Nodules and mammography

Jan Kybic

2020

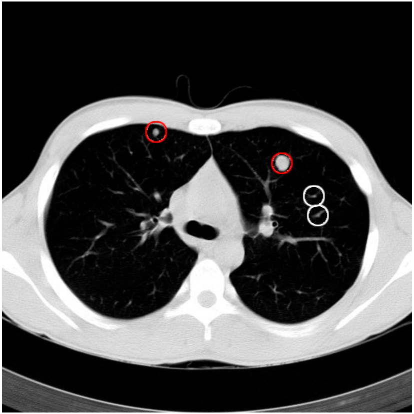
Murphy et al: A large-scale evaluation of automatic pulmonary nodule detection in chest CT using localfeatures and k-nearest-neighbor classification

Key points

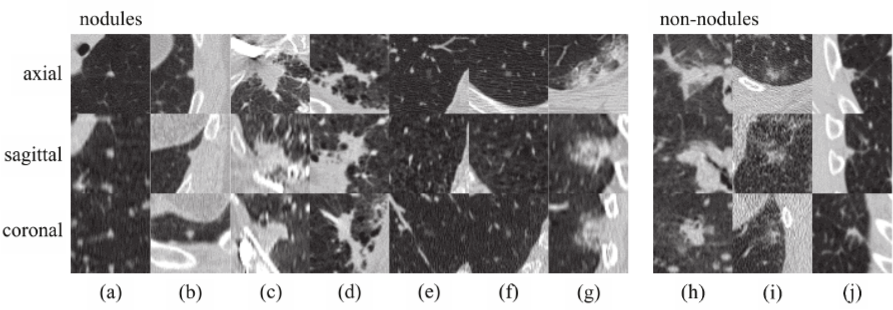
- Nodule (pre-cancer) detection
- Handcrafted features, simple classifier

Pulmonary nodules

- small bright spots on thoraci CT (often round but not always)
- mostly benign but some may lead to cancer
- earlier detection $\rightarrow$better prognosis



Nodule examples



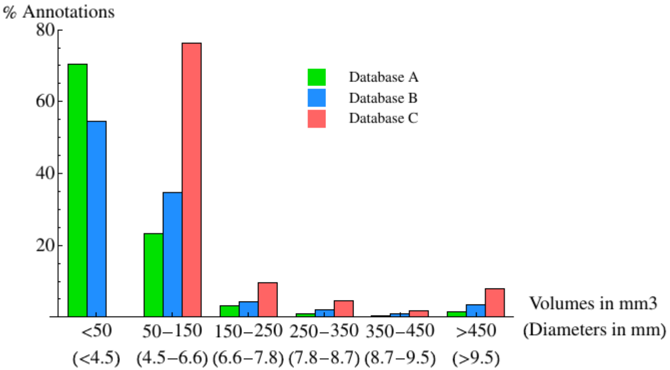
Nelson trial data

- 512×512pixels, $306 \sim 860$ slices
- two observers to mark nodules
- small nodules (diameter $< 3\,\text{mm}$) may not be marked
- if several scans per patient - the earliest chosen
- TP $=$ within 7 pixels
- Datasets: A - all scans, B - all scans with at least one big nodule, C - only big nodules

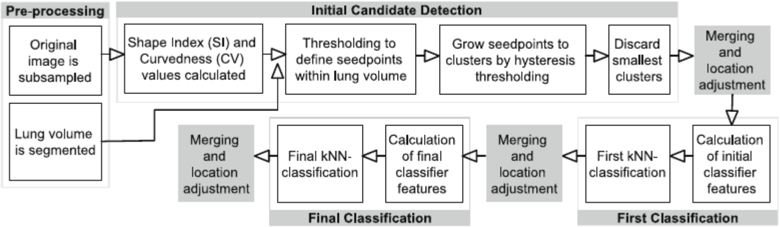
Table 1
Statistics on the number of scans in the three databases.

	A	B	C
#Scans before checks	1588	1158	1158
#Scans with lung segmentation failures	53	37	37
#Scans after removing failures	1535	1121	1121
#Scans in training set	722	580	580
#Scans in test set	813	541	541
#Nodules in final training set	1369	1763	760
#Nodules in final test set	1525	1688	768

Size distribution

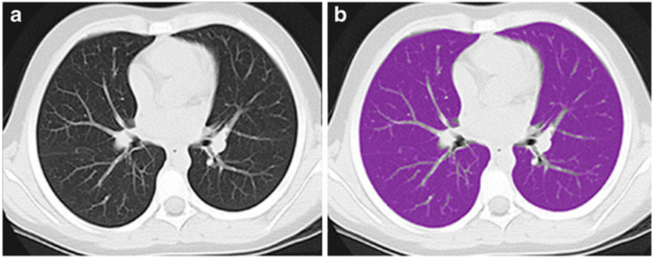


Flowchart


Pre-processing
Original image is subsampled
Lung volume is segmented
Initial Candidate Detection
Shape Index (SI) and Curvedness (CV) values calculated
Thresholding to define seedpoints within lung volume
Grow seedpoints to clusters by hysteresis thresholding
Discard smallest clusters
Merging and location adjustment
Calculation of initial classifier features
First kNN-classification
First Classification
Merging and location adjustment
Calculation of final classifier features
Final kNN-classification
Final Classification
Merging and location adjustment

Preprocessing

- downsampling to 256×256
- lung segmentation



from Jill Stein et al. DOI: 10.1007/s00247-016-3686-8

Shape index and curvedness

$$\mathrm{SI} = \frac{2}{\pi} \arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right)$$

$$\mathrm{CV} = \sqrt{k_1^2 + k_2^2}$$

Principal curvatures k_1,k_2

- minimum and maximum curvatures of the isosurface
- can be calculated from Hesssian with $\sigma = 1$

Shape index

- 1 local maximum = bright blob, 0.5 bright tubular structure, 0 saddle/flat...

