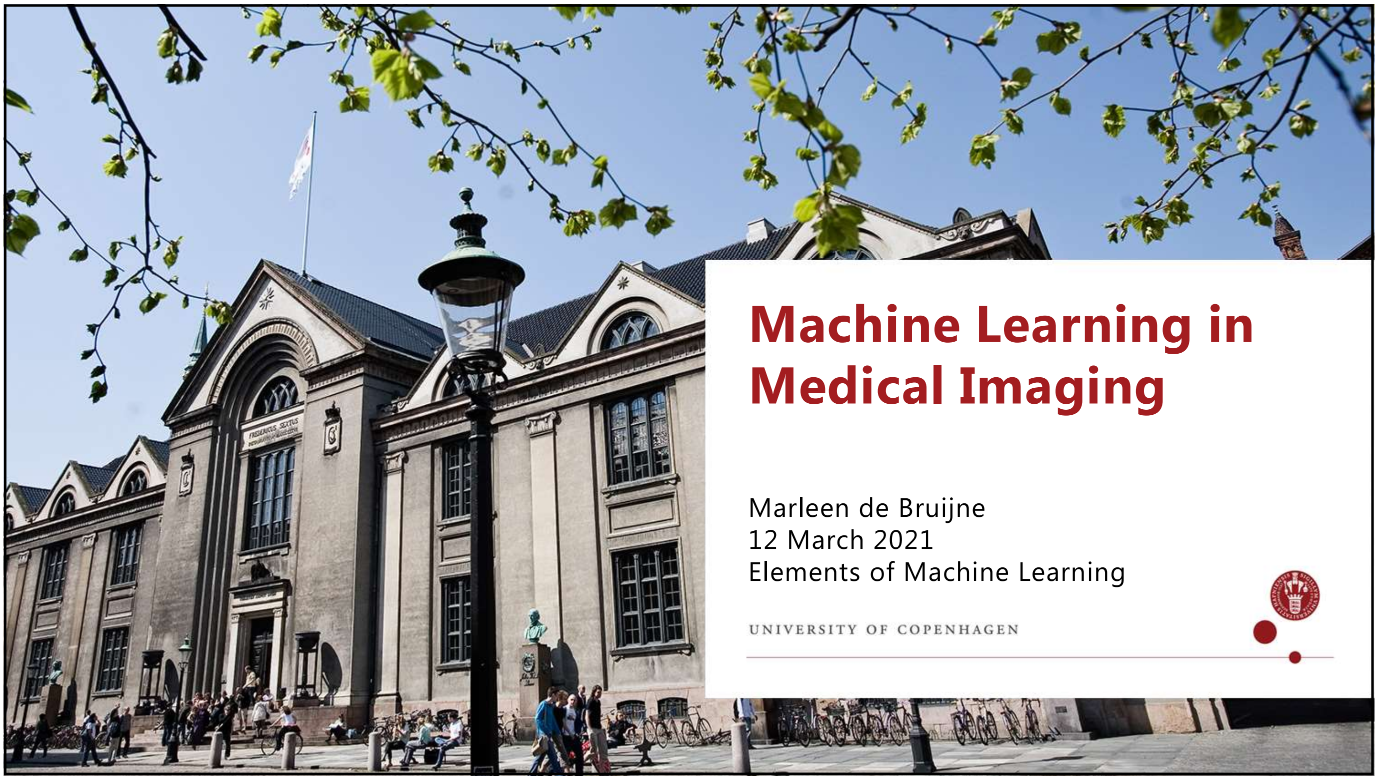


Machine Learning in Medical Imaging

Marleen de Bruijne

12 March 2021
Elements of Machine Learning

UNIVERSITY OF COPENHAGEN

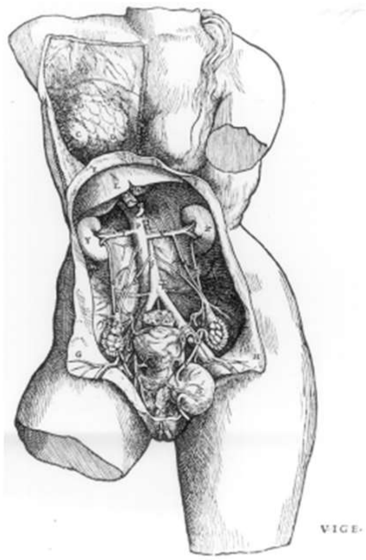
1

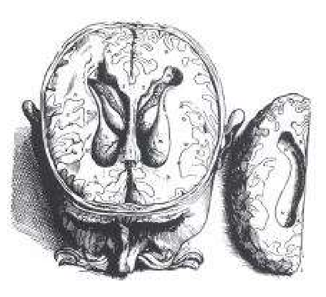
This lecture

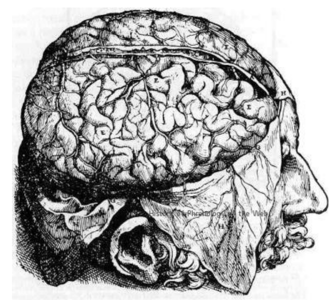
- A Brief History of Medical Imaging
- Why (automated) medical image analysis?
- Basic principle of machine learning as used in medical imaging
- Examples of own (and other group's) research
- Where are we in clinical practice?
- Open questions

2

Early medical imaging







3

Non-invasive imaging

The first X-ray image:
Mrs. Röntgen's hand (1895)

One of the first clinical examples: Shot of hail in hand (1896)

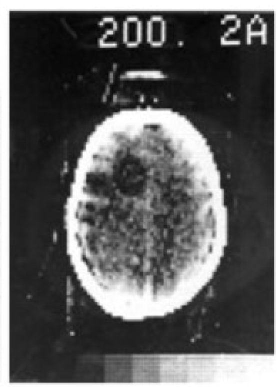
4

The first successful brain scan (ex vivo)

- 1968
- 80x80 pixels
- Predecessor of the Computed Tomography (CT) scan

5

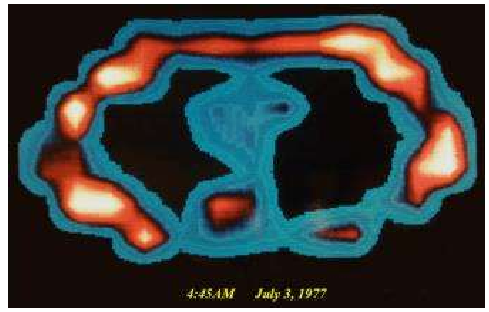
1st Clinical CT scan, 1971



- 3-dimensional! 80x80x9 pixels

6

First MRI scan: 1977

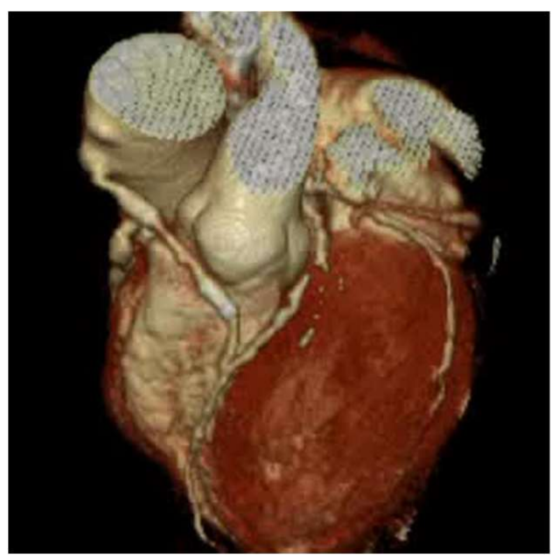


This makes it possible to distinguish between soft tissues

7

Modern medical imaging: We see hearts beat

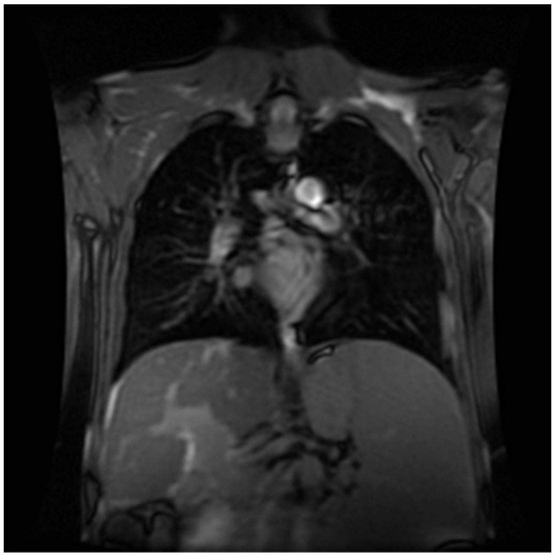
Computed tomography (CT)



Courtesy Nico Mollet, Radiology, Erasmus MC

8

We see lungs breathe

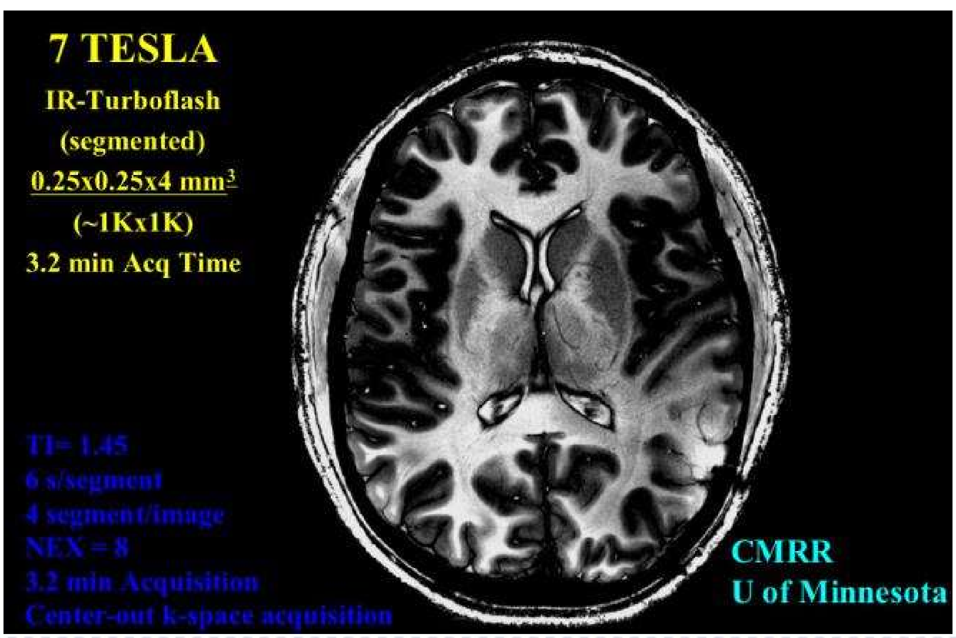


MRI

Courtesy Harm Tiddens, Erasmus MC-Sophia Children's hospital

9

The newest MRI scanners yield very high resolution and soft-tissue contrast



10

Big data

Thousands of scans each day

In Denmark alone:

991,683 CT exams per year

498,708 MRI exams per year*

US, X-ray even more frequent

Ideal application area for machine learning?

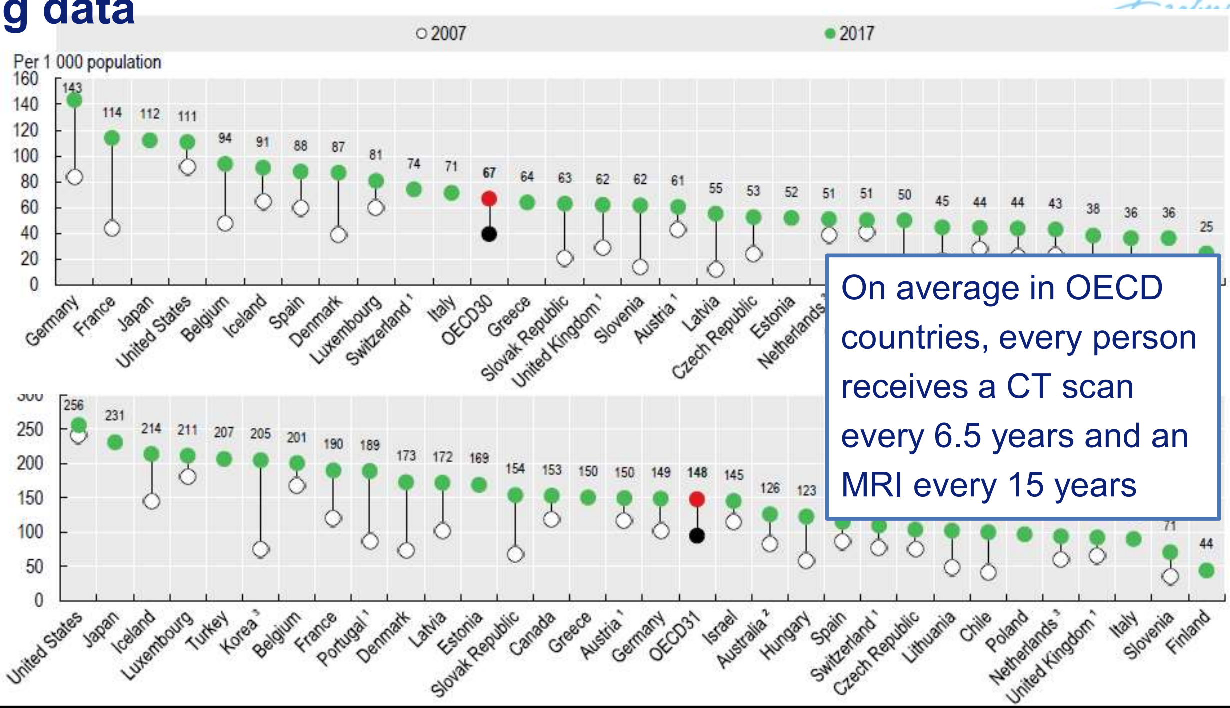
**2017 numbers, Health at a Glance 2019*

11

Big data

MRI

CT



On average in OECD countries, every person receives a CT scan every 6.5 years and an MRI every 15 years

12

What do we do with all those images?

Erasmus MC

13

Erasmus MC

Humans (and especially radiologists) are extremely good at interpreting images

BUT…

14

Quantification is more difficult

Erasmus MC

Humans are not very good at judging color (or intensity values)

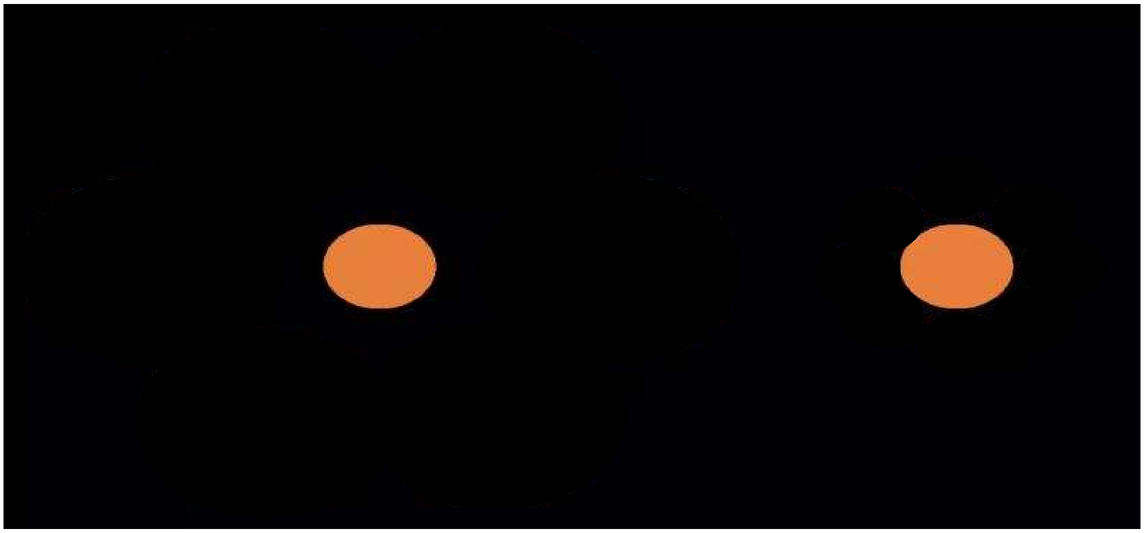
15

Quantification is more difficult

Humans are not very good at judging intensity values

16

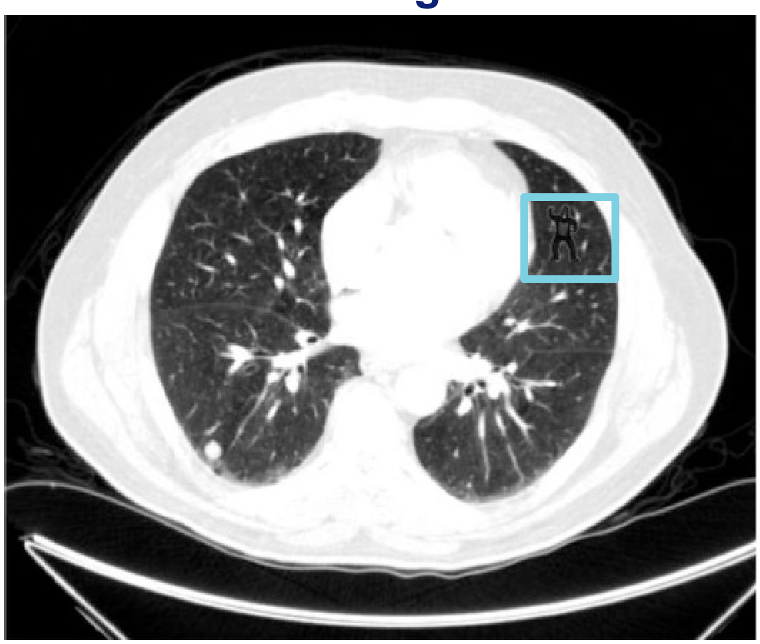
Quantification is more difficult



Humans are not very good at judging dimensions

17

Information overload:
Each scan has hundreds of images to assess



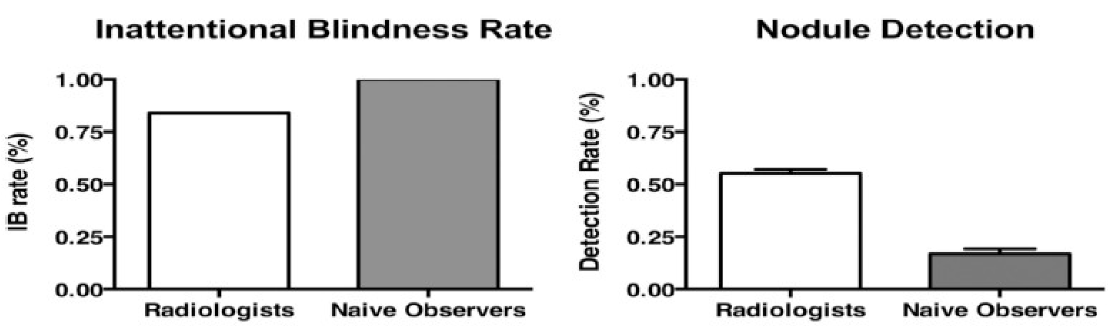
18

Erasmus MC

19

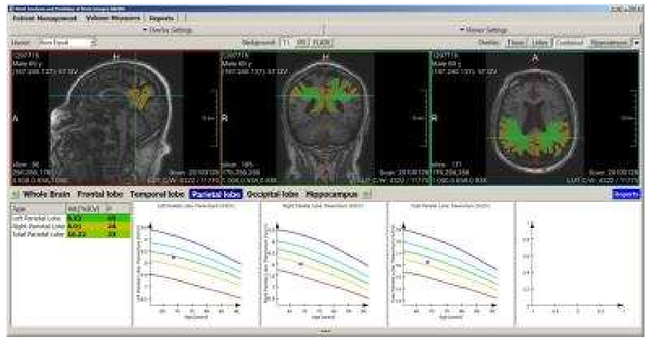
Erasmus MC

83% of radiologists did not see the gorilla!



