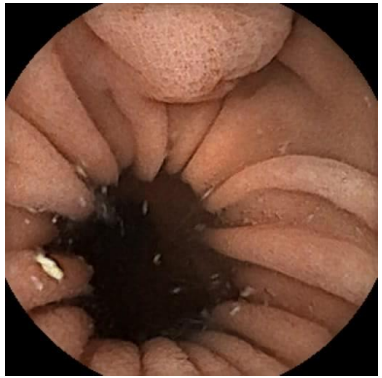
AI Vision for Endoscopy

Automated Image Analysis



Andrew Forrester

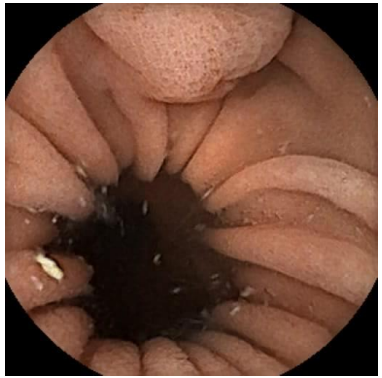
INSIGHT
AI Fellow

github.com/oneforawe/ai-vision-for-endoscopy

AI Vision for Endoscopy

Automated Image Analysis

☑



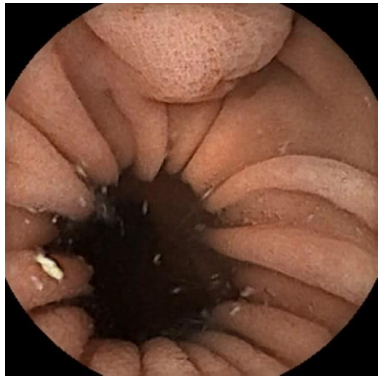
Andrew Forrester

INSIGHT
AI Fellow

github.com/oneforawe/ai-vision-for-endoscopy

AI Vision for Endoscopy

Automated Image Analysis



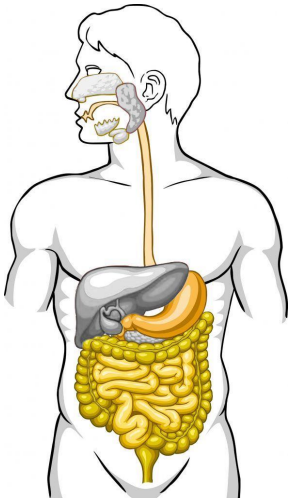
Andrew Forrester

INSIGHT
AI Fellow

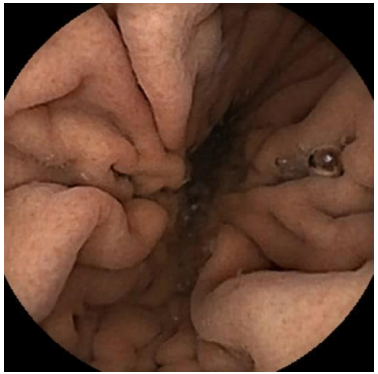
github.com/oneforawe/ai-vision-for-endoscopy

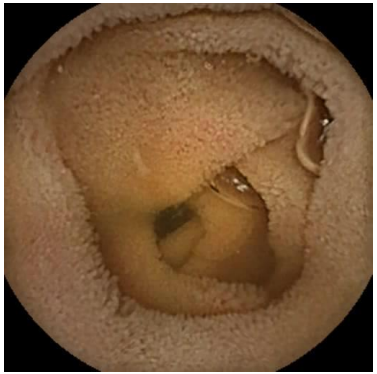
Context

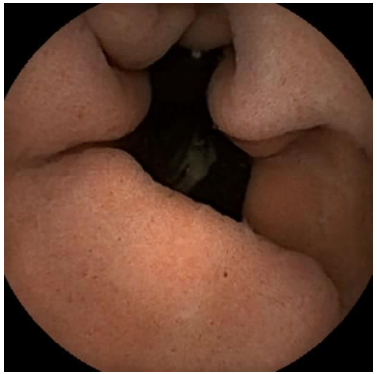
Capsule Endoscopy



(~2 FPS)·(~12 hours) = (~100k images)







Solution

Python package:

- framework for rapid iteration
- automated evaluation
- functional classifier (CNN)

Pipeline

Data

- Docbot
  raw pill images
  (140k labeled)