Seed point detection

- Cluster formation

Table 2
Initial seed thresholds.

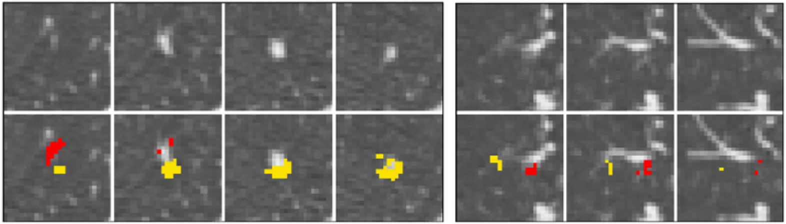
Value	Upper threshold	Lower threshold
SI	1	0.8 (near pleural surface) 0.9 (elsewhere)
CV	1	0.3

Table 3
Hysteresis thresholds.

Value	Upper threshold	Lower threshold
SI	1	0.7 (near pleural surface)
CV	1.3	0.2

- Cluster merging (distance $<$ voxels). Small objects ($<$15 voxels) discarded.
- Candidate location $=$ highest locally averaged intensity

Merging examples



(a) (b)

Successive axial slices. (a) TP, (b) FP.

False positive reduction

- Classify candidates
- *k*-NN classifier
- Two stages (15 and 50 features)
- Final stage on full resolution images
- Feature selection (Sequential forward floating selection)
- Operating point: sensitivity 90%

Training set generation

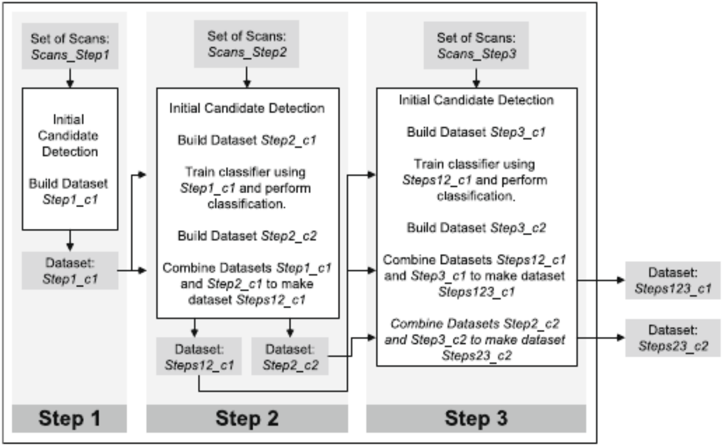


Fig. 4. Generation of training sets.

Subsample N class, N:P ratio 3:1, preserving pdf.

Features (1)

Table 4
The features calculated for the first kNN classifier. See text in Section 2.3.2.

ID	Description
Features of the voxel cluster	
a1	Cluster size (number of voxels)
a2	Compactness1, $\frac{ClusterSize}{(dim_x)(dim_y)(dim_z)}$
a3	Compactness2, $\frac{ClusterSize}{max\_dim^3}$
a4	Ratio *max_dim*:*min_dim*
a5	Ratio *max_dim*:*med_dim*
a6	Ratio A_{med}:A_{max} where A_{max}, A_{med} and A_{min} are the eigenvalues for the eigenvectors of the clu
a7	Ratio A_{min}:A_{max}
a8	Sphericity, $\frac{num\_cluster\_voxels\_in\_sphere\_S}{vol\_sphere\_S}$ where *sphere_S* is a sphere at the candidate location with r
a9	Ratio Sphericity:*r*
Features of voxels in spherical kernels at the candidate location	
a10-a18	On grey-values over spherical kernels *K*: Average, Median, Standard-Deviation

Features (2)

Table 5
The features calculated for the final kNN classifier. See text in Section 2.3.3

ID	Description	Notes
Features of the voxel cluster		
b1-b9	Features a1-a9 as described in Table 4	
b19	$min\_dim = min_i(dim_i)$	dim_i = width in di
b20	$max\_dim = max_i(dim_i)$	dim_i = width in di
Features of voxels in spherical kernels at the candidate location		
b10-b18	Features a10-a18 as described in Table 4	
b21-b26	On grey-values over spherical kernels *K*: Min, Max	Halfsizes of *K*: 1 (
b27-b36	On SI over spherical kernels *K*: Average, Median, Std-Dev, Min, Max	Halfsizes of *K*: 3 (
b37-b46	On CV over spherical kernels *K*: Average, Median, Std-Dev, Min, Max	Halfsizes of *K*: 3 (
Features calculated on randomly chosen points on a spherical surface around the candidate location.		
b47-b76	Features of Gradient orientation values: Average(Avg), Median, Max, Min, Std-Dev, Coefficient of Variation, Ratio Max:Min, Ratio Std-Dev:Median, Ratio Median:Avg, Ratio Median:Max	30 points on sphe b66), 50 points o
b77-b106	Features of Gradient magnitude values: Average(Avg), Median, Max, Min, Std-Dev, Coefficient of Variation, Ratio Max:Min, Ratio Std-Dev:Median, Ratio Median:Avg, Ratio Median:Max	30 points on sphe b96), 50 points on sphere
Features of voxels in the candidate segmentation		
b107-b115	Features a1-a9 as described in Table 4 but calculated this time over the segmented v	
b116	$min\_dim = min_i(dim_i)$	
b117	$max\_dim = max_i(dim_i)$	
b118-b122	On grey-values over segmented voxels: Average, Median, Std-Dev, Min, Max	
b123-b127	On SI of segmented voxels: Average, Median, Std-Dev, Min, Max	
b128-b132	On CV of segmented voxels: Average, Median, Std-Dev, Min, Max	
b133	Ratio Num segmented voxels: Num ROI voxels	
b134	Ratio (Distance from candidate location to the farthest point in the segmentation): (Number of voxels in the segmentation)	
Other features		
b135	Posterior probability of being a true nodule from the first classication step	