20

From visual assessment to quantitative analysis aided by computers



- Provide numbers (imaging biomarkers) to help physicians make better decisions
- Monitor subtle changes: disease progression, treatment effect
- Can we use this data even better?

21

Medical Image Analysis: The classical approach

→

Image processing
Segmentation

→

Detection and volume measures e.g.:

- Tumor
- Brain tissues
- MS lesions
- Calcifications
- Brain structures
- Vessel lumen
- …

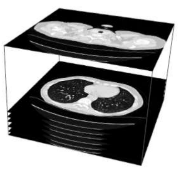
→

- Diagnosis
- Risk profile
- Treatment plan
- Prognosis

22

Machine Learning in Medical Image Analysis

Erasmus MC



→

~~Image processing~~
~~Segmentation~~

Machine learning

→

Detection and volume measures e.g.:
- Tumor
- Brain tissues
- MS lesions
- Calcifications
- Brain structures
- Vessel lumen
- …

→

- Diagnosis
- Risk profile
- Treatment plan
- Prognosis

23

Machine Learning in Medical Image Analysis
Learning Imaging Biomarkers

Erasmus MC

→

~~Image processing~~
~~Segmentation~~

Machine learning

→

~~Detection and volume measures e.g.:~~
- ~~Tumor~~
- ~~Brain tissues~~
- ~~MS lesions~~
- ~~Calcifications~~
- ~~Brain structures~~
- ~~Vessel lumen~~
- ~~…~~

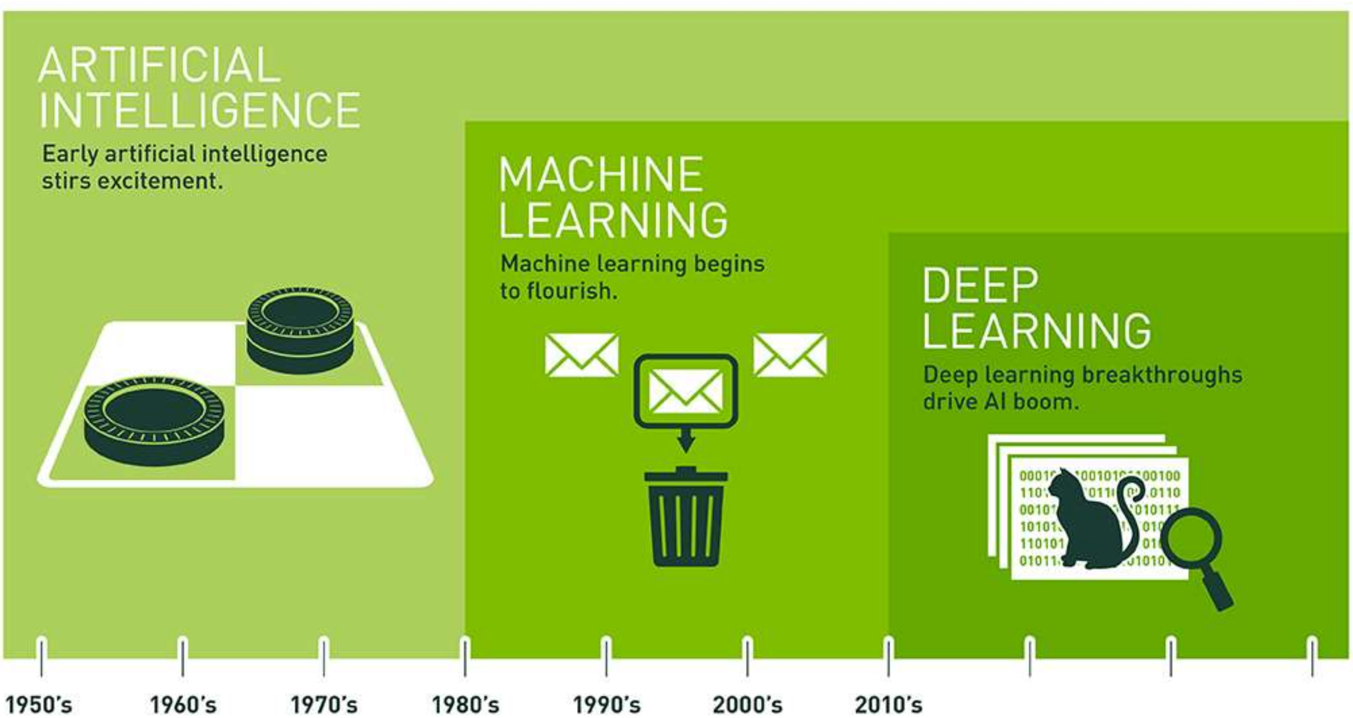
→

- Diagnosis
- Risk profile
- Treatment plan
- Prognosis

24

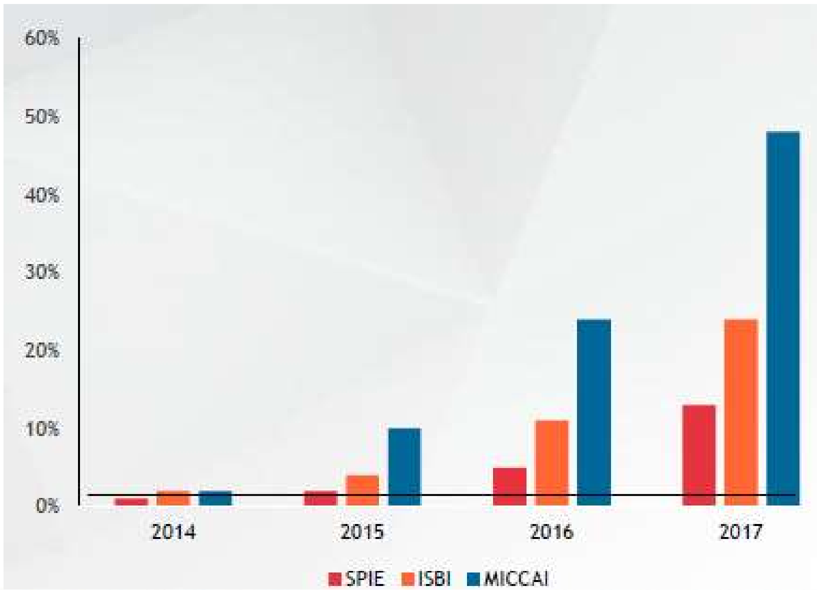
MACHINE LEARNING IN RADIOLOGY: THE PRINCIPLE

25



26

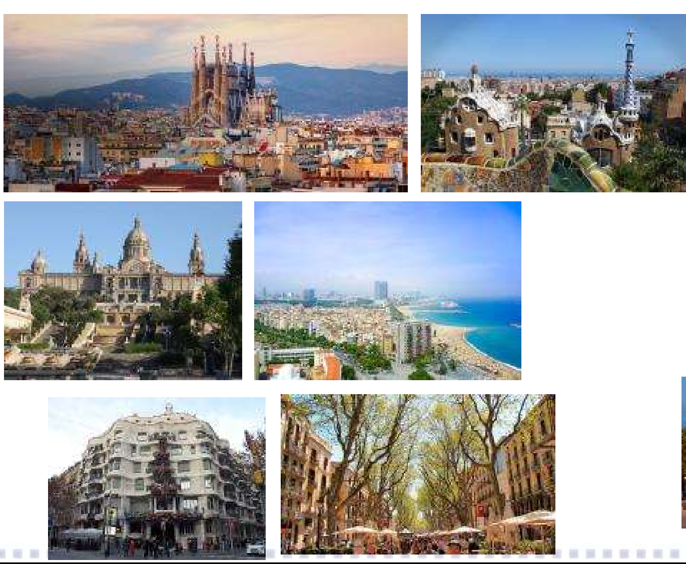
Deep learning is now the main component in most new medical image analysis techniques

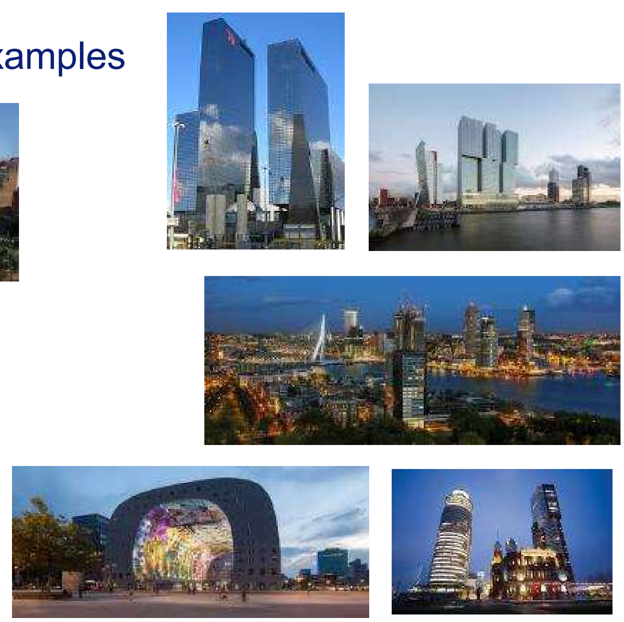


27

What is machine learning?

Letting the computer learn based on examples

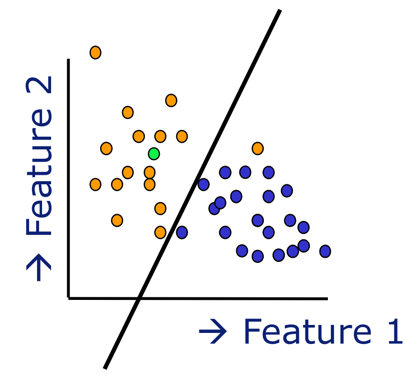




28

What is Machine learning?

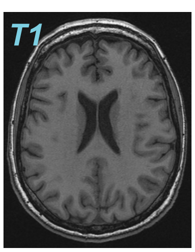
- Letting the computer learn based on examples
- The classical approach:

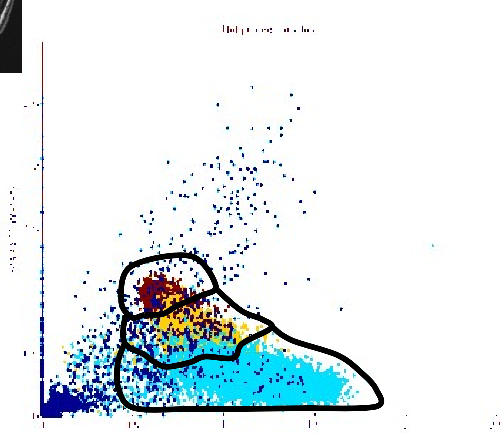


Input (image) ⇨ *Compute descriptive features* ⇨ *Train a classifier* ⇨ *Output*

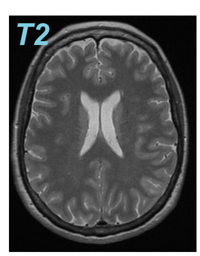
29

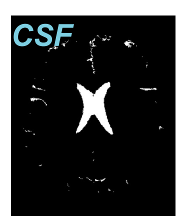
A simple example: 2 features





BG **WM** **GM** **CSF**





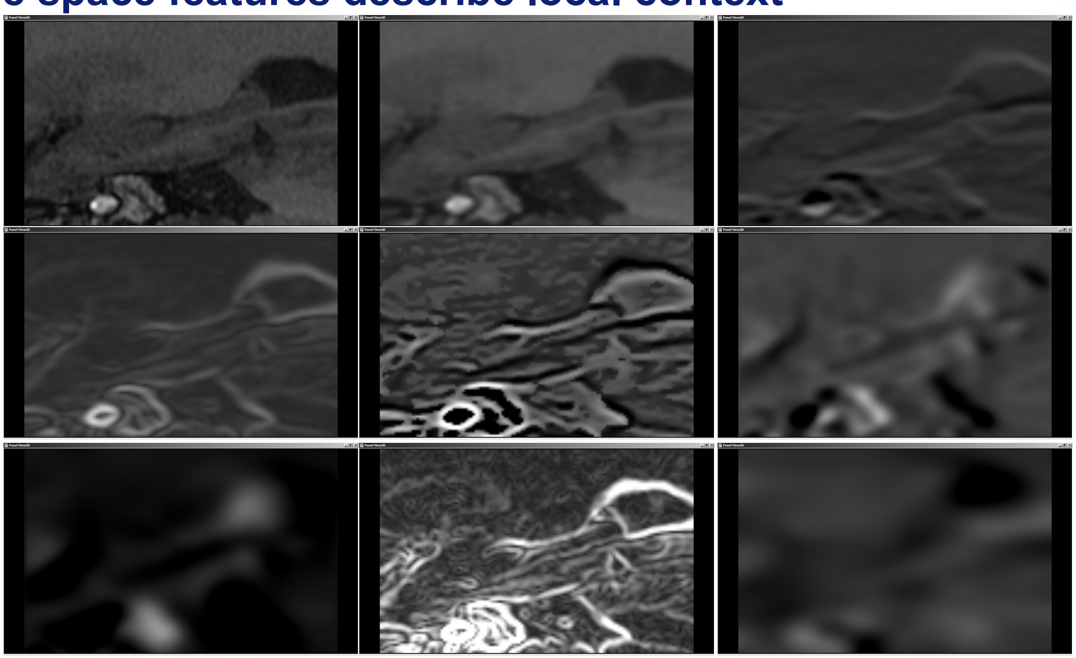
30

Example: Hippocampus segmentation
Intensity is not enough

Image *Manual segmentation* *Intensity threshold*

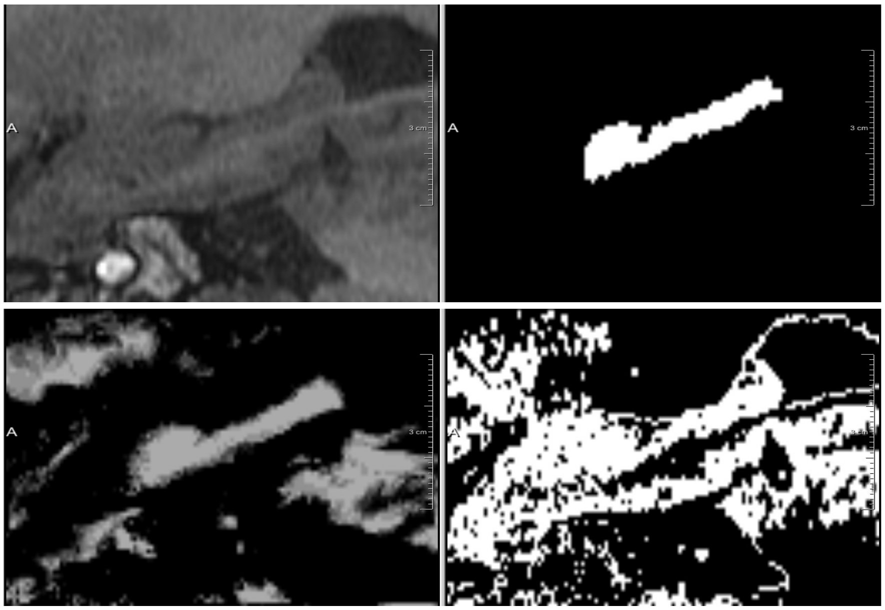
31

Beyond intensity:
Scale space features describe local context



32

Example: Hippocampus segmentation



Van der Lijn et al, TMI 2012

33

Deep learning: what is new?

The classical picture:

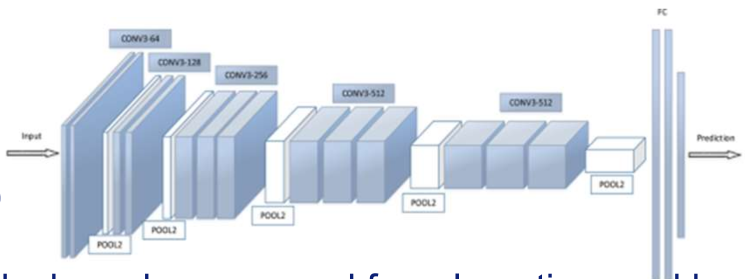
Input (image) ⇨ *Compute descriptive features* ⇨ *Train a classifier (predictor)* ⇨ *Output*

34

Deep learning: what is new?