Processing

- auto-EDA
- group by patient
- auto-subset, partition
- auto crop and mask

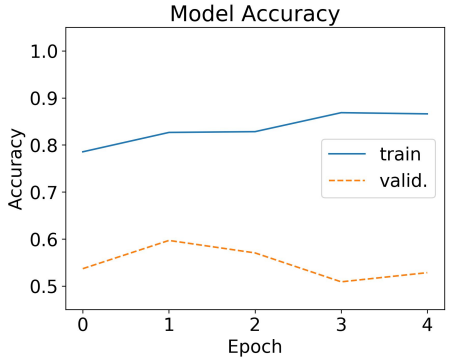
Models

- MobileNet,
  Xception, etc
  (transfer-learn)
- add layers,
  fine-tune,
  regularize
- augment

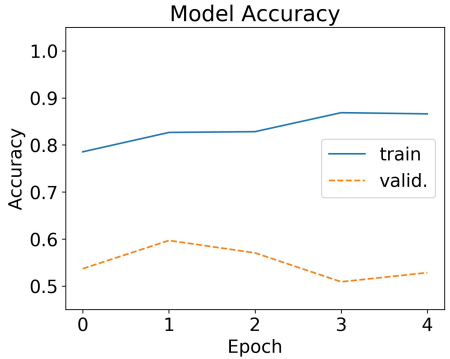
Analysis

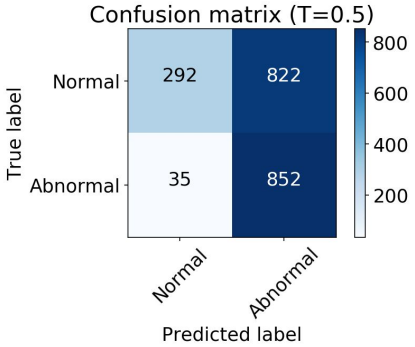
- auto-plots:
  training acc/loss
  ROC
  confusion matrix
- auto-metrics
  accuracy, etc
- threshold