Results

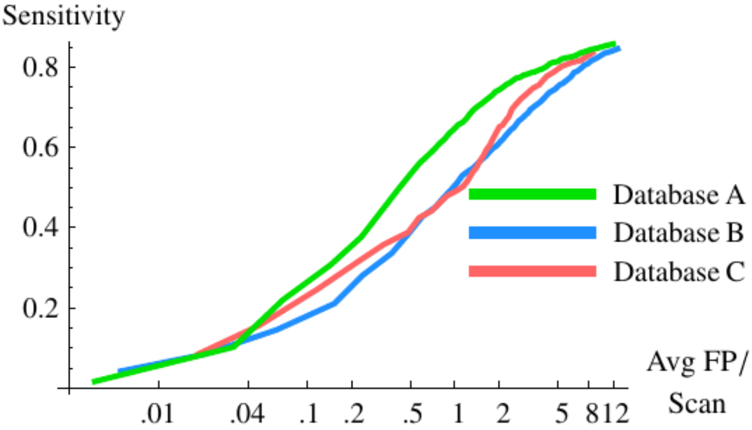
Table 7
Results for experiments on database A.

Number of Scans	813	
Number of annotations	1525	
	Sensitivity	FP per scan
After initial candidate detection	97.2%	649.0
After first classification	92.3%	77.3
After final classification		
– At around 4 FP per scan	80.0%	4.2

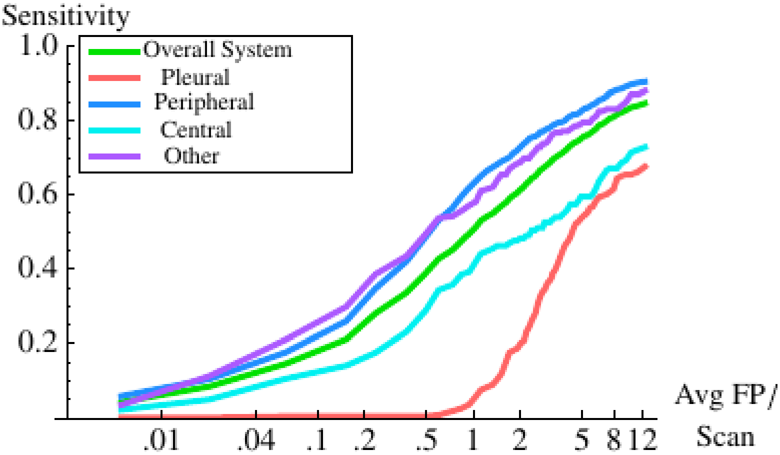
Table 9
Results for experiments on database C.

Number of scans	541	
Number of annotations	768	
	Sensitivity	FP per scan
After initial candidate detection	98.2%	752.1
After first classification	92.2%	51.2
After final classification		
– At around 4 FP per scan	77.7%	4.2

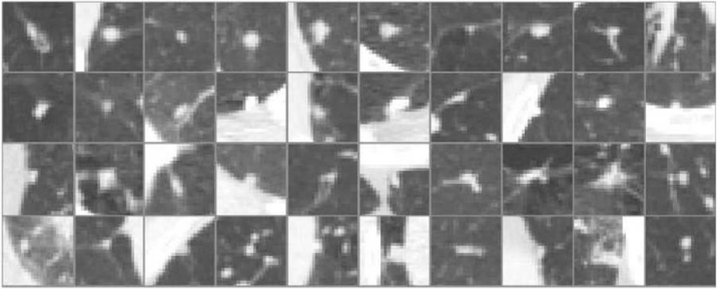
FROC curve


Sensitivity
0.8
0.6
0.4
0.2
.01
.04
.1
.2
.5
1
2
5
8
12
Avg FP/
Scan
Database A
Database B
Database C

FROC by location

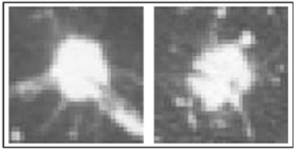


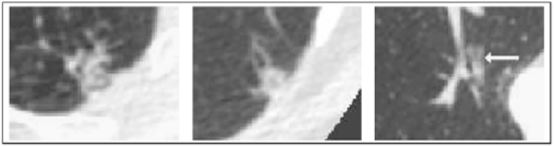
Example nodules



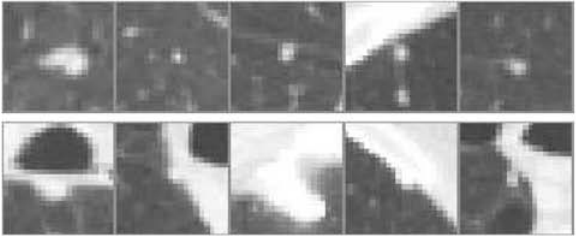
top row - easy detections ($p > 0.9$), bottom row - not detected ($p < 0.35$)