- Work directly on original image data
- Apply filters in multiple layers (-> more complex features)
- **Learn the filters:** change model parameters to minimize prediction error on training data

Input (image)



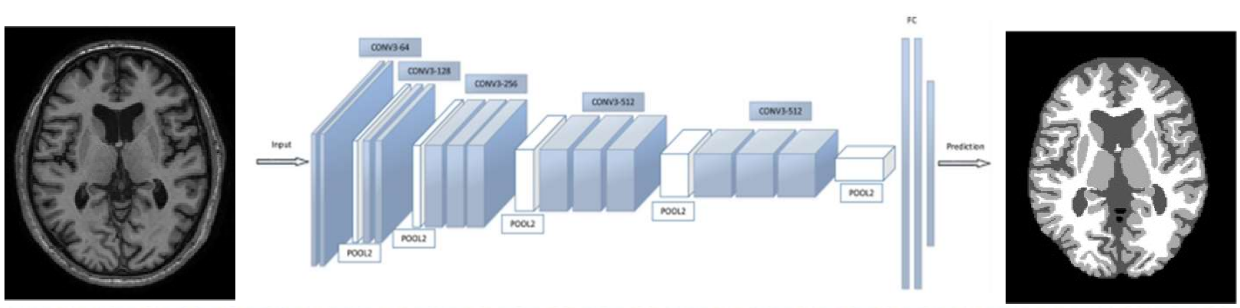
Output

Deep networks have been around for a long time and have only in the past few years become really successful

35

What can we learn?

- Image segmentation

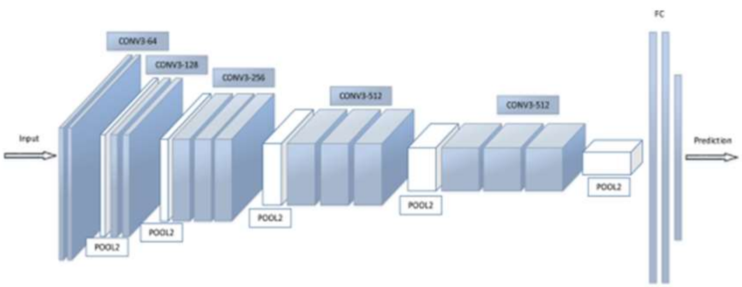


36

Machine learning in radiology: the principle

- Most common: deep learning
- Apply filters in multiple layers (derive complex image characteristics)
- **Learn the filters:** change model parameters to minimize prediction error on training data

Input (image)

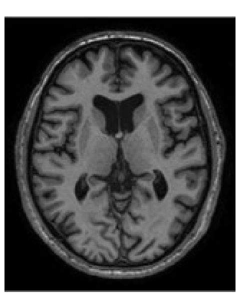


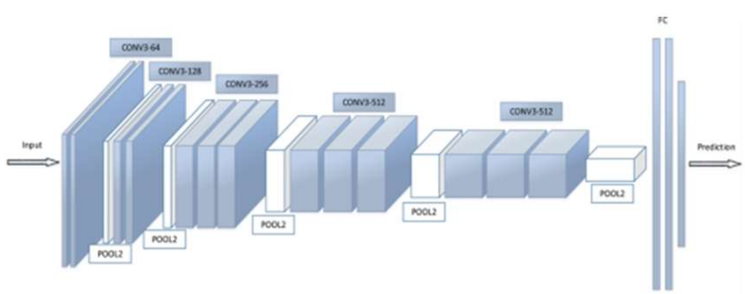
Output

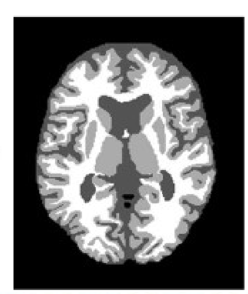
37

What can we learn?

- Image segmentation





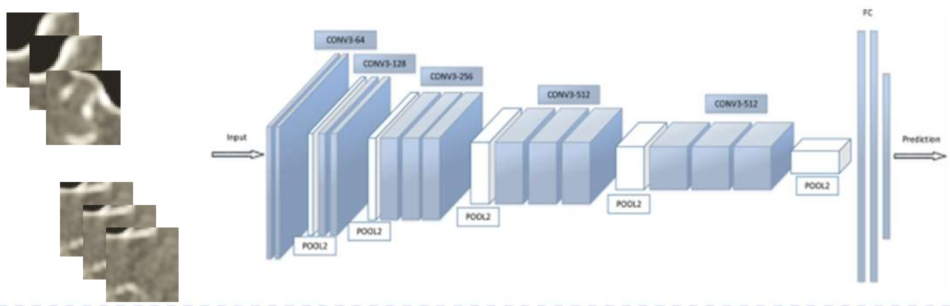


38

What can we learn?

Erasmus MC

- Image segmentation
- Lesion detection



Lesion

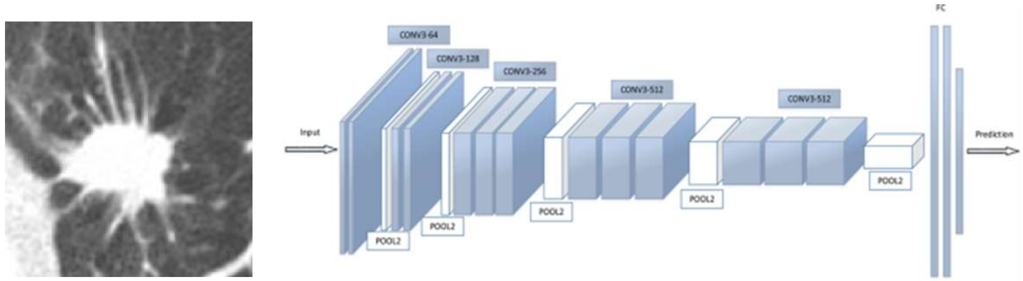
Not lesion

39

What can we learn?

Erasmus MC

- Image segmentation
- Lesion detection
- Lesion characterization



Malignant

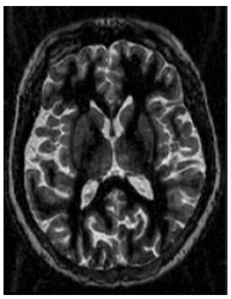
Benign

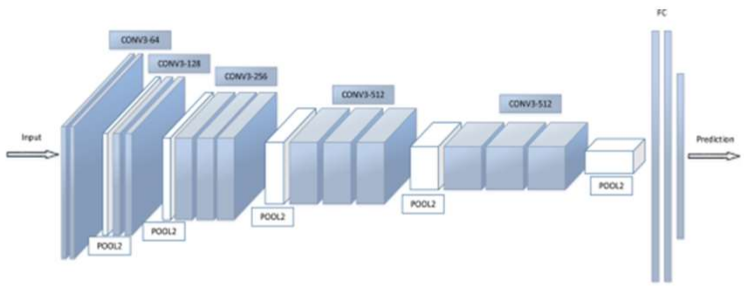
40

What can we learn?

Erasmus MC

- Image segmentation
- Lesion detection
- Lesion characterization
- Diagnosis/prognosis





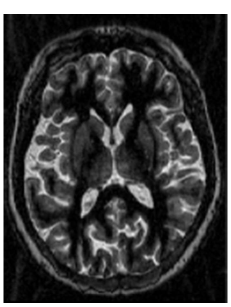
Atrophy

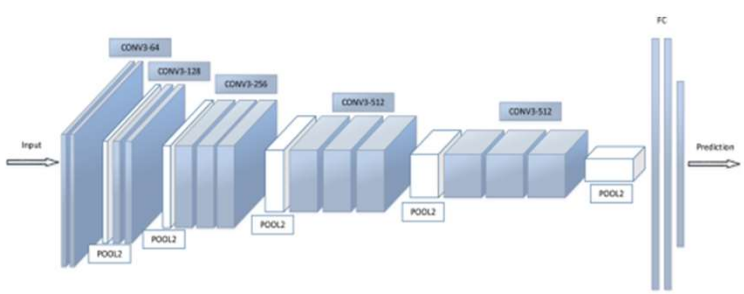
41

What can we learn?

Erasmus MC

- Image segmentation
- Lesion detection
- Lesion characterization
- Diagnosis/prognosis





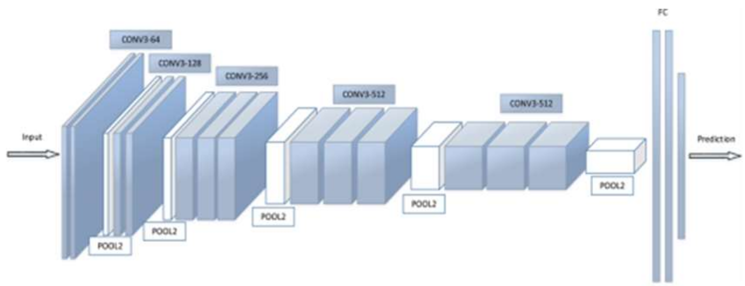
Likely to develop dementia within 5 yr

42

What can we learn?

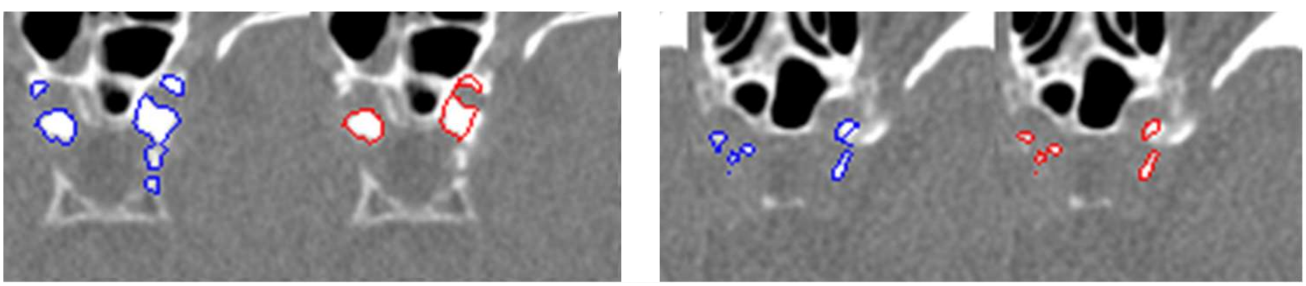
- Image segmentation
- Lesion detection
- Lesion characterization
- Diagnosis/prognosis
- Image synthesis / reconstruction

Low quality image



High quality image

Deep networks can solve many challenging medical imaging tasks as good as humans can

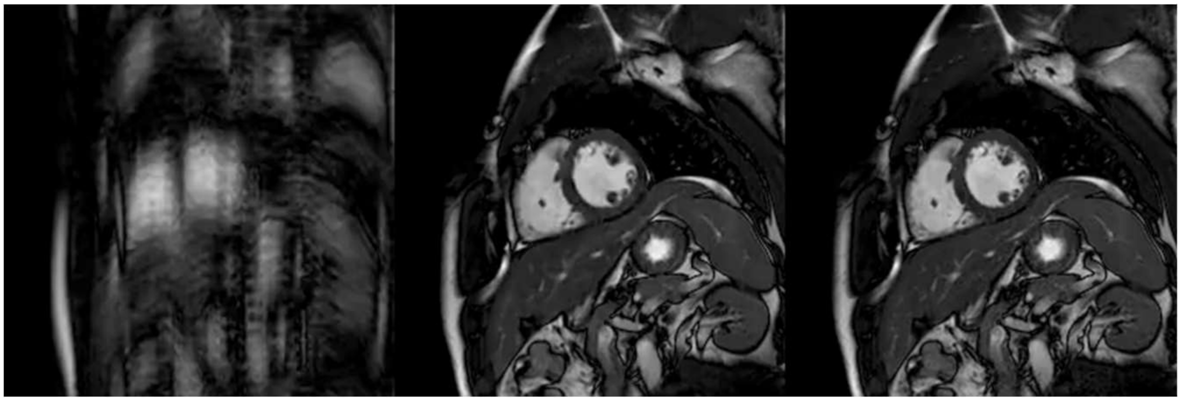


Manual — Automatic

- Segmentation of intracranial calcifications in CT
- Agreement manual/automated scores ICC 0.98
- On average, automated segmentation slightly preferred

Bortsova et al 2021, MICCAI 2017, ECR 2018

MRI reconstruction

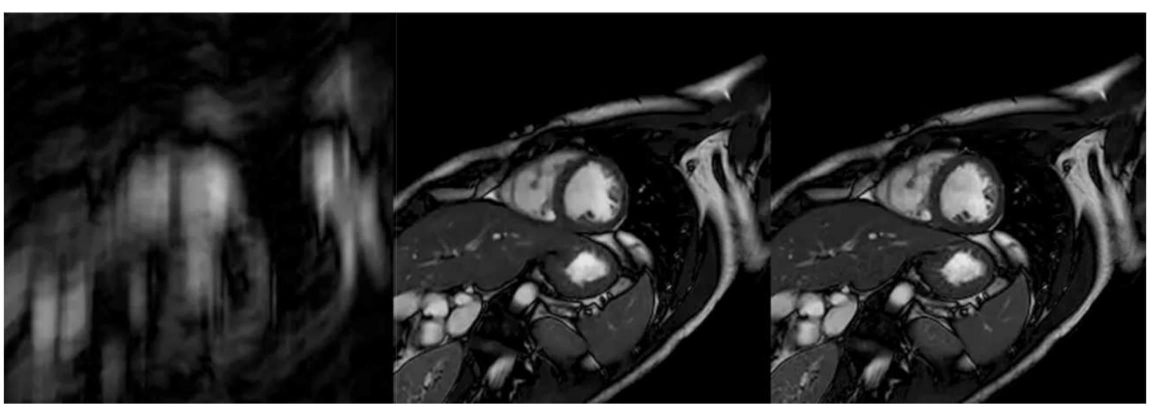


6 x undersampled | CNN reconstruction | Ground truth

Schlemper et al. IEEE TMI 2017. Images Courtesy of Daniel Rueckert

45

MRI Reconstruction

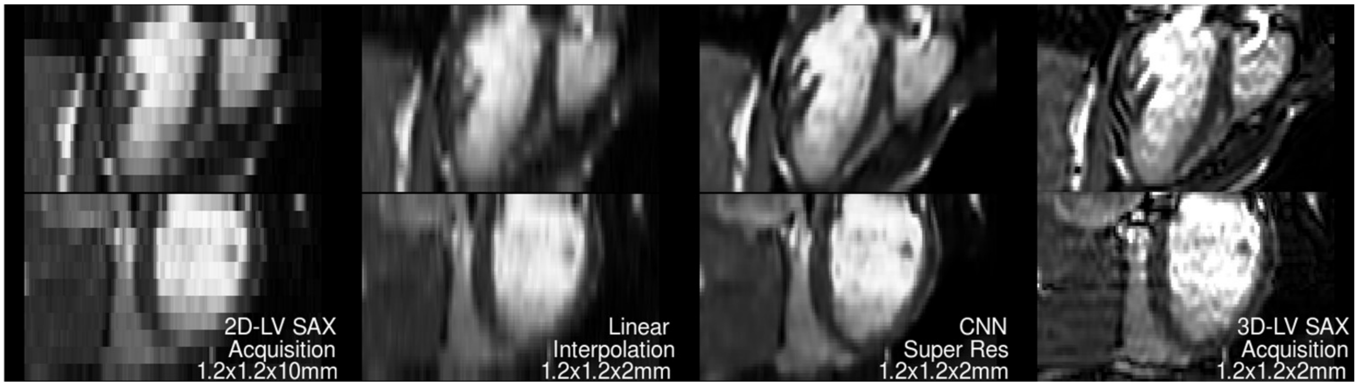


11 x undersampled | CNN reconstruction | Ground truth

Schlemper et al. IEEE TMI 2017. Images Courtesy of Daniel Rueckert

46

Deep learning for image super-resolution



O. Oktay et al. IEEE TMI 2017. Images Courtesy of Daniel Rueckert

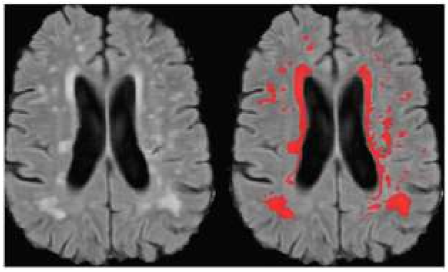
47

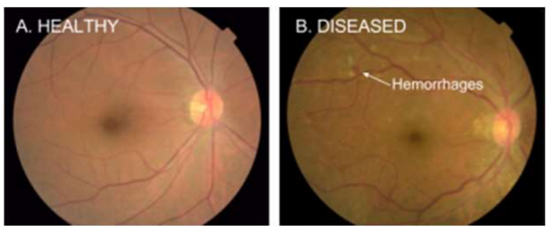
MACHINE LEARNING IN RADIOLOGY: WHERE ARE WE NOW?

