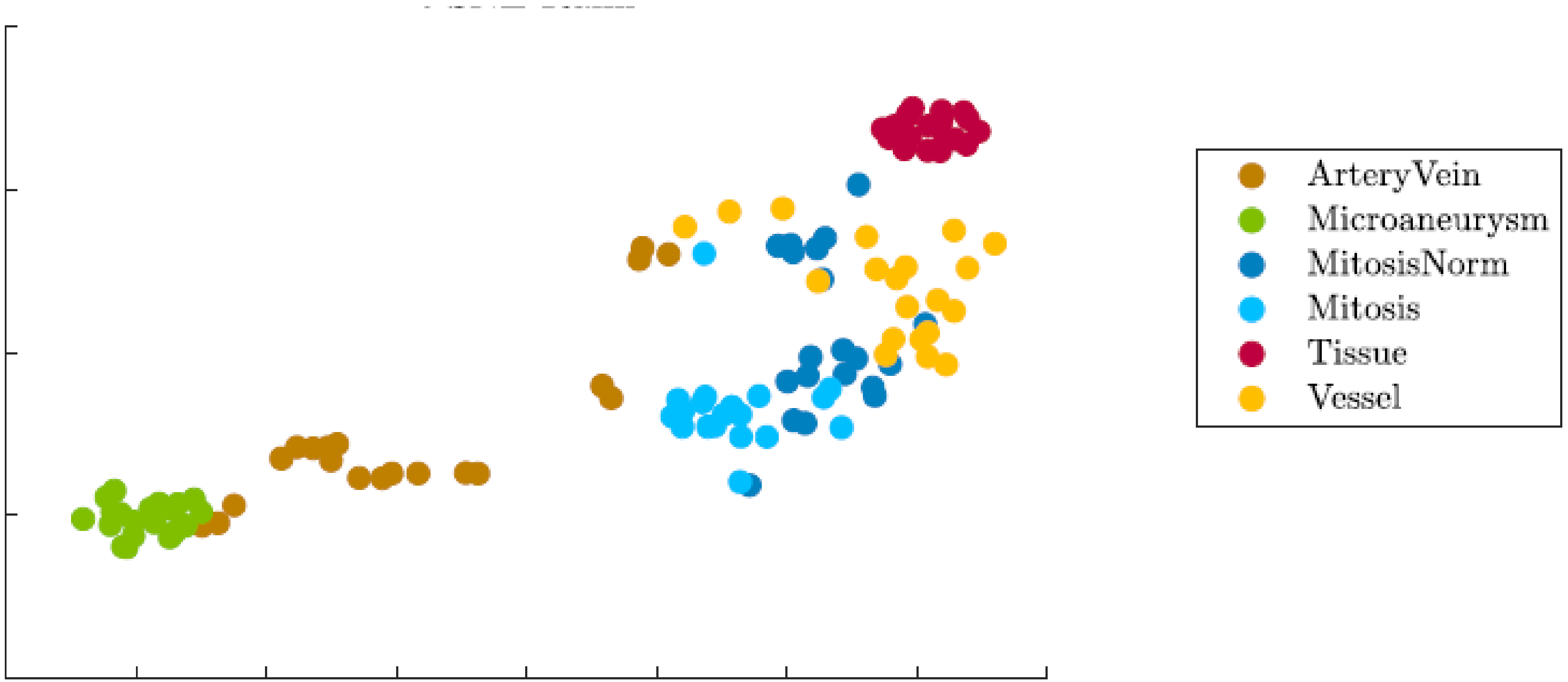
Transfer learning from (dis)similar datasets



Cats or CAT scans: transfer learning from natural or medical image source datasets?

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

Meta-learning: how to quantify similarity of data?

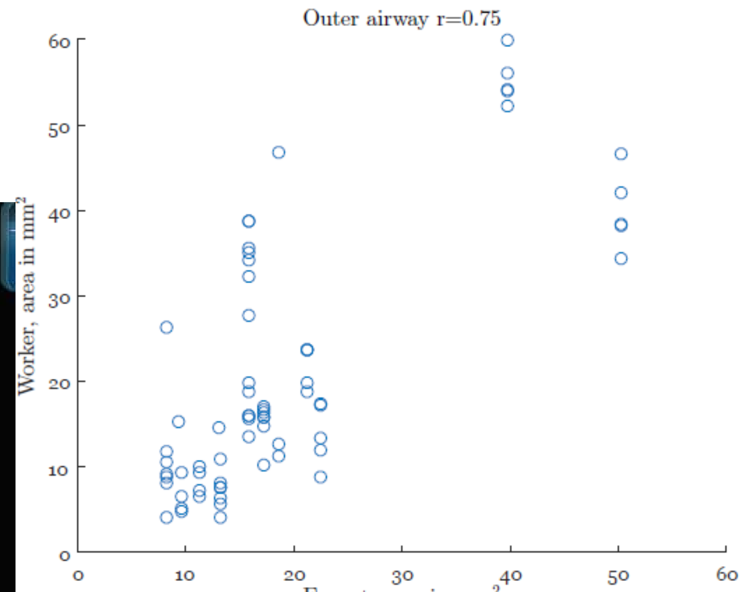


Cheplygina, V., Moeskops, P., Veta, M., Bozorg, B. D., & Pluim, J. (2017). Exploring the similarity of medical imaging classification problems. In Large-Scale Annotation of Biomedical Data and Expert Label Synthesis (MICCAI LABELS) (pp. 59-66)