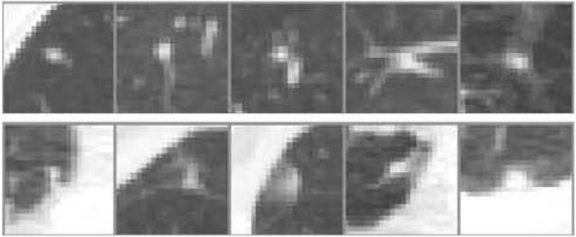
Missed nodules





False positives





Setio et al: Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. IEEE TMI 2016

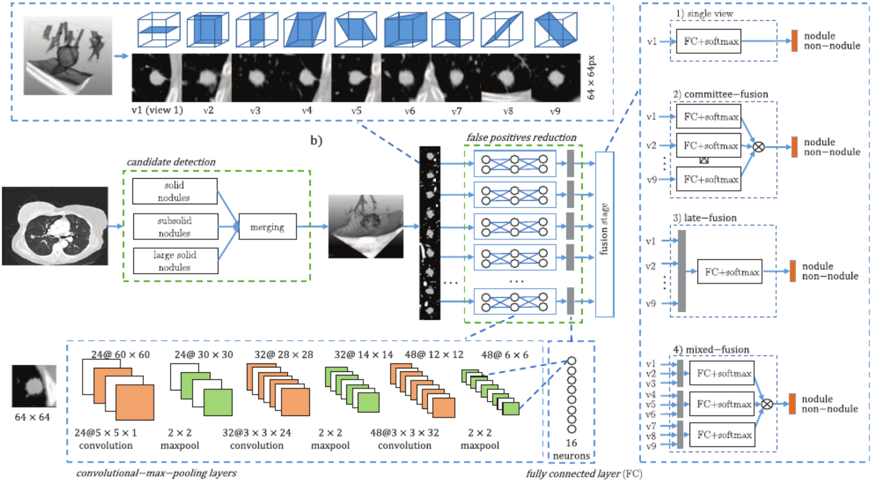
Key points

- Nodule detection from 3D CT
- Candidate detection (by 3 specialized detection)
- CNN for FP reduction
- 2D patches/planes + fusion

Datasets

- LIDC - 1018 scans, 888 retained (ignore thick-slice cases), 4 observers
- ANODE09 - 55 scans, 2 observers,
- DLCST - 612 scans, 2 observers, 898 nodules
- considered nodules $>$ 3mm

Flowchart



v1 (view 1)
v2
v3
v4
v5
v6
v7
v8
v9
64 × 64px
b)
candidate detection
solid nodules
subsolid nodules
large solid nodules
merging
false positives reduction
fusion stage
1) single view
FC+softmax
nodule
non−nodule
2) committee−fusion
3) late−fusion
4) mixed−fusion
24@ 60 × 60
24@ 30 × 30
32@ 28 × 28
32@ 14 × 14
48@ 12 × 12
48@ 6 × 6
64 × 64
24@5 × 5 × 1 convolution
2 × 2 maxpool
32@3 × 3 × 24 convolution
2 × 2 maxpool
48@3 × 3 × 32 convolution
2 × 2 maxpool
16 neurons
convolutional−max−pooling layers
fully connected layer (FC)

Candidate detection

- **Solid nodules** - Murphy's detector (shape index, curvedness, thresholding, clustering)
- **Subsolid nodules** (pure and part-solid ground-glass) - thresholding, morphological opening, connected components, segmentation
- **Large nodules** ($>$10mm, possibly attached to pleura) - lung segmentation, rolling-ball segmentation smoothing, density thresholding, multi-scale morphological opening

DETECTION SENSITIVITY OF CANDIDATE DETECTION ALGORITHMS

Total number of CT scans: 888

Total number of nodules: 1,186

Candidate detection	Detected nodules	Sensitivity (%)	False Positives (FPs)	FPs per scan
Solid	1,016	85.7	292,413	329.3
Subsolid	428	36.1	255,027	287.2
Large solid	377	31.8	41,816	47.1
Combined set	1,120	94.4	543,160	611.7
Reduced set	1,106	93.3	239,041	269.2

Patch classification

Patch extraction

- 50×50mm, 64×64pixels, nine planes

CNNs

- 3 convolutional layers, 3 max-pooling layers
- testing - 1 s per scan on a GPU

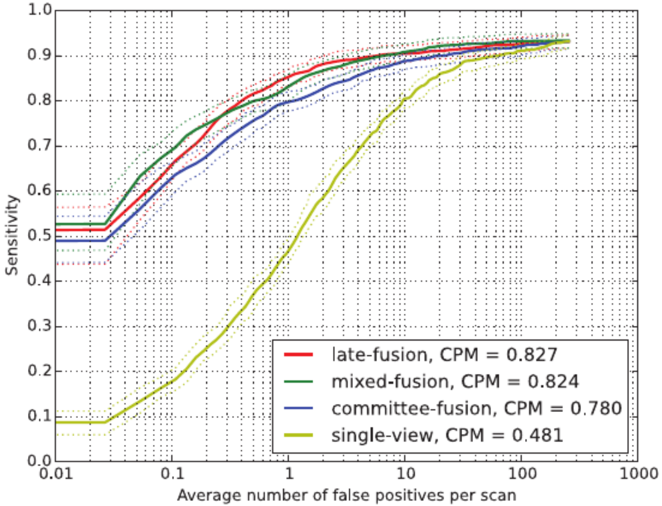
Fusion

- **Committee fusion**
  - FC layer + softmax + product rule
  - each stream trained separately
- **Late fusion**
  - concatenate FC layer outputs
  - FC layer + softmax
- **Mixed fusion**
  - group 9 patches into 3 groups of orthogonal views
  - contenate within group (as in late fusion)
  - FC layer + softmax + product rule (as in committee fusion)