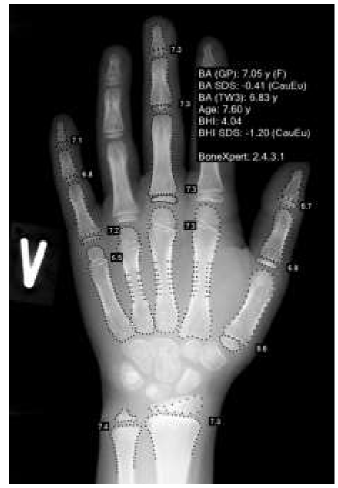
48

Automated medical image analysis

In some applications, automated analysis is now as good as a human

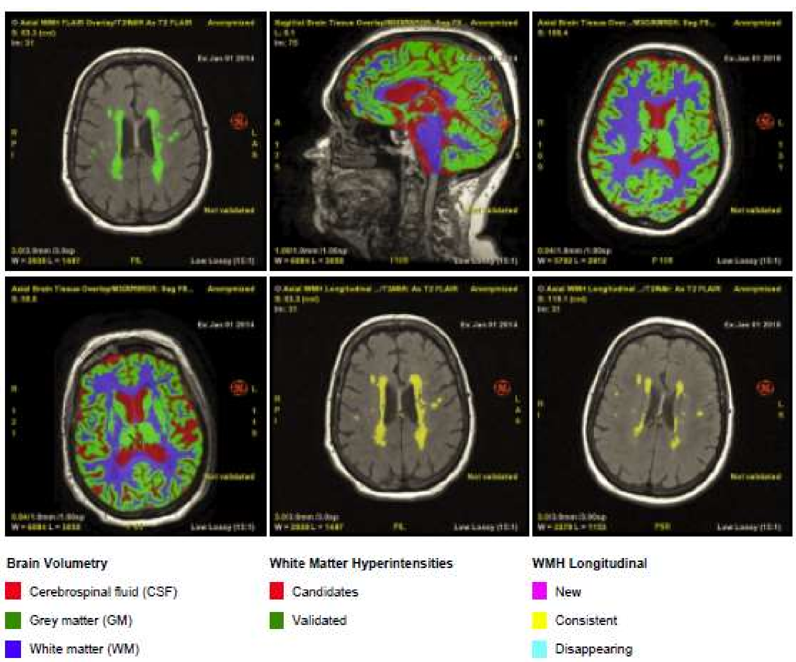






Examples by Ghafoorian et al, Gulshan et al, Thodberg (BoneXpert)

49



Quantib Brain sample report, courtesy Quantib

50

Deep learning is extremely successful in medical image analysis

- However:
  - It requires (a lot of) representative, annotated training data
  - In the medical domain, high quality annotated data is difficult and expensive to obtain

- **How can we learn reliable models with less/lower quality data?**

51

WHAT IF LABELS ARE BAD?

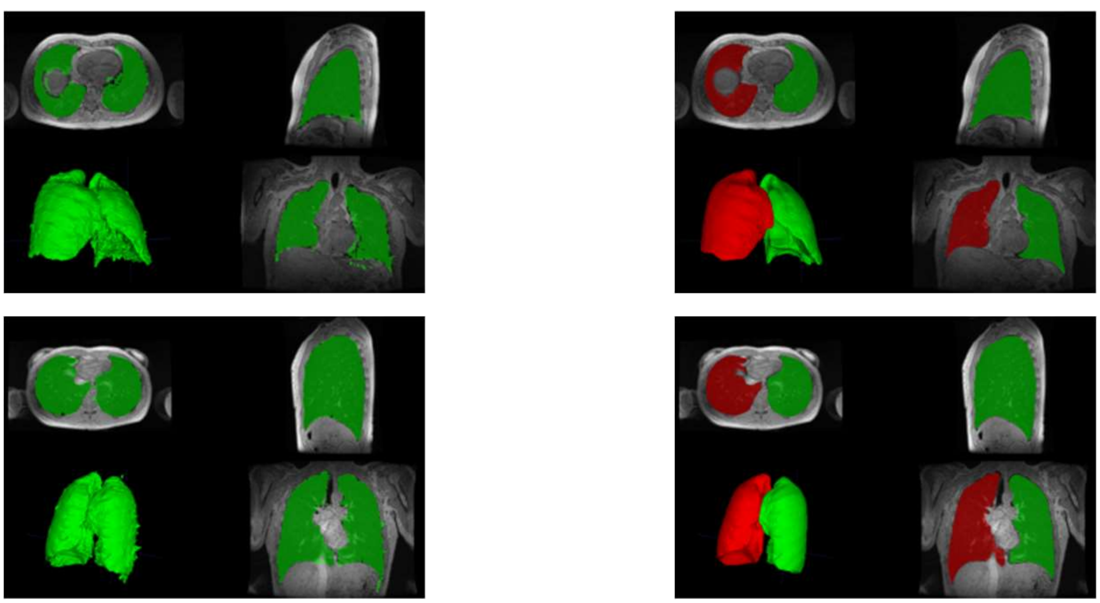
52

Garbage in = garbage out?

Tim Noble and Sue Webster, Dirty White Trash (With Gulls)

53

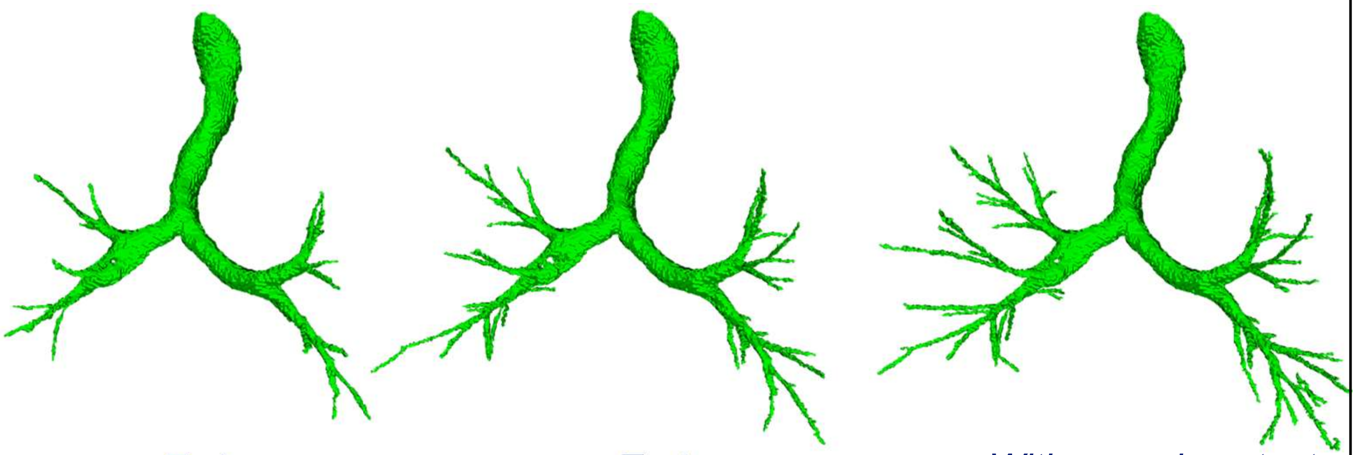
Noise in training labels may be corrected



Train *Test* *Van Tulder 2019*

54

Training data does not have to be perfect



Train

Test

With vessel context

Lo et al, MEDIA 2010

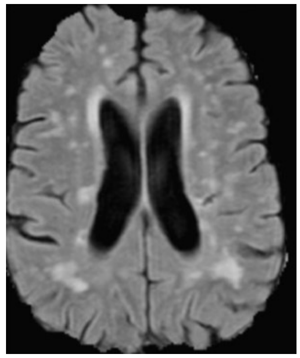
55

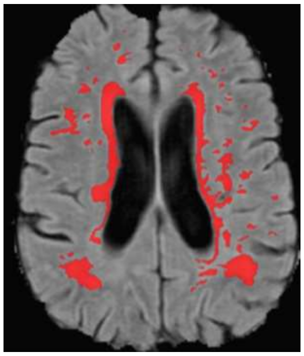
One solution to the “small data” problem

LEARNING FROM WEAK LABELS

56

Weak labels?





Strong

Weak

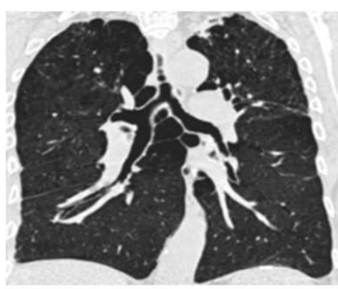
“Moderate lesion load”

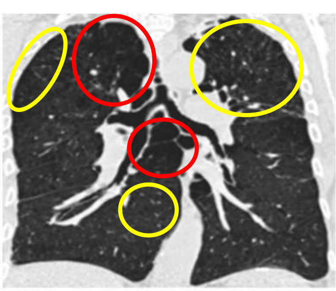
“White matter hyperintensities present”

“3 lesions visible in middle slice”

57

Weak labels?





Strong

Weak

“Emphysema present”

“Upper-lobe predominant mixed-type emphysema”

FEV1/FVC<70%

58

If this works, we can learn from…

- Labels that are easy/quick/cheap to obtain:

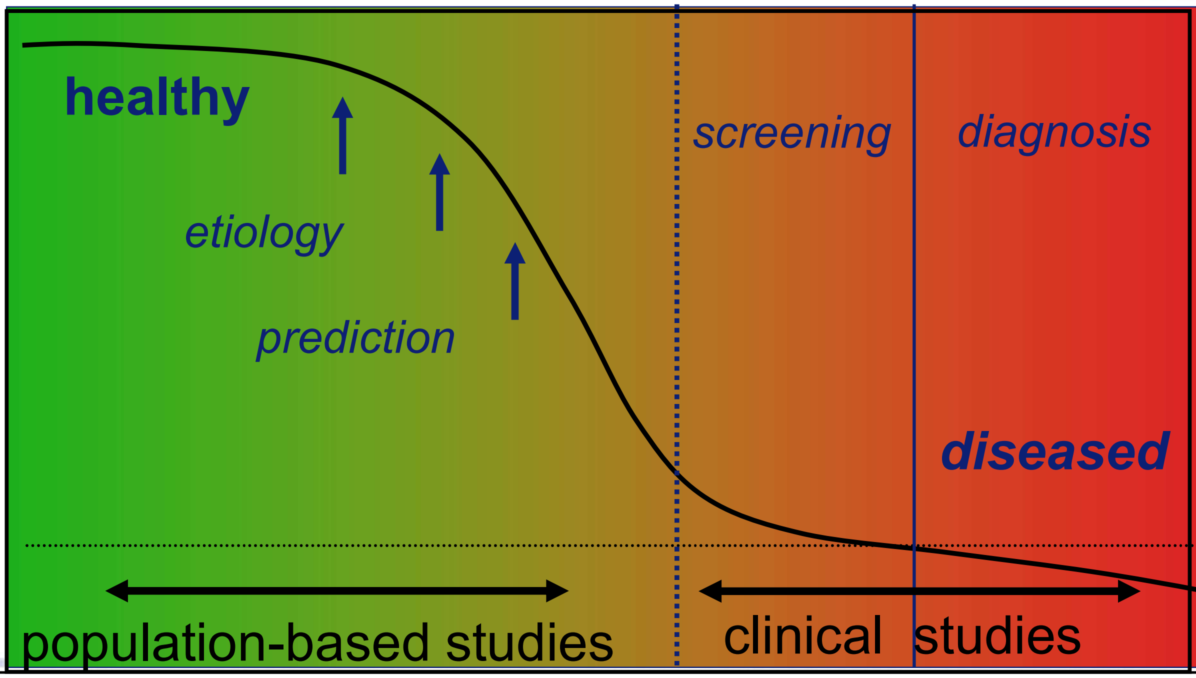
- Available data with semi-quantitative analyses
- Radiology reports
- Clinical databases

59

EXAMPLE: IMAGING BIOMARKERS OF DEMENTIA

60

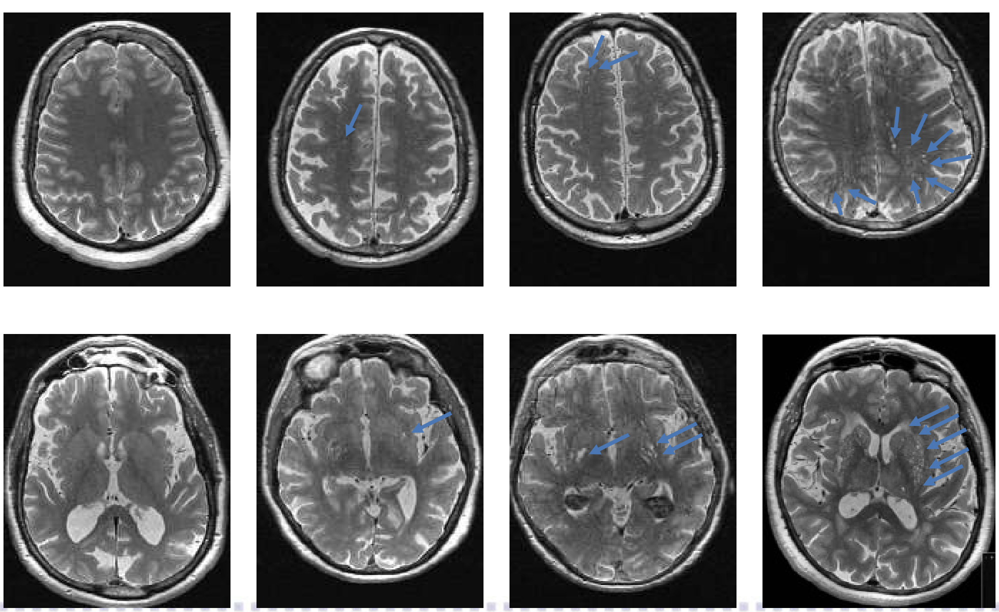
Discover imaging biomarkers for early detection



61

Quantifying enlarged perivascular spaces -- A marker of cerebral small vessel disease

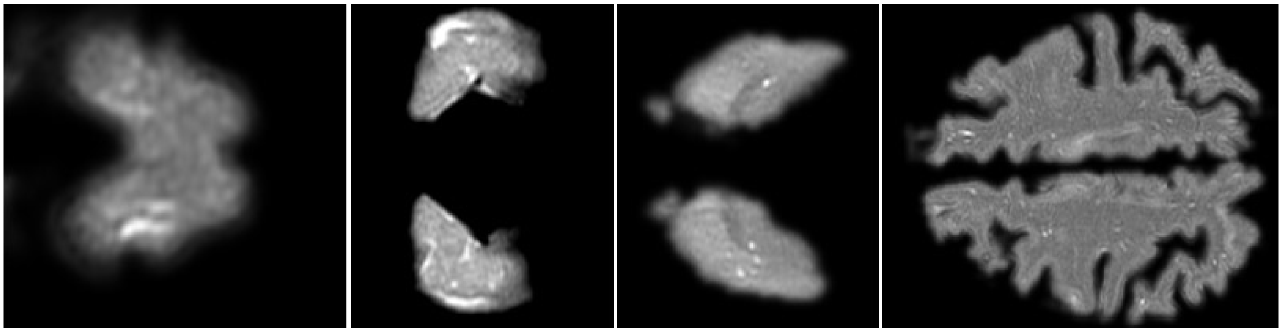
No lesions



Many lesions

62

Current assessment: visual scoring

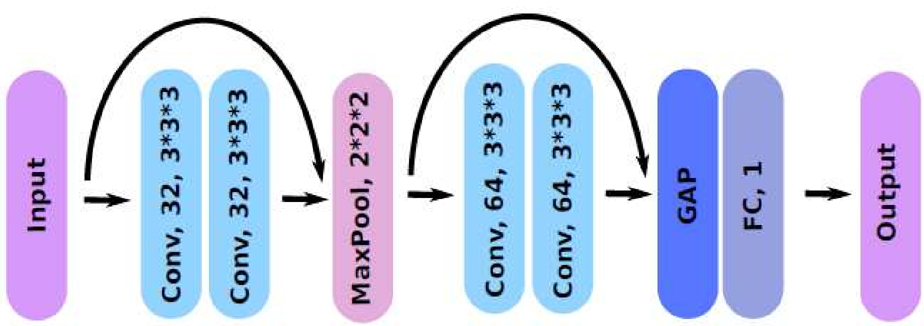


midbrain *hippocampus* *basal ganglia* *white matter*

- Count number of lesions in certain regions in certain slices
- Automated quantification: 3D CNN regression to predict this number

63

Simple regression network



64

Good agreement with expert scores

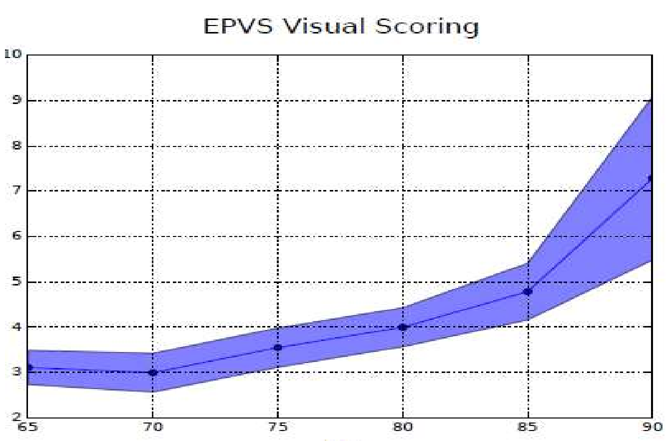
Region	Inter-observer Agreement	Trained on 1600 scans	Trained on 400 scans
Midbrain	0.75	0.75	0.74
Hippocampi	0.82	0.88	0.74
Basal Ganglia	0.62	0.82	0.73
Centrum Semiovale	0.80	0.86	0.80

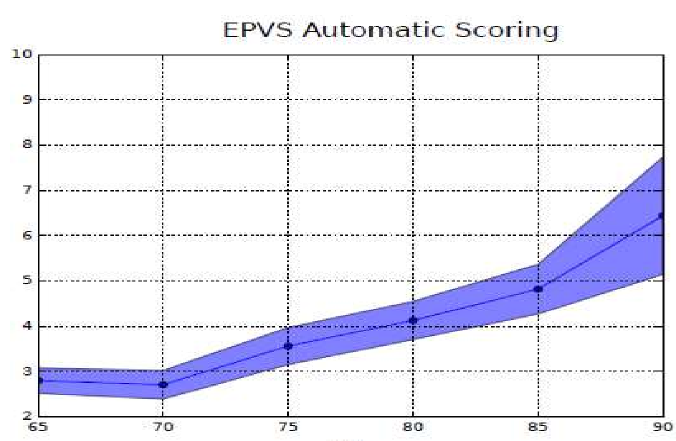
- Human-level performance in all regions
- Still ok when training on 400
- High scan-rescan reproducibility (better than intra-observer agreement)

F. Dubost et al, Medical Image Analysis 2018; Neuroimage 2019

65

Associations with risk factors very similar for automated and visual scores





F. Dubost et al, Neuroimage 2019

66

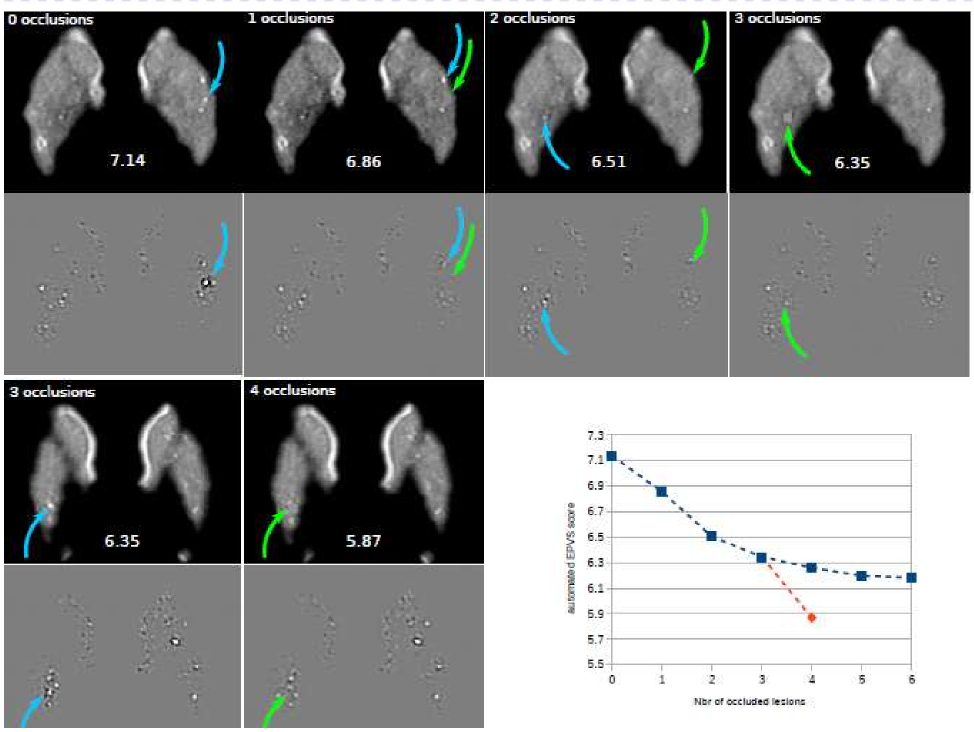
What has the network learned?