A Survey of Crowdsourcing in Medical Image Analysis

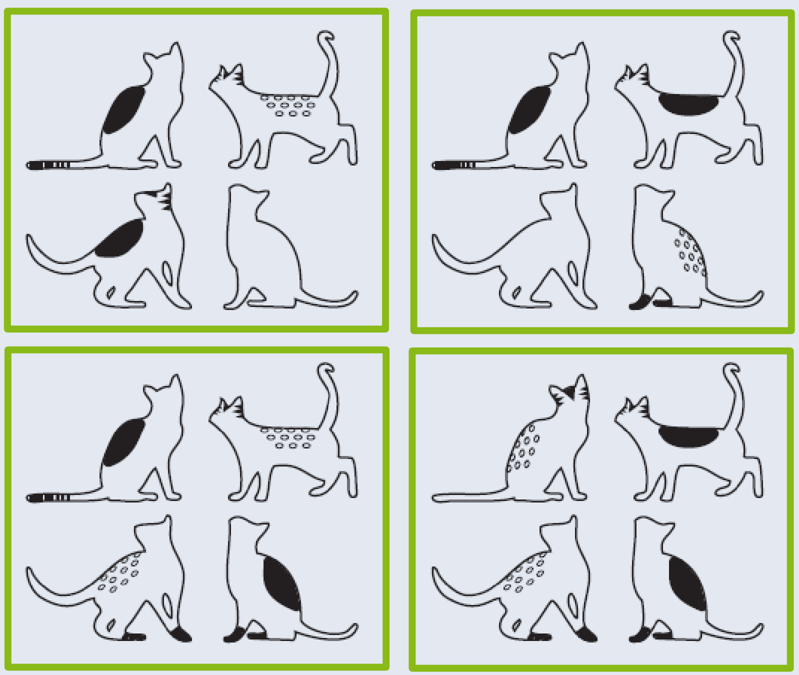
Silas Ørting[1]✉, Andrew Doyle[2,*], Matthias Hirth[3,*] Arno van Hilten[4,*], Oana Inel[5,7,*], Christopher R. Madan[6,*], Panagiotis Mavridis[7,*], Helen Spiers[8,9,*], and Veronika Cheplygina[10]✉

[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



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.



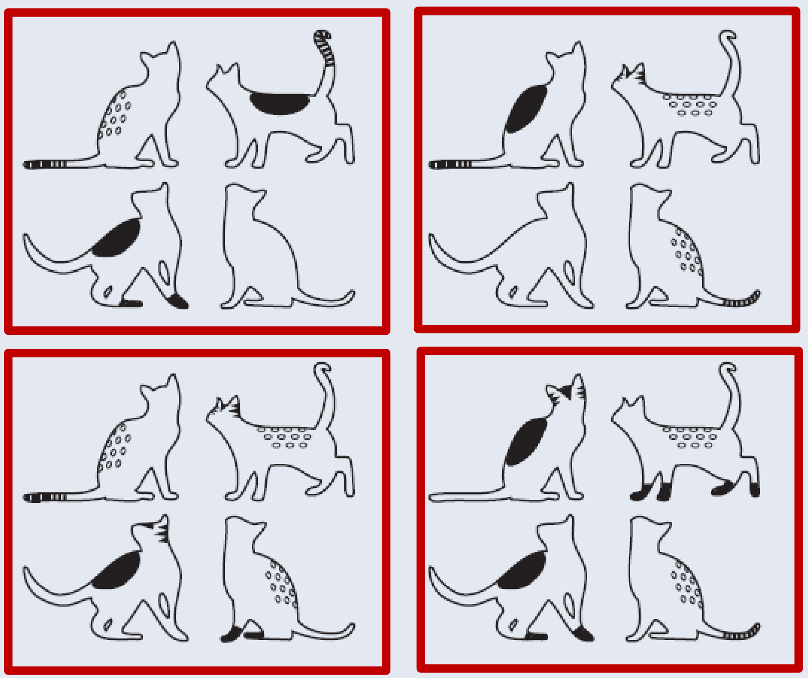
ISBN 978-94-6295-192-1

Dissimilarity-Based Multiple Instance Learning

Veronika Cheplygina

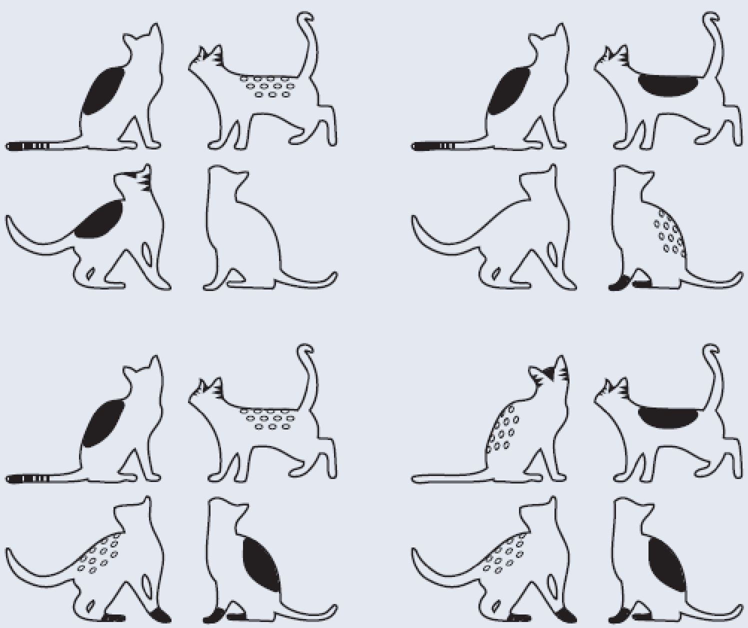
Dissimilarity-Based Multiple Instance Learning

Veronika Cheplygina



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