Training

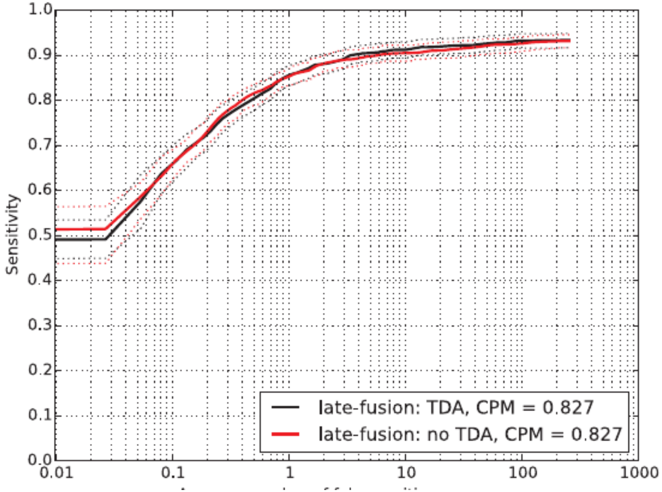
- Negative training data pruning
  - preliminary classification by existing algorithms
  - eliminate candidates with low nodule probability
- 5-fold cross-validation on LIDC (3/5 training, 1/5 validation, 1/5 testing)
- cross-entropy error
- RMSprop
- random initialization
- dropout regularization
- augmentation of nodules (shift, scaling)
- random upsampling of nodules for training
- test-data augmentation (scaling)

FROC fusion

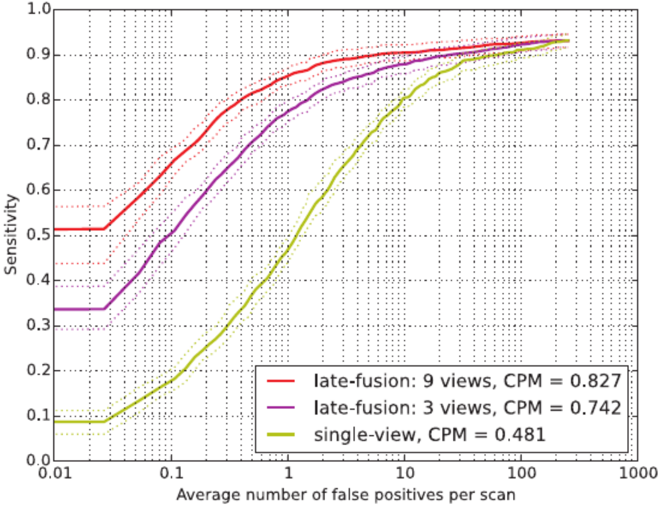


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Sensitivity
0.01
0.1
1
10
100
1000
Average number of false positives per scan
late-fusion, CPM = 0.827
mixed-fusion, CPM = 0.824
committee-fusion, CPM = 0.780
single-view, CPM = 0.481

FROC test augmentation



FROC number of views

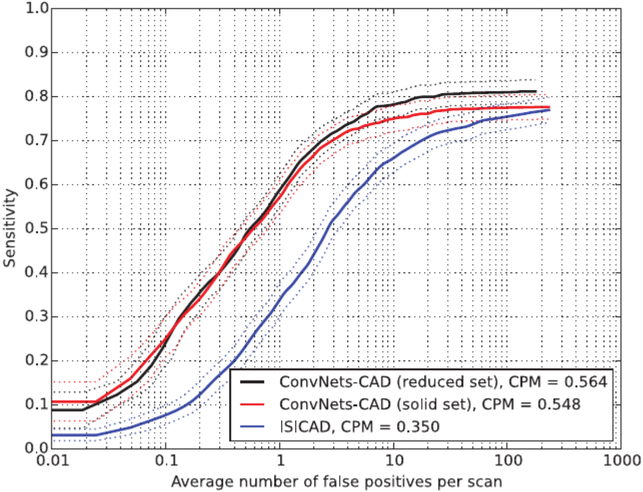


LIDC results

TABLE III
PERFORMANCE BENCHMARK OF CONVNETS CONFIGURATIONS ON LIDC-IDRI DATASET. THE BEST SCORE FOR EACH PERFORMANCE METRIC IS MARKED IN BOLD. FOR COMPARISON PURPOSES, THE PERFORMANCE OF THE COMBINED ALGORITHMS [3], [5], [27] IS INCLUDED

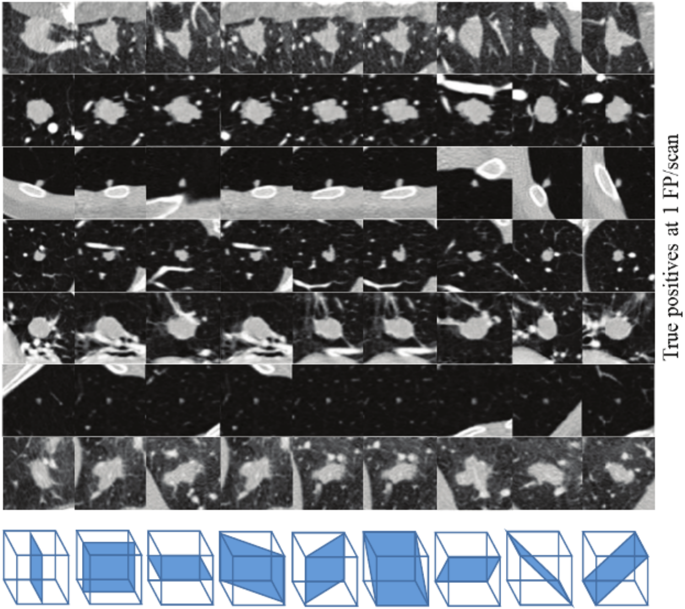
Configuration	Number of views	AUC	CPM
combined algorithms	-	0.969	0.573
single-view	1	0.969	0.481
committee-fusion	3	0.981	0.696
	9	0.987	0.780
late-fusion	3	0.987	0.742
	9	0.993	**0.827**
mixed-fusion	3*3	**0.996**	0.824