- Something that correlates very well with visual scoring

67

Annotated slice

Other slice



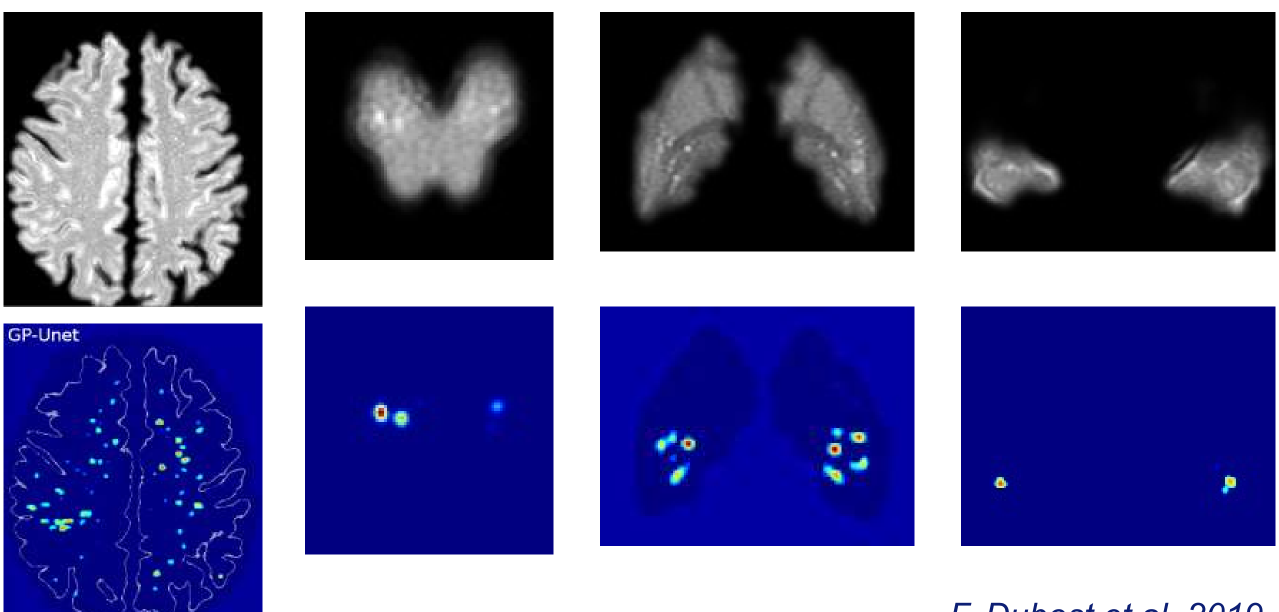
68

What has the network learned?

- Something that correlates very well with visual scoring
- Whole region information instead of single slice

69

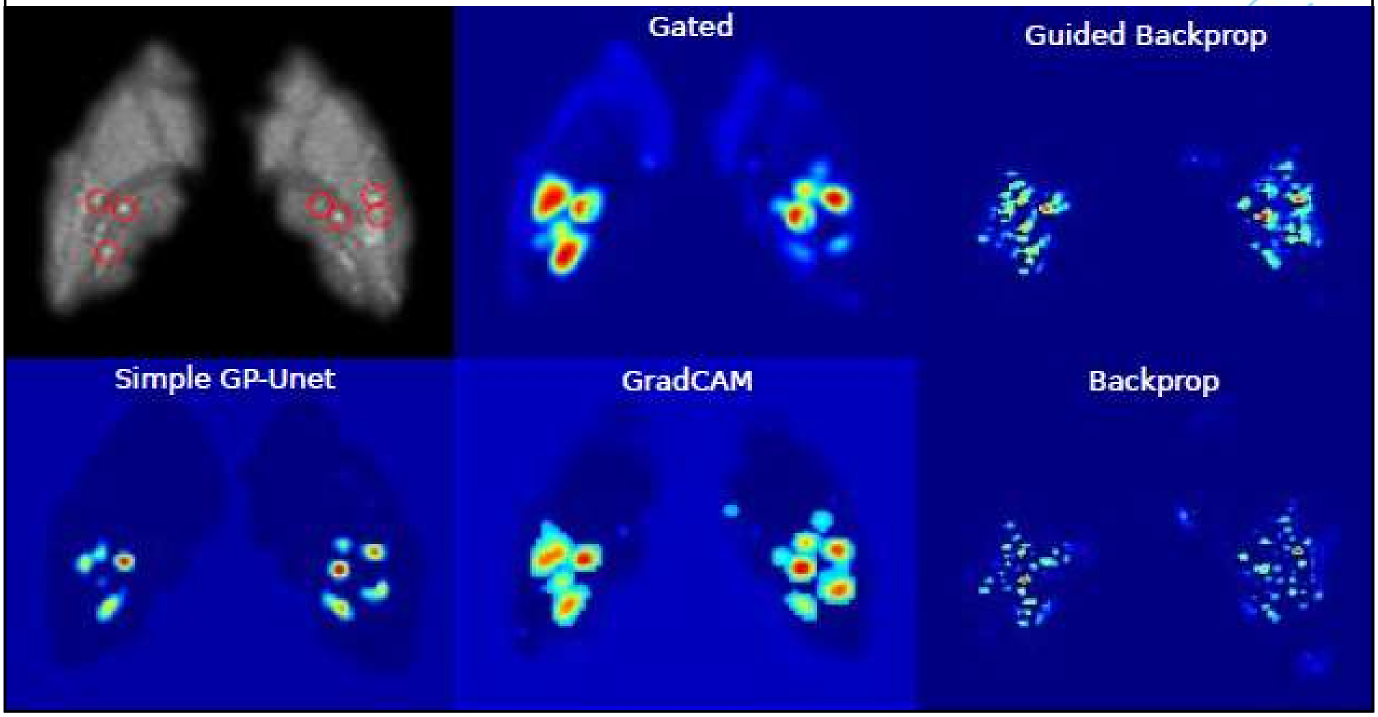
What does the network focus on?



F. Dubost et al, 2019

70

Many techniques to visualize network predictions…



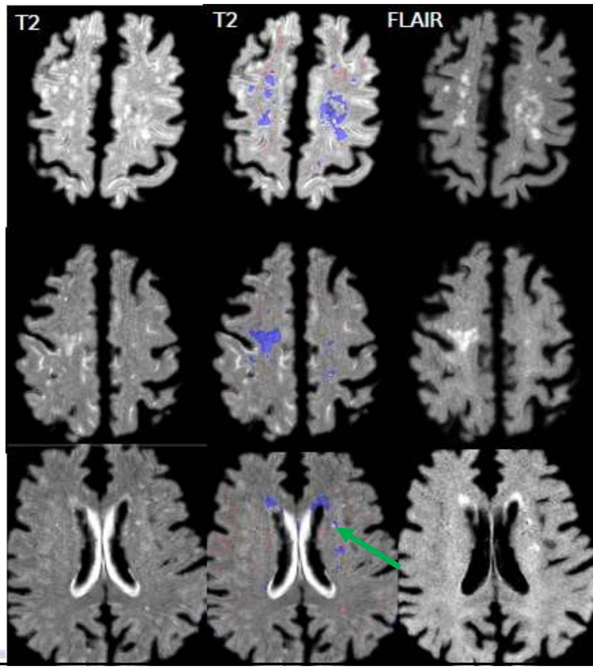
71

What has the network learned?

- Something that correlates very well with visual scoring
- Whole region information instead of single slice
- Seems to focus on the lesions

72

Can discriminate between types of lesions



Very little overlap between EPVS (red) and other types of lesions - white matter hyperintensities (blue) or lacune (green arrow)

73

What has the network learned?

- Something that correlates very well with visual scoring
- Seems to focus on the lesions
- Whole region information instead of single slice
- Can discriminate between types of lesions
- → Could replace visual assessment in large studies
- Potential to improve on visual assessment by using full 3D information and quantifying shape, volume of lesions from activation maps

74

MACHINE LEARNING IN RADIOLOGY: RECENT ADVANCES

75

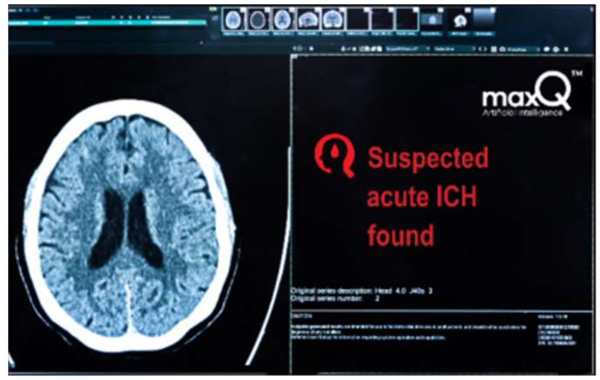
AI to automate diagnosis

- 2018: The first ever autonomous AI system cleared by the FDA to provide a diagnostic decision

76

AI based triage

- Detect urgent cases faster
- Multiple approved AI systems to detect brain bleeds in CT



Medtronic partners with Viz.ai to accelerate adoption of stroke triage software in the hospitals across the nation

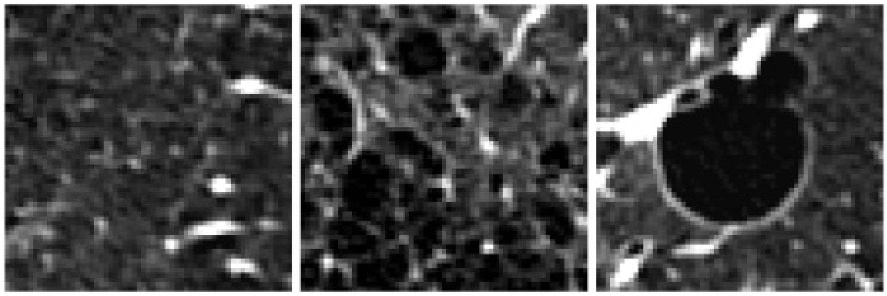
Aidoc gets CE Mark for deep learning solution for head and neck CT imaging

77

OR (IN RESEARCH): AI TO UNCOVER NEW PATTERNS

78

AI to improve diagnosis: uncovering new patterns



- Chronic obstructive pulmonary disease (COPD)
- Various disease subtypes, with different patterns in CT
- Can CT pattern predict lung function?

79

Chronic Obstructive Pulmonary Disease (COPD)

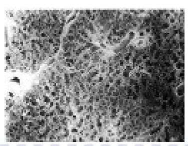
- Major cause of death worldwide
- Early stages severely underdiagnosed
- Poorly understood
- Different phenotypes may require different treatment

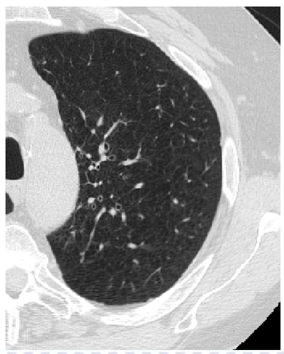
80

Quantifying COPD

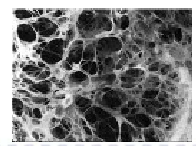
- Lung function tests are the gold standard for COPD diagnosis and grading
- For subtyping and early assessment, we need imaging (CT)
- Signs of COPD:
  - Destruction of lung tissue (emphysema)
  - Airway wall thickening

NORMAL





COPD



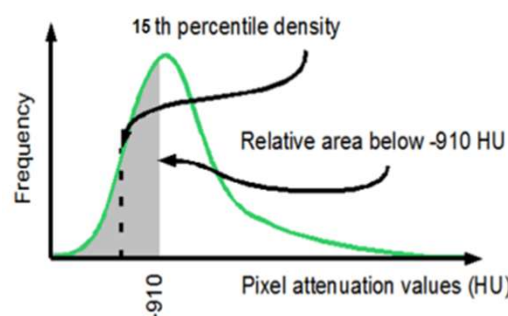
Quantifying emphysema

Current CT quantification of emphysema is based on CT density values

Cannot discriminate patterns

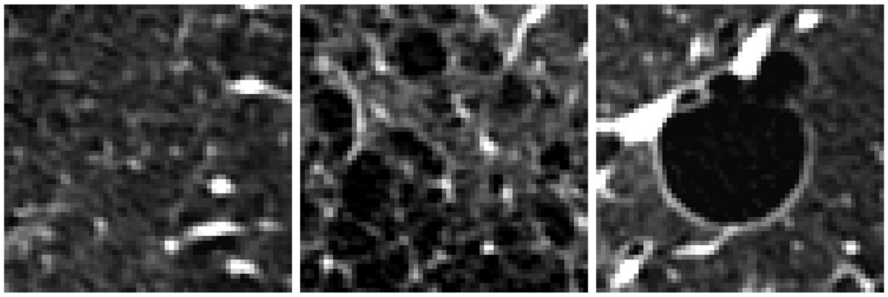
Sensitive to changes in inspiration level, scan protocol

Relatively low correlation between densitometry and visual assessment



Beyond density: can we do better?

Erasmus MC

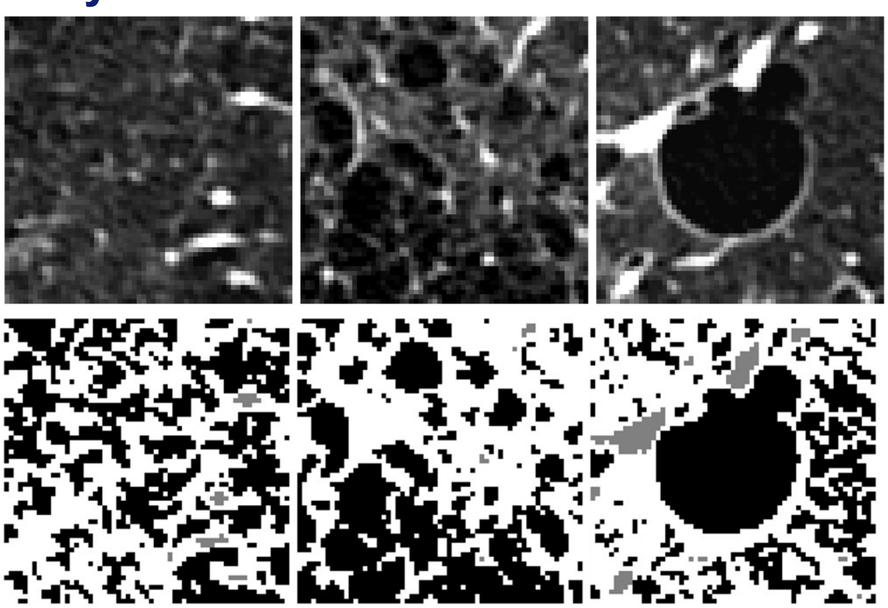


Emphysema has various disease subtypes, with different patterns in CT

83

Different patterns may have identical densitometry values!

Erasmus MC



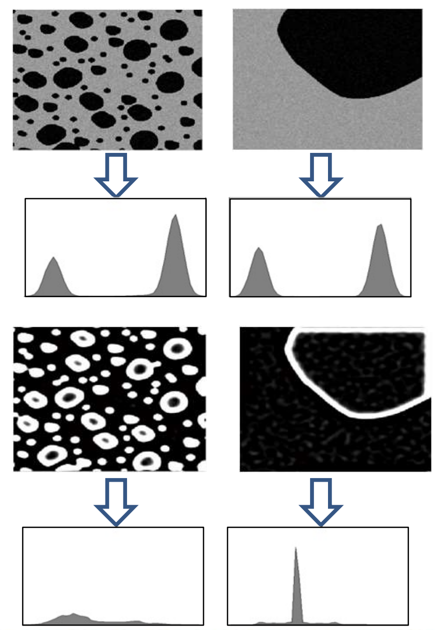
84

... but we can dicsriminate them by their texture

Appearance differs by sub-type and severity

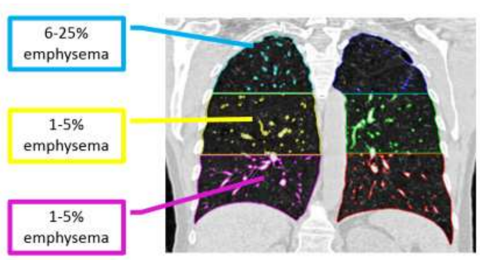
CT lung density can not capture this

A proper texture descriptor can

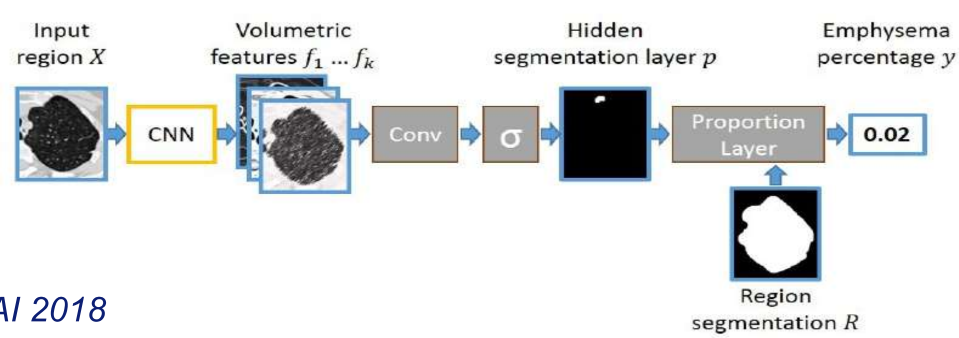


85

Learning from visual scoring



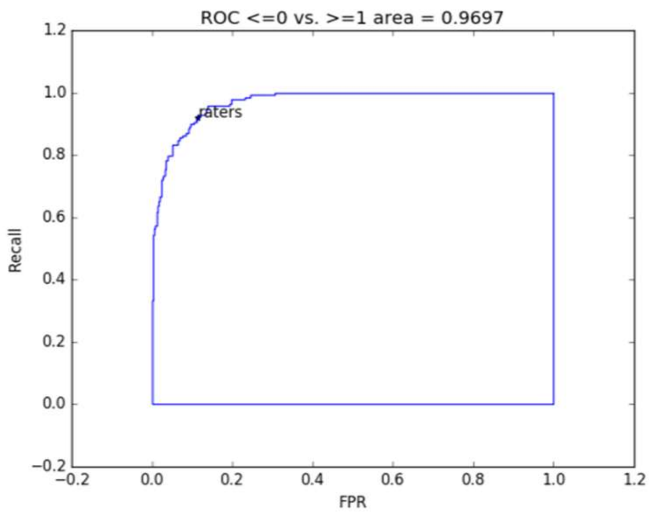
- Train a network to reproduce these scores
- Via an intermediate (segmentation) image with the right proportions



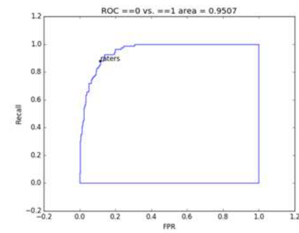
Bortsova et al MICCAI 2018

86

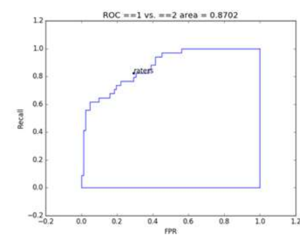
Performs similar to trained observers



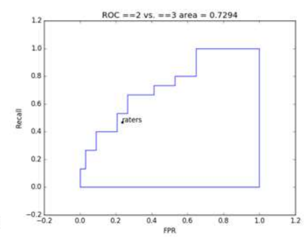
Emphysema detection



Healthy vs. mild (grade 1)



Grade 1 vs. 2



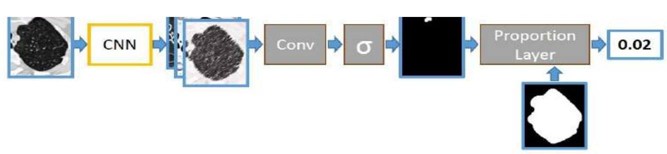
Grade 2 vs. 3

Gerda Bortsova, ECR 2018, MICCAI2018

87

ProportionNet consistently better than regression network -- especially in smaller training sets

Architecture:	GAPNet		ProportionNet	
Training set size\Task:	Presence	Extent	Presence	Extent
small sets (50, 75, 100)	0.90 ± 0.04	0.74 ± 0.06	**0.94 ± 0.01**	**0.79 ± 0.02**
medium sets (150, 200, 300)	**0.96 ± 0.01**	0.80 ± 0.02	**0.96 ± 0.01**	**0.84 ± 0.01**
large set (700)	0.96	0.79	**0.97**	**0.86**



Bortsova et al MICCAI 2018

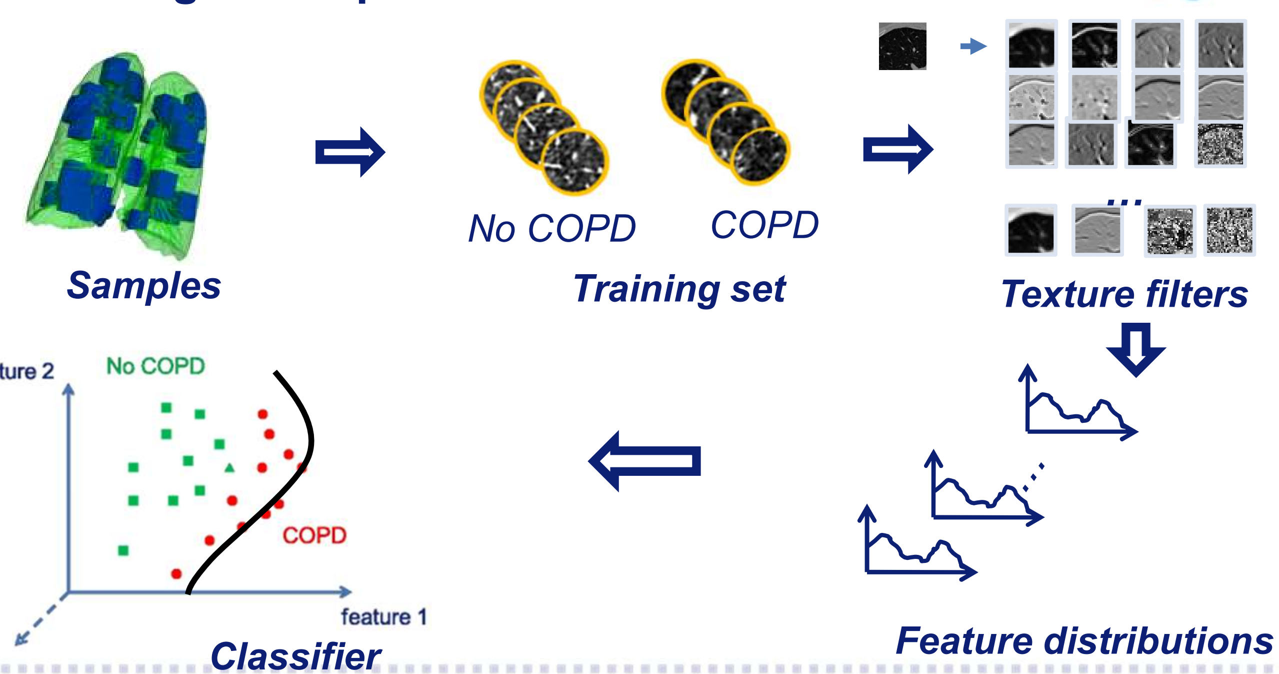
88

Comparison with other methods

Labels:	LLP								MIL	
Training set size:	100 subjects				700 subjects				700 subjects	
Region:	RU		LU		RU		LU		RU	LU
Metric:	ICC	r_s	ICC	r_s	ICC	AUC	ICC	AUC	AUC	AUC
Densitometry	-	0.23	-	0.14	-	0.59	-	0.54	0.59	0.54
[3] and [4]	0.72	-	0.63	-	-	-	-	-	0.89	0.87
GAPNet+RMS	-	-	-	-	0.79	0.93	0.76	0.90	-	-
GAPNet+LPI_6	0.77	0.62	0.74	0.52	0.82	0.96	0.76	0.94	**0.96**	**0.94**
ProportionNet	**0.87**	**0.73**	**0.81**	**0.66**	**0.87**	**0.97**	**0.85**	**0.95**	**0.96**	**0.94**

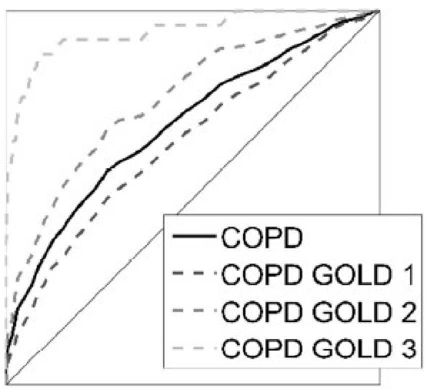
89

Learning COPD patterns

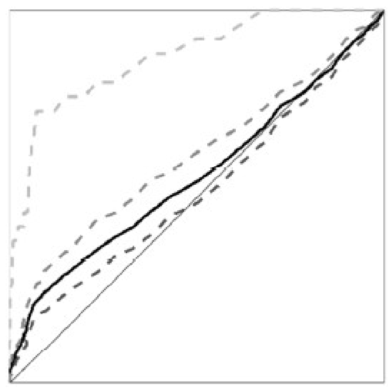


90

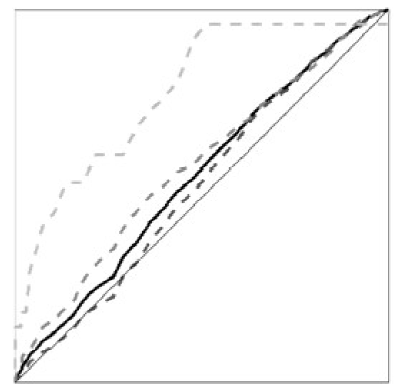
Machine learning detects COPD better than conventional density measures



Machine learning



Area below -950 HU

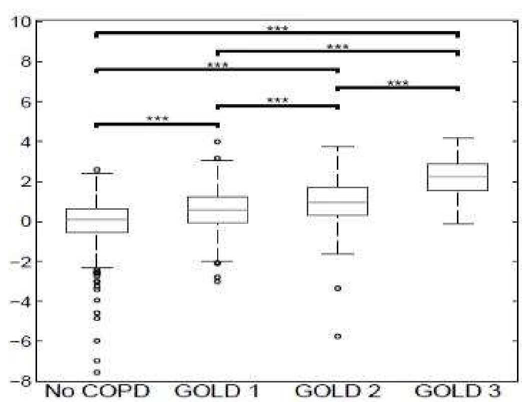


15th percentile

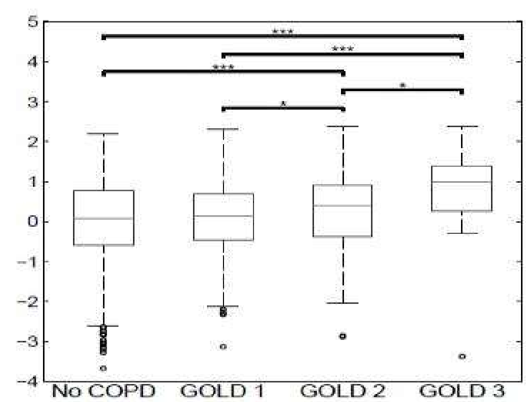
Sorensen et al '19

91

COPD Severity staging: Machine learning can detect early stages



Machine learning



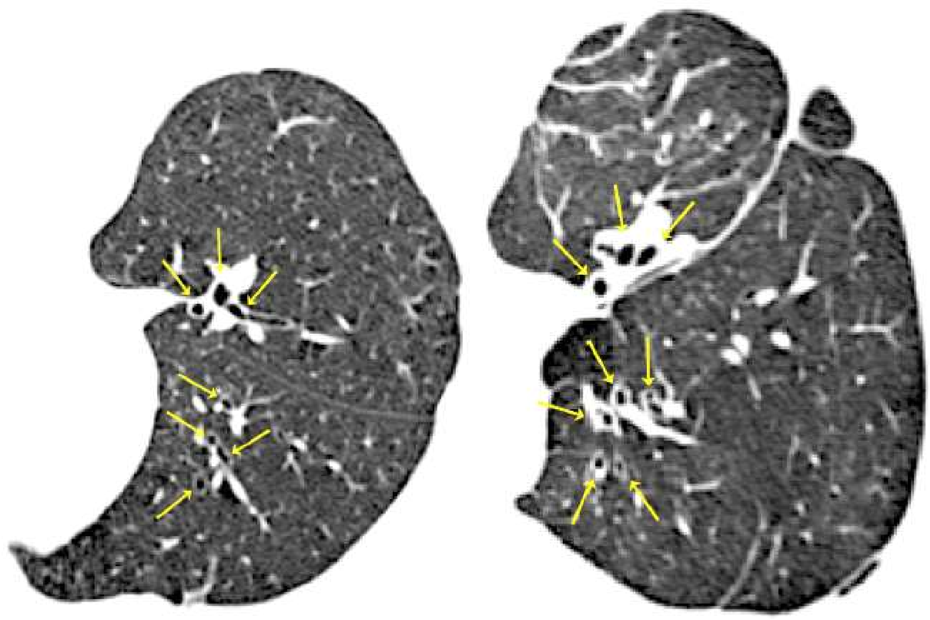
Densitometry (RA 950)

ML could also predict future development of COPD (AUC=0.6, p<0.001) whereas densitometry could not (AUC~0.5, p>0.05)

Sørensen et al '19

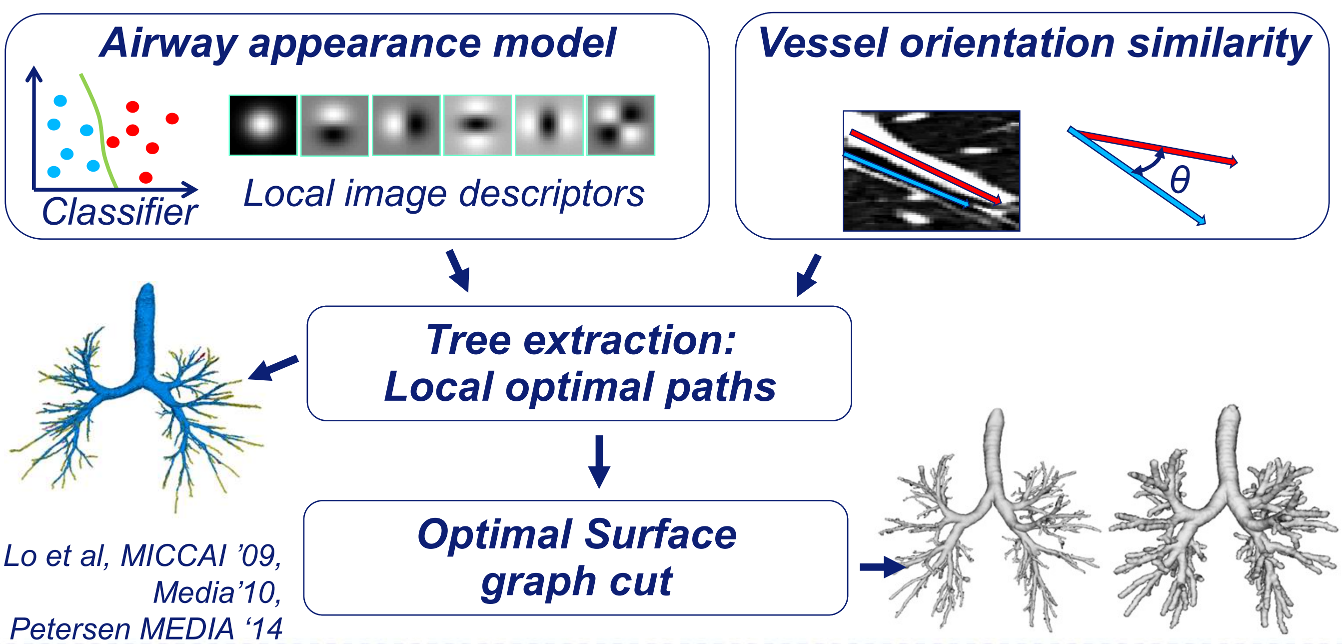
92

COPD on CT: Airway wall thickening



93

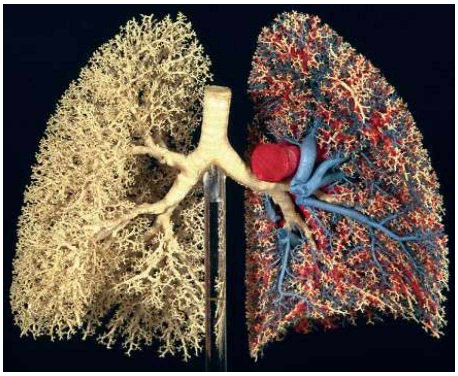
Airway tree segmentation: A learning approach



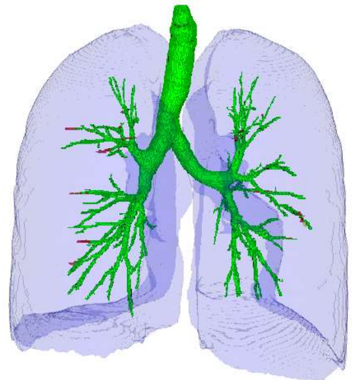
Lo et al, MICCAI ’09,
Media’10,
Petersen MEDIA ‘14

94

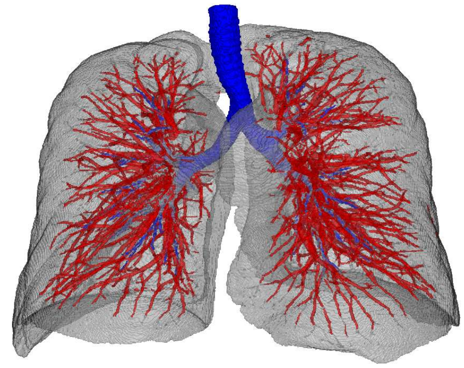
Airways and vessels in CT



Plastic cast of airways (white) and vessels



Extracted airways



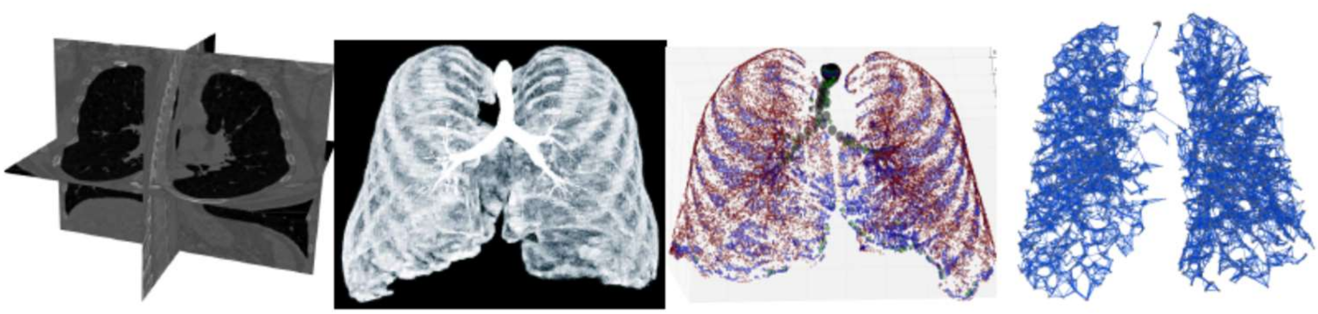
Extracted airways and vessels

Many smaller (peripheral) airways are not detected in CT!