DLCST results

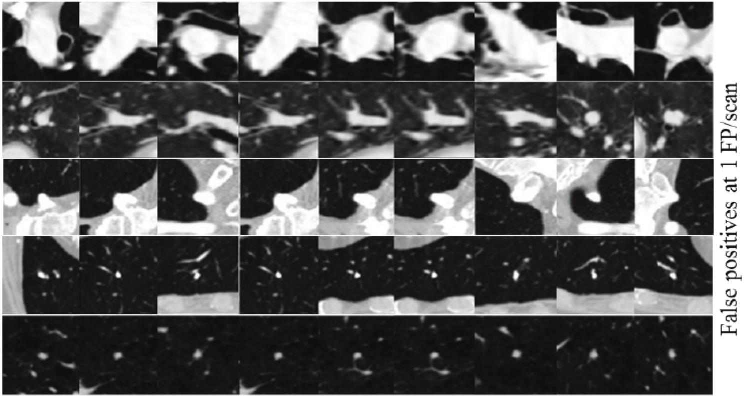


sensitivity 76.5% at 6 FPs/scan, which is 94% of the true candidate nodules

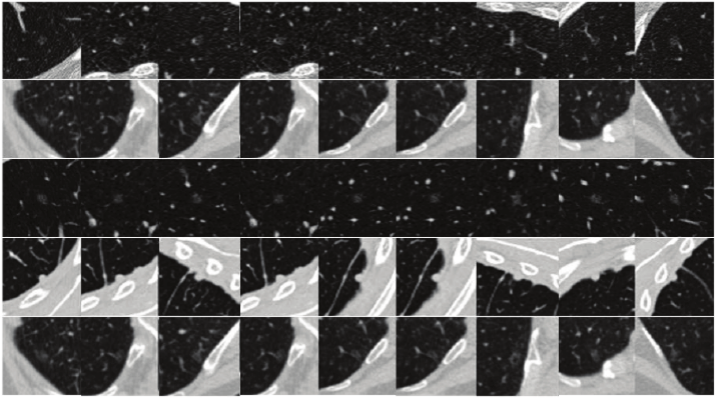
True positives



False positives



False negatives



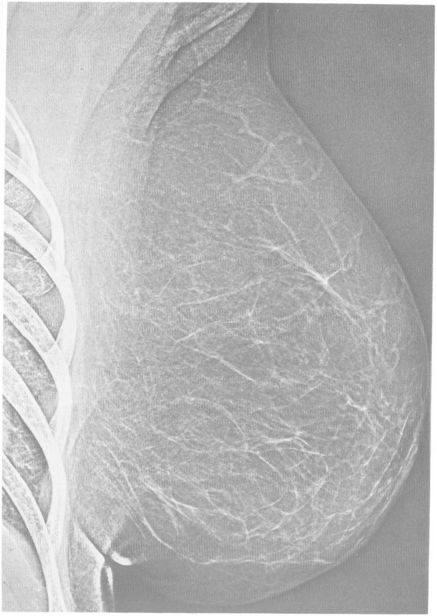
False negatives at 4 FPs/scan

Kooi: Large scale deep learning for computer aided detection of mammographic lesions. MIA 2017

Key points

- detect lesions from mammographs
- candidate detection learned
- classification to reduce FPs
- combine deep and manual features

Mammography



Data overview

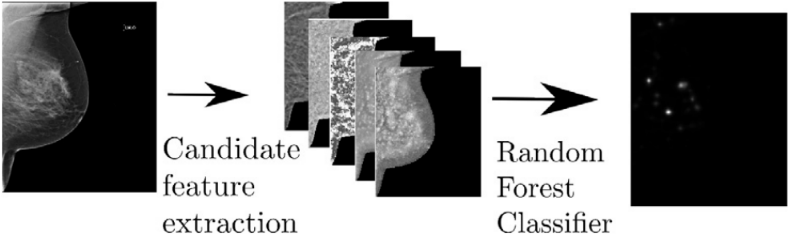
Table 1
Overview of the data. Pos refers to the amount of malignant lesions and neg to the amount of normals.

	Cases		Exams		Images		Candidates	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Train	296	6433	358	11,780	634	39,872	634	213,450
Valid.	35	710	42	1247	85	4218	85	19,460
Test	124	2064	124	5317	271	18,182	271	180,777

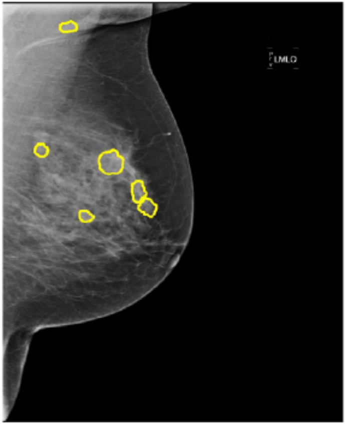
Candidate detection

- 5 features based on Gaussian derivative kernels
  - center of mass
  - size
  - spiculation (spikes or points)
- random forest classifier
- training data
  - positive samples from annotated lesions
  - negative samples randomly
- test time - apply RF to all pixels

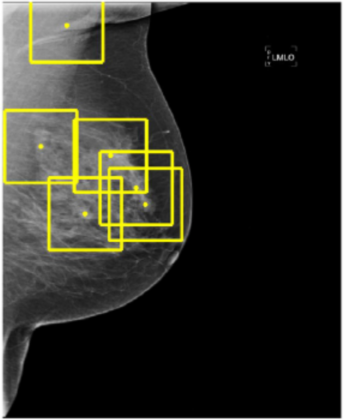
Candidate detection examples



Patches for CNN



(a) Illustration of segmentations for the reference system.



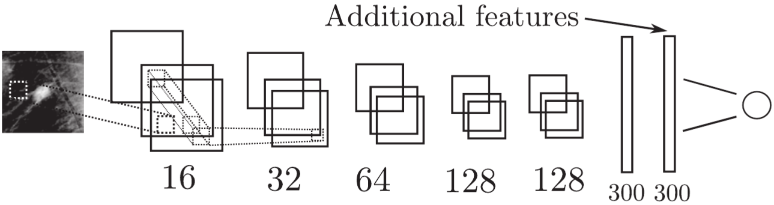
(b) Illustration of extracted patches for the CNN.

Baseline (classical system)

- mass segmentation by dynamic programming in polar coordinates
- 74 features:
  - candidate detector features,
  - contrast features,
  - texture features,
  - geometry features
  - context features (rest of the breast)
  - patient features
- RF classifier (also tested SVM, gradient boosted tree, MLPs)

CNN

- ReLU
- Binary cross-entropy loss
- Data augmentation (scale, translation, flip)
- scaled-down VGG model (6 layers with 3×3 kernels, 2×2 max-pooling), FC layer with 300 neurons
- learned features also extracted and a classifier trained
- positive samples randomly oversampled
- deep networks tried but did not improve the results



Feature importance

Table 3
Overview of results of the CNN combined with individual feature sets.

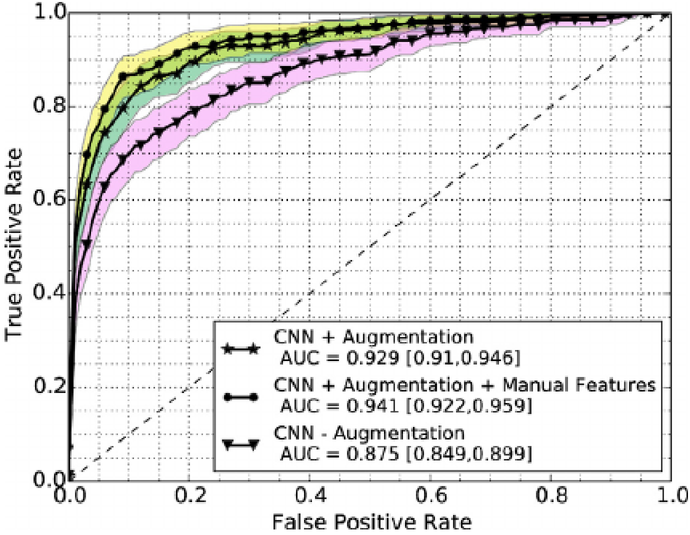
Feature group added to CNN	AUC	CI
CNN Only	0.929	[0.897, 0.938]
Candidate detector	0.938	[0.919, 0.955]
Contrast	0.931	[0.91, 0.949]
Texture	0.933	[0.912, 0.950]
Geometry	0.928	[0.907, 0.946]
Location	0.933	[0.913, 0.950]
Context	0.934	[0.914, 0.952]
Patient	0.929	[0.908, 0.947]
All	0.941	[0.922, 0.958]

Dataset size importance

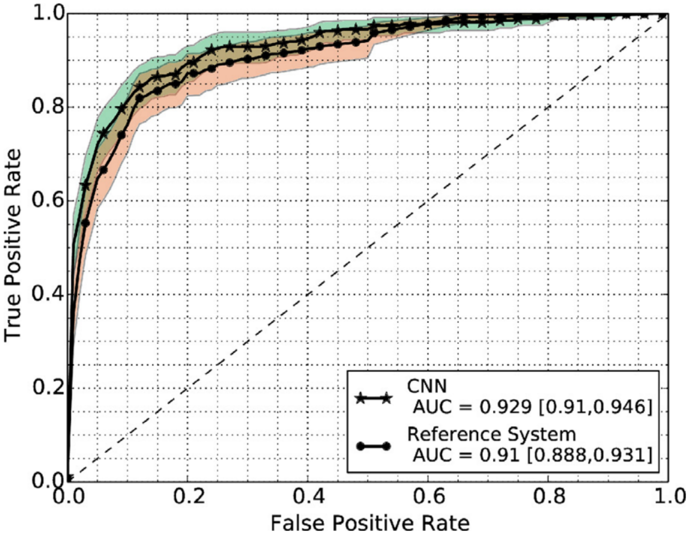
Table 4
AUC values obtained when training the network on subsets of malignant lesions in the training set, keeping the same amount of normals.