Cast Figure by Weibel, 2009, Swiss Med Wkly

95

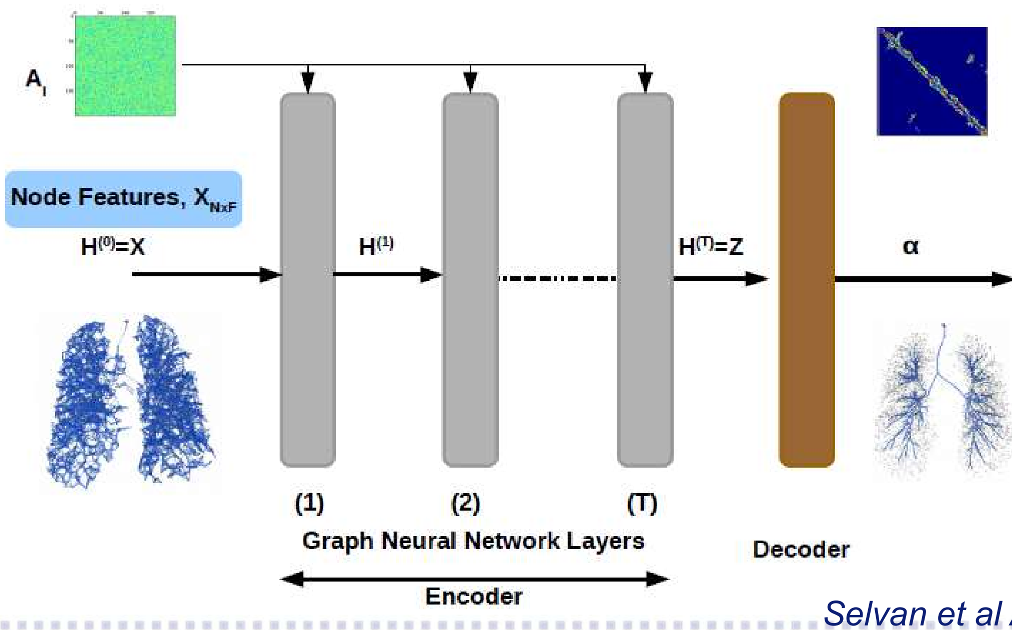
More exploratory airway segmentation using graph convolutional networks



- Construct **overcomplete** graph including many potential airway branches
- Extract airway tree using graph refinement
- Not sequential, may find more complete trees

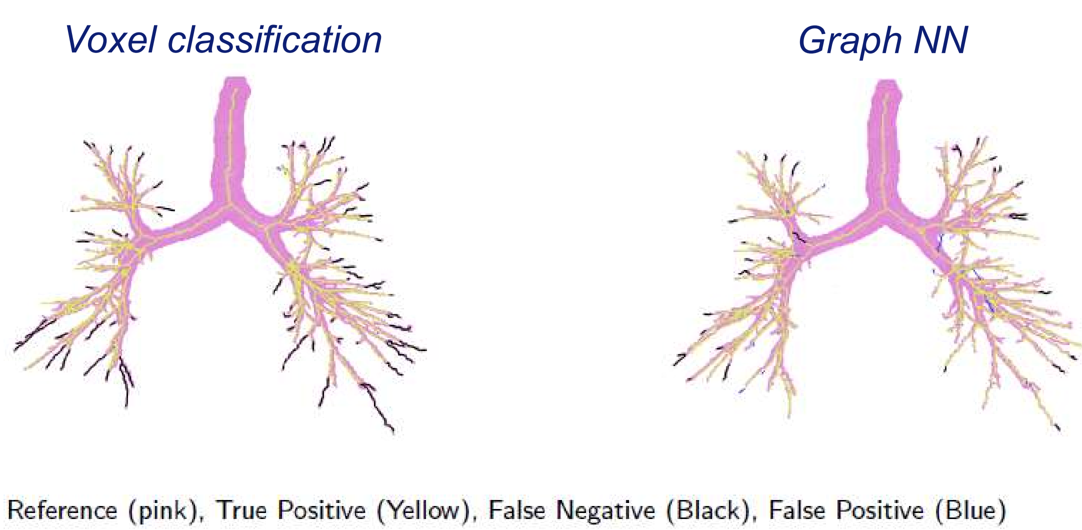
96

Graph refinement using Graph NN



Selvan et al ArXiv '18

97



Selvan et al '18

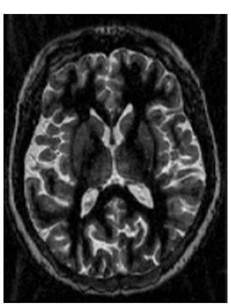
98

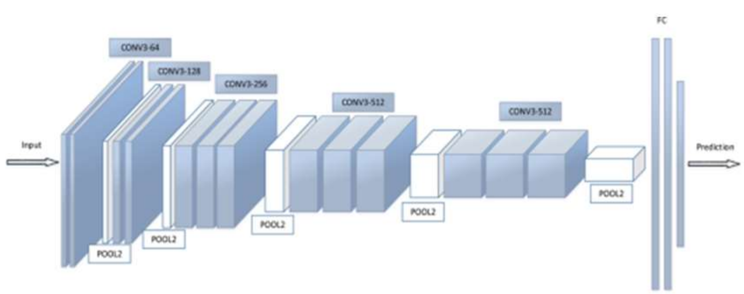
MACHINE LEARNING IN RADIOLOGY: THE FUTURE

99

AI for risk stratification and clinical decision support

Note: many conventional imaging biomarkers are predictive of disease as well. The difference is that we now **optimize** for prediction

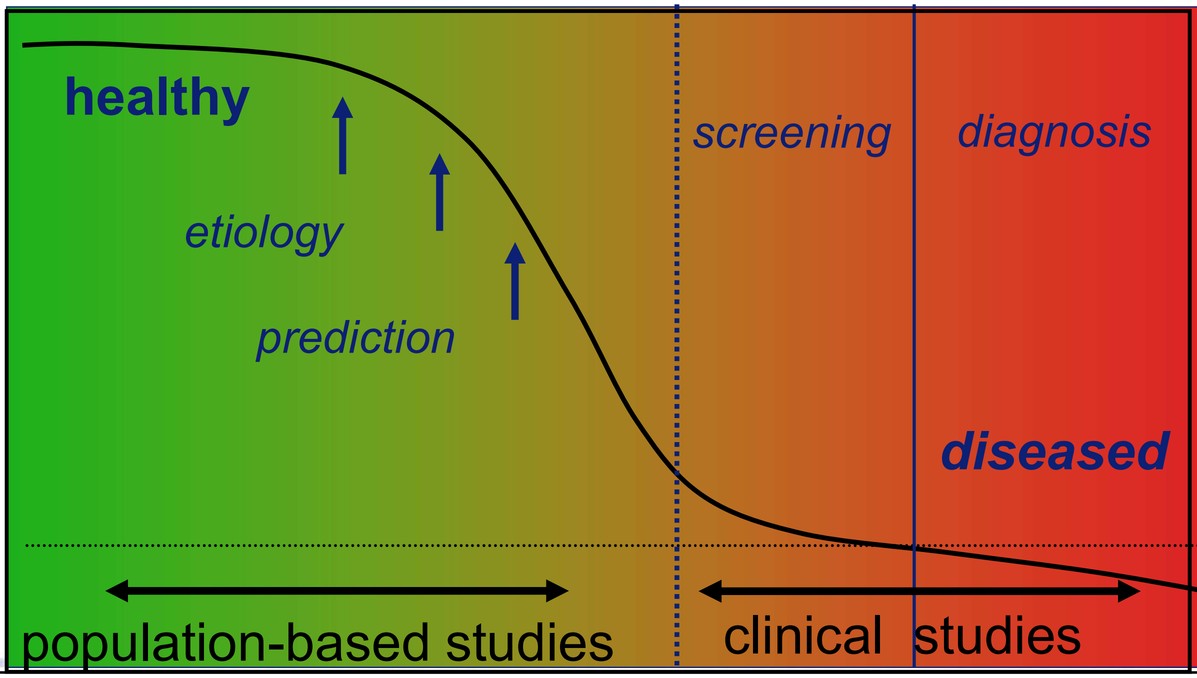




Risk of developing dementia within 5 yr

100

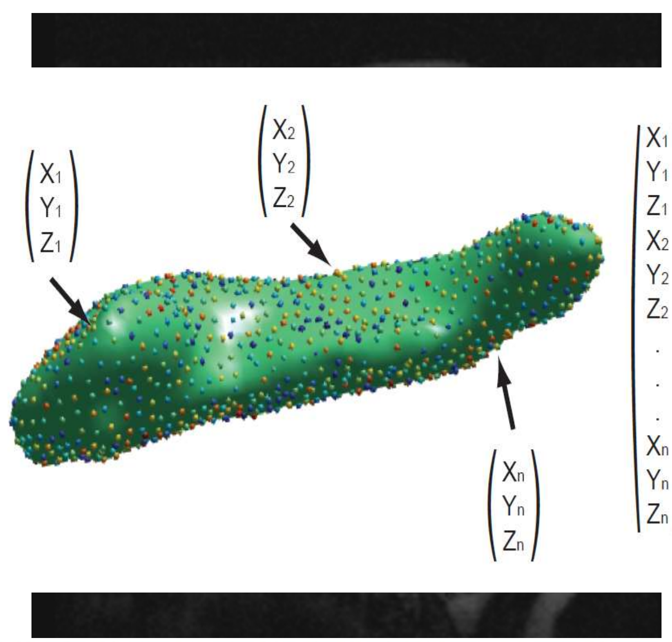
Learning imaging biomarkers for early detection



101

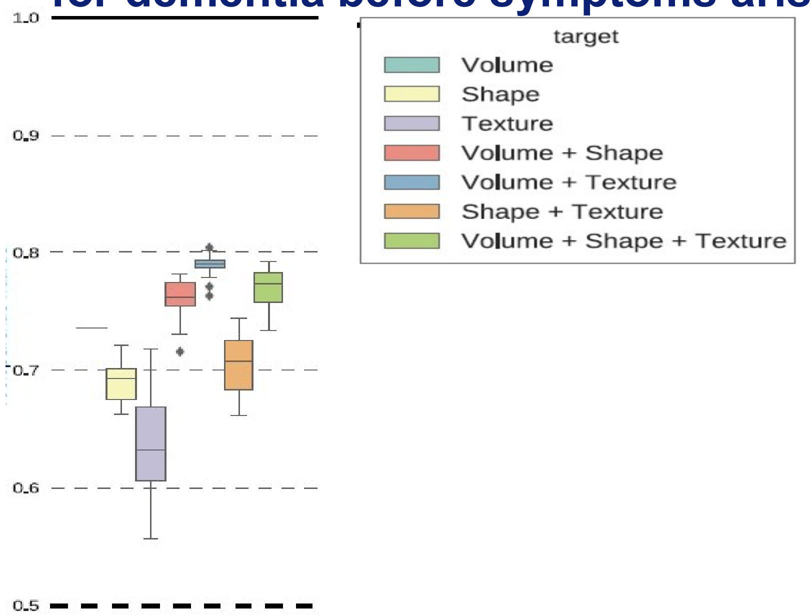
Example: learning imaging biomarkers of dementia

- Large group of **healthy volunteers age 60+;** Some develop dementia, most do not
- Segment hippocampus
- Determine corresponding points on the shape surface (features)
- Compute texture features
- Train a classifier to recognize dementia pre-stage



102

Hippocampus shape and texture are predictive for dementia before symptoms arise



Achterberg et al

103

Patterns of shape differences

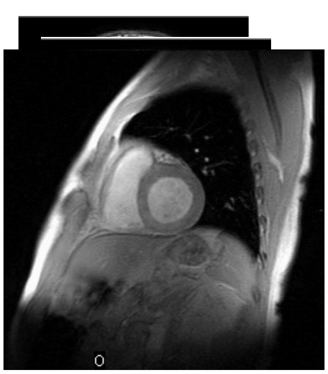
Achterberg et al, HBM 2013

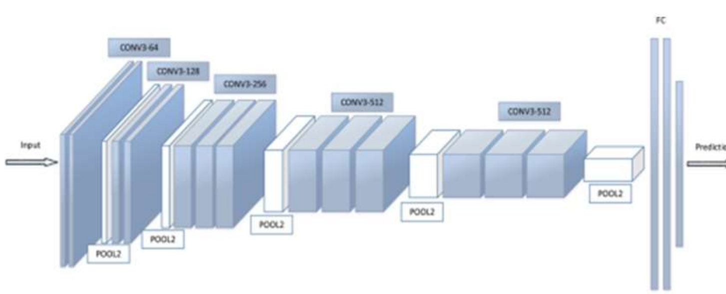
104

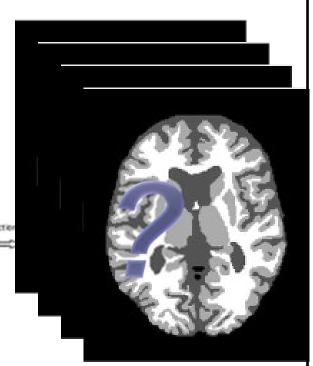
OPEN PROBLEMS

105

What if training data is not representative?

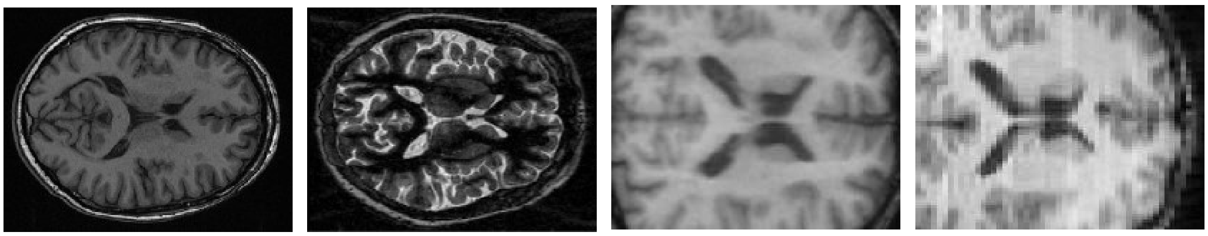


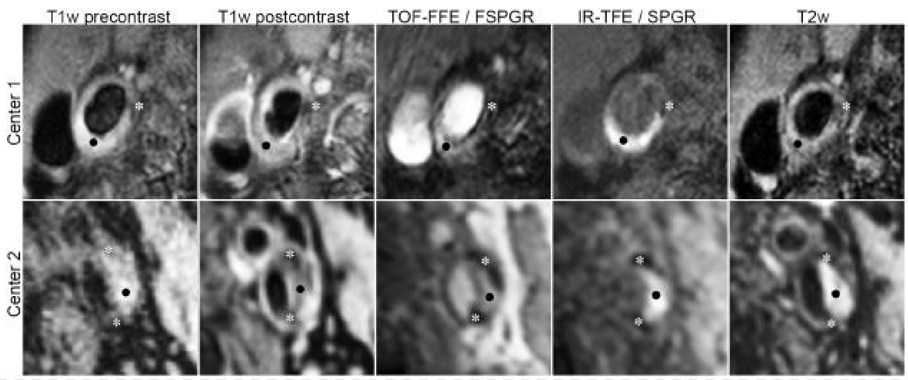




106

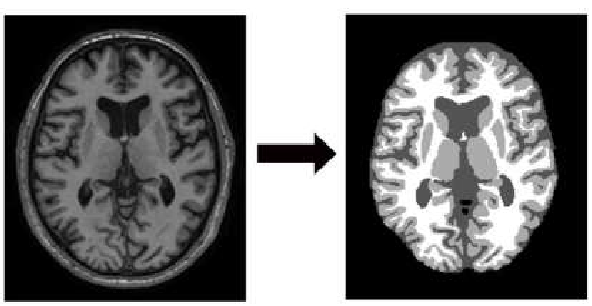
Challenge: Difference between scan protocols



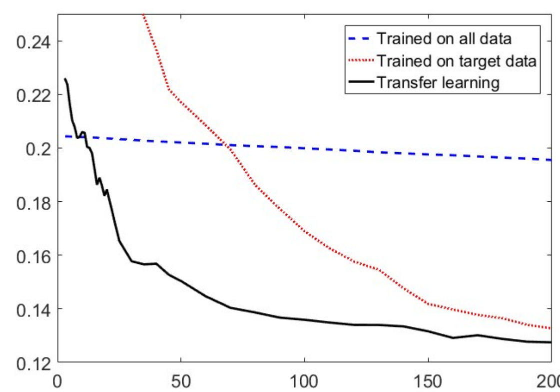


107

Transfer Learning Improves Supervised Image Segmentation Across Imaging Protocols