Data Augmentation	60%	All
With	0.842	0.929
Without	0.685	0.875

Augmentation importance



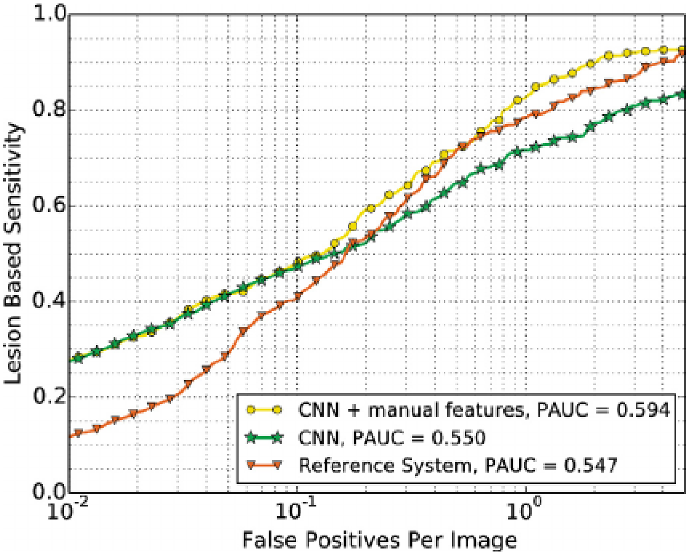
CNN versus baseline



excluding context, location, patient information

CNN versus baseline

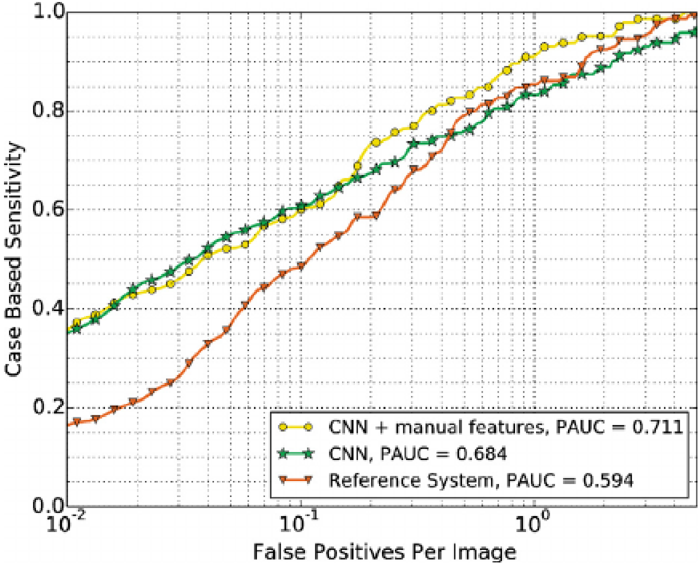
Lesion FROC



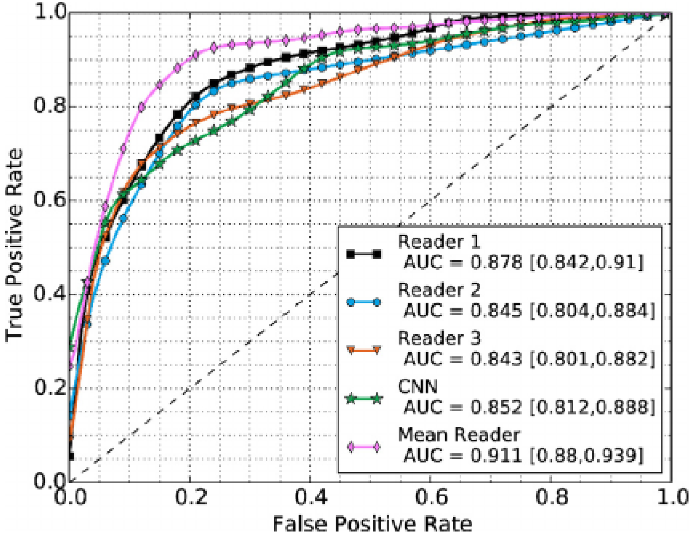
including context, location, patient information

CNN versus baseline

Case FROC

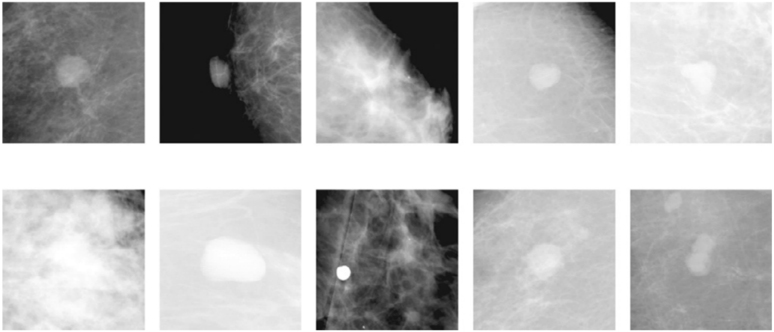


CNN versus human readers

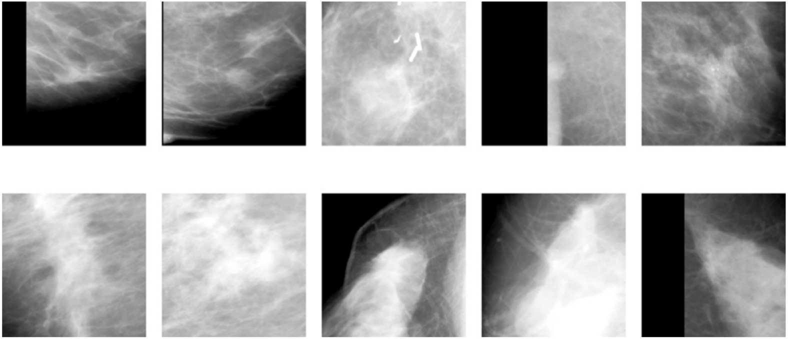


no significant difference between CNN and any readers
difference with mean of readers significant

False positives



False negatives



mostly very large lesions