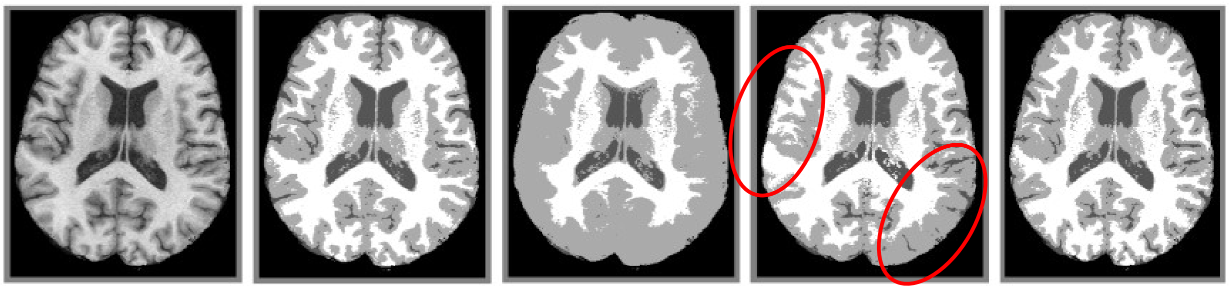
Brain tissue segmentation



Van Opbroek et al, MLMI 2012, TMI 2014

108

Segmentation across scan protocols

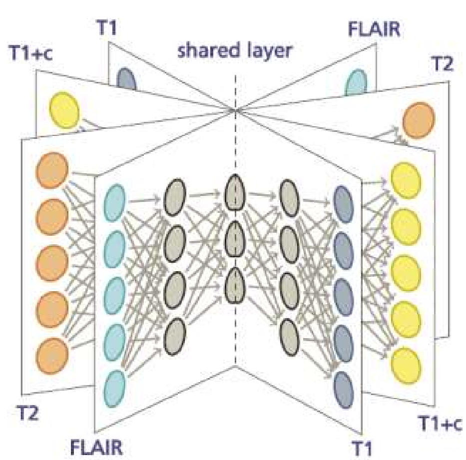


MRI	Manual Segmentation	Trained on other sources+ (few) labeled target data	Trained on (few) labeled target data	Transfer Classifier

Van Opbroek 2018

109

Learning modality-invariant representations

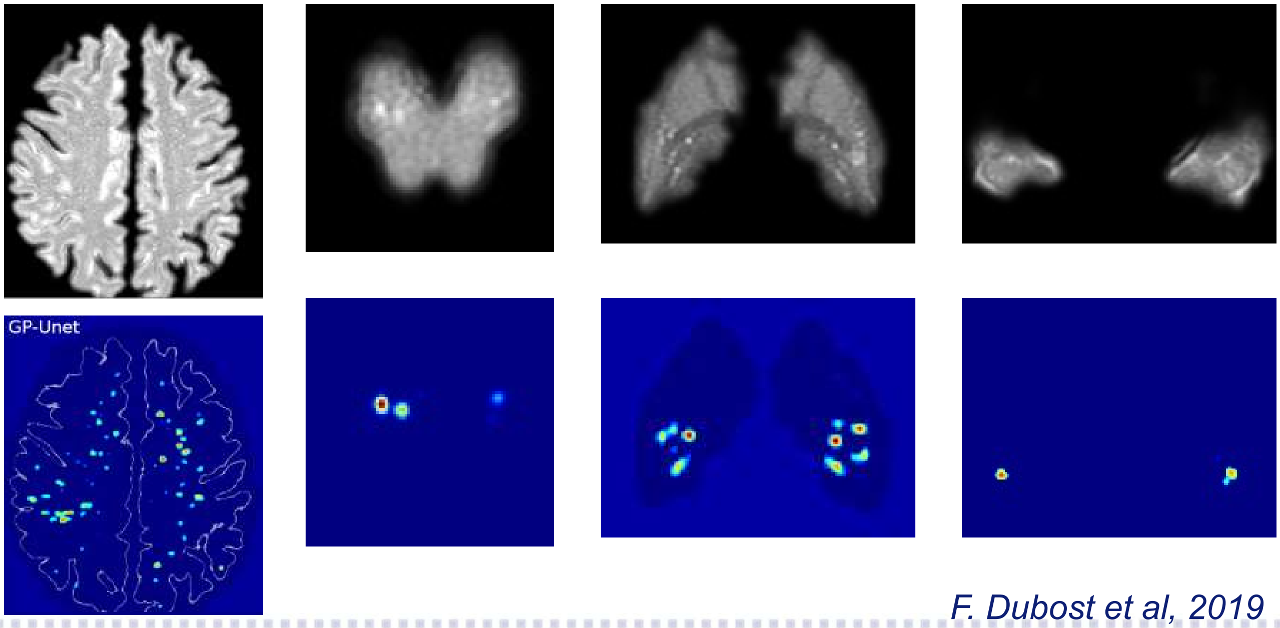


- Use *paired* data -> exploit spatial correspondence between modalities
- Shared convolutional decoder-encoder network
- Similarity term in shared layer
- Feature normalization
- Modality dropout
- Improves cross-modality classification while maintaining within modality performance

Van Tulder and de Bruijne, MCV 2016. IEEE TMI 2018

110

Model interpretation



F. Dubost et al, 2019

111

Risk of biased decisions

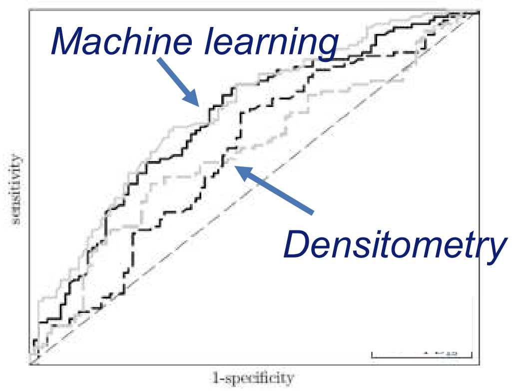
112

Challenge: modeling with confounders

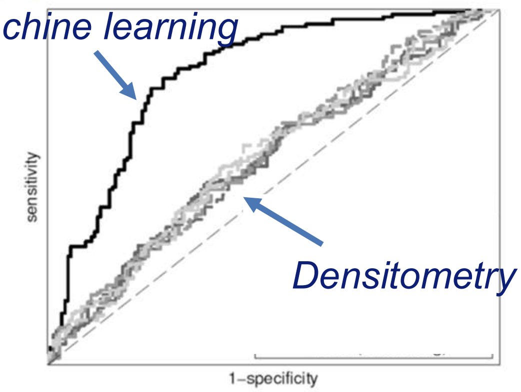
Model can fit to – possibly irrelevant – features that correlate with outcome

Example: COPD classification based on image texture

Results are very different when balancing gender



Balanced dataset



Biased dataset

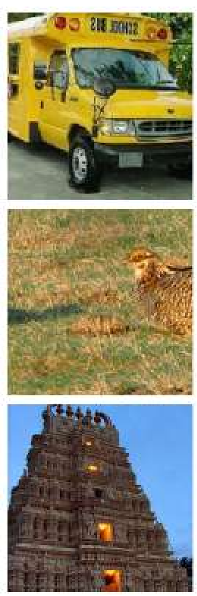
Sorensen et al, TMI'12

113

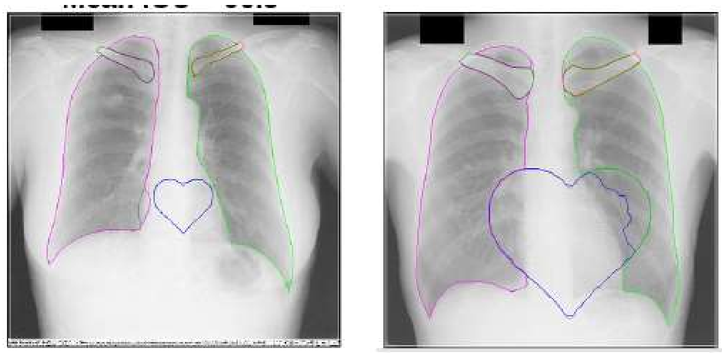
AI systems can be fooled

The machines can not "see" like us

But they are becoming very good at predicting our decisions



114



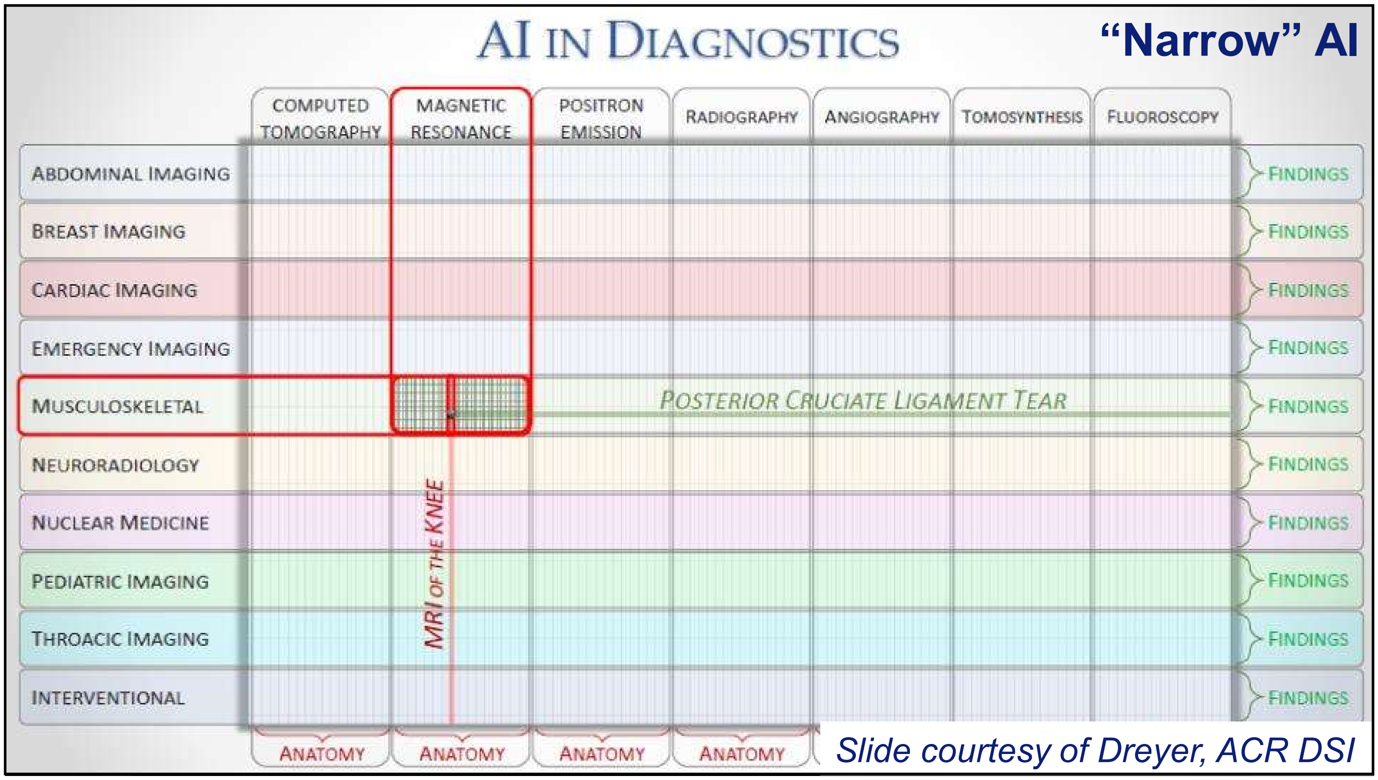
Work by Gerda Bortsova

115

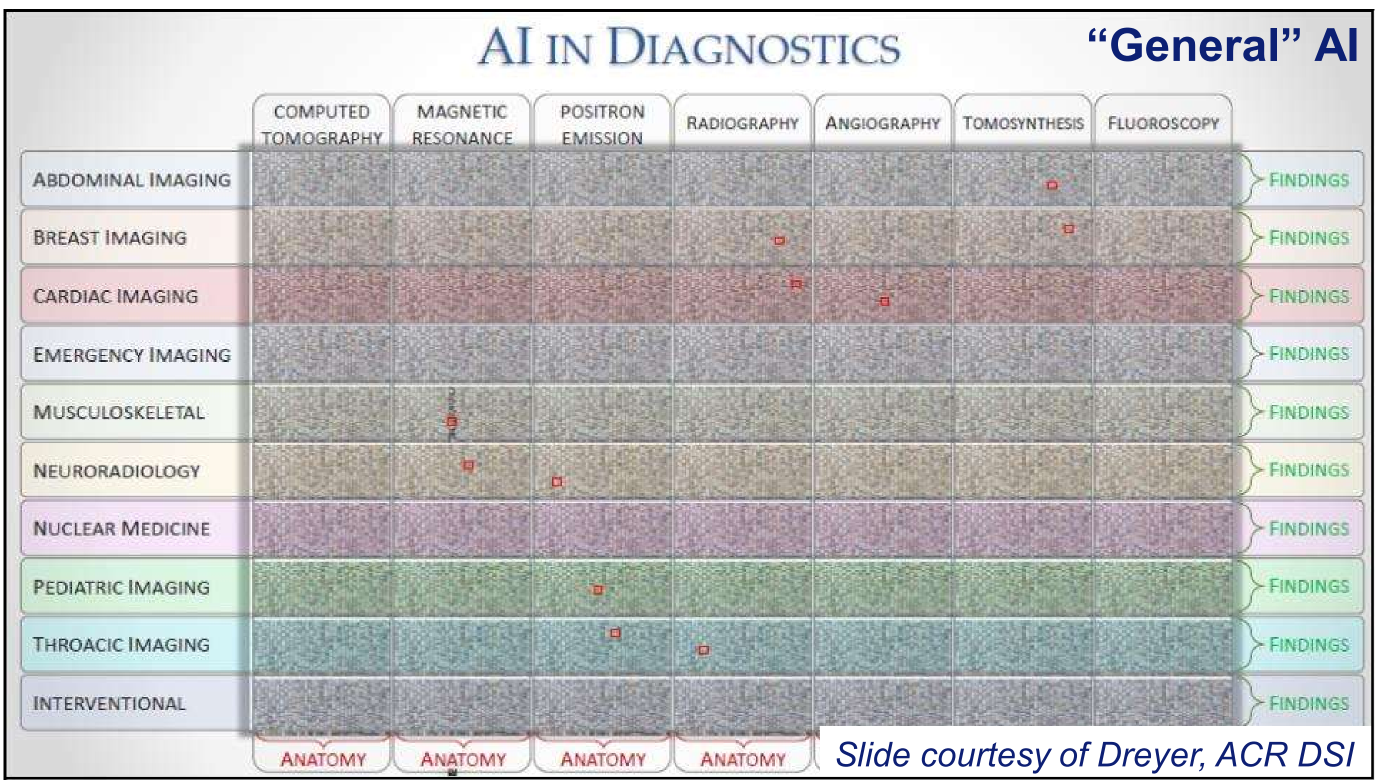
Will AI replace radiologists?

- Maybe, in some areas
- But:

116


AI IN DIAGNOSTICS
"Narrow" AI
COMPUTED TOMOGRAPHY
MAGNETIC RESONANCE
POSITRON EMISSION
RADIOGRAPHY
ANGIOGRAPHY
TOMOSYNTHESIS
FLUOROSCOPY
ABDOMINAL IMAGING
BREAST IMAGING
CARDIAC IMAGING
EMERGENCY IMAGING
MUSCULOSKELETAL
NEURORADIOLOGY
NUCLEAR MEDICINE
PEDIATRIC IMAGING
THROACIC IMAGING
INTERVENTIONAL
POSTERIOR CRUCIATE LIGAMENT TEAR
MRI OF THE KNEE
FINDINGS
ANATOMY
Slide courtesy of Dreyer, ACR DSI

117


AI IN DIAGNOSTICS
"General" AI
COMPUTED TOMOGRAPHY
MAGNETIC RESONANCE
POSITRON EMISSION
RADIOGRAPHY
ANGIOGRAPHY
TOMOSYNTHESIS
FLUOROSCOPY
ABDOMINAL IMAGING
BREAST IMAGING
CARDIAC IMAGING
EMERGENCY IMAGING
MUSCULOSKELETAL
NEURORADIOLOGY
NUCLEAR MEDICINE
PEDIATRIC IMAGING
THROACIC IMAGING
INTERVENTIONAL
FINDINGS
ANATOMY
Slide courtesy of Dreyer, ACR DSI

118

To conclude

- Machine learning/deep learning are very powerful techniques for (medical) image interpretation
- Versatile: can perform segmentation, detection, tissue characterization, diagnosis, synthesis, etc…
- Can automate repetitive tasks
- Can uncover new imaging biomarkers

- Typically require (a lot of) annotated and representative data
- Can not deal with situations very different from the training data
- Can be difficult to interpret

119

Thanks to…

Adria Perez
Andres Arias
Annegreet van Opbroek
Antonio Garcia-Uceda Juarez
Arna van Engelen
Deep Kayal
Dirk Poot
Fedde van der Lijn
Florian Dubost
Francesco Ciompi
Gijs van Tulder
Hakim Achterberg
Jens Petersen
Lauge Sørensen
Nora Baka
Pechin Lo
Raghavendra Selvan
Sepp de Raedt
Shuai Chen
Silas Ørting
Veronika Cheplygina
Vladlena Gorbunova
Zahra Sedghi Gamechi

Aasa Feragen
Jon Sporring
Mads Nielsen
Marco Loog
Stefan Klein
Theo van Walsum
Wiro Niessen

Aad van der Lugt
Arfan Ikram
Asger Dirksen
Harm Tiddens
Haseem Ashraf
Hieab Adams
Jolien Roos
Jesper Pedersen
Klaus Kofoed
Mathilde Winkler Wille
Meike Vernooij
Saher Shaker
Zaigham Saghir

Netherlands Organization for Scientific Research (NWO)
Netherlands Organisation for Health Research and Development
Center for Translational Molecular Medicine (CTMM)
Danish Strategic Research Council
Danish Council for Independent Research (DFF)
Innovative Medicines Initiative (IMI)
Astra Zeneca
COSMONiO
Quantib
Vertex

Questions?

marleen.debruijne@erasmusmc.nl

120