Results


Model Accuracy
Accuracy
Epoch
train
valid.
1.0
0.9
0.8
0.7
0.6
0.5
0
1
2
3
4

Results

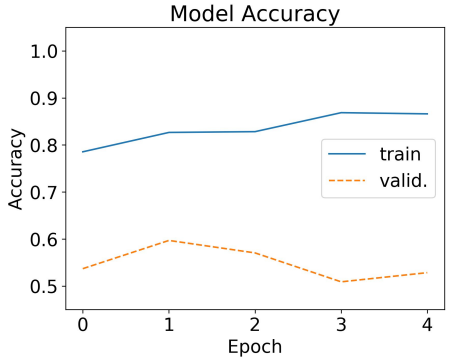


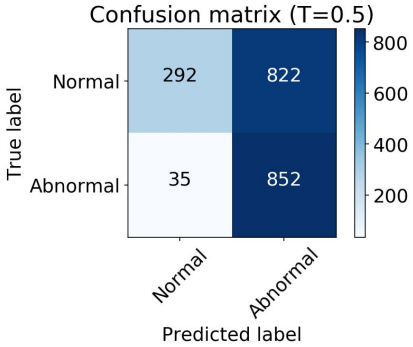


Accuracy

57%

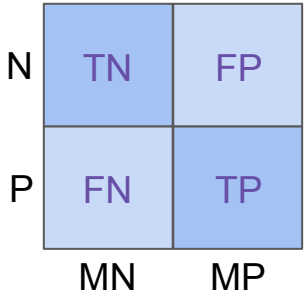
Results



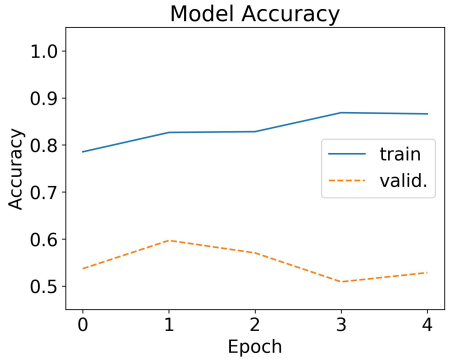


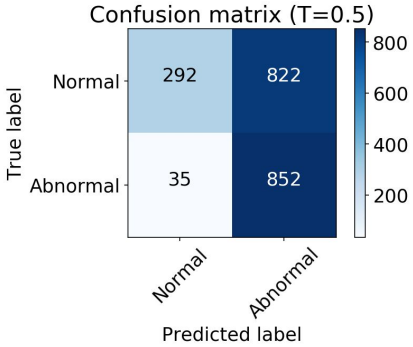
Accuracy

57%



Results



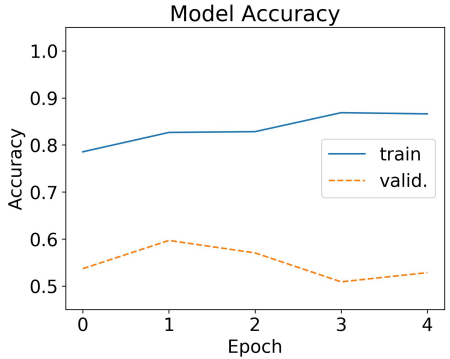


Accuracy

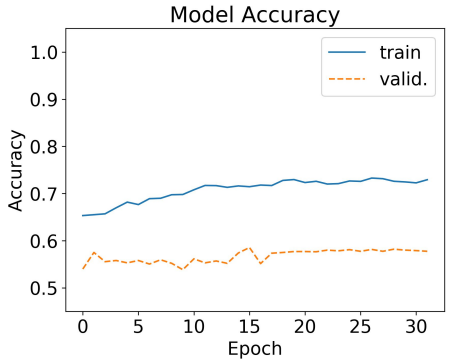
57%

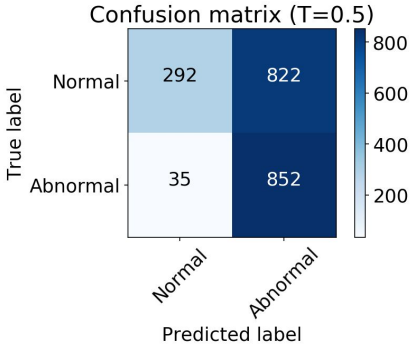
N	Saving time	Wasting time
P	Losing lives	Saving lives
	MN	MP

Results



More regularized



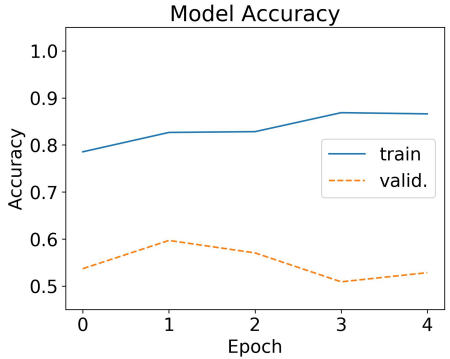


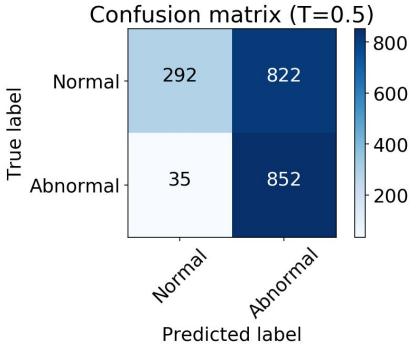
Accuracy

57%

N	Saving time	Wasting time
P	Losing lives	Saving lives
	MN	MP

Results



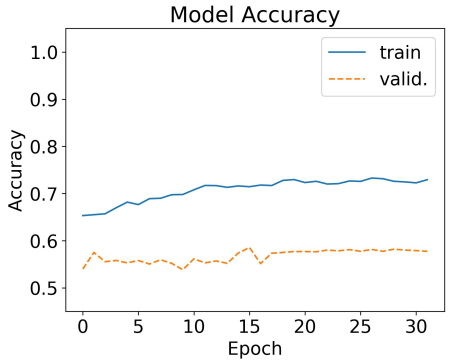


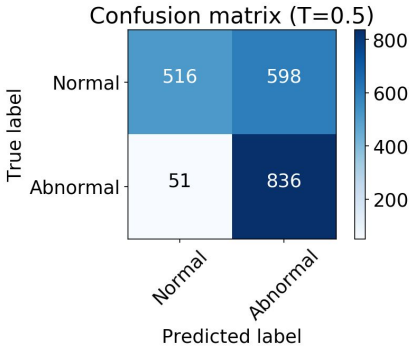
Accuracy

57%

	MN	MP
N	Saving time	Wasting time
P	Losing lives	Saving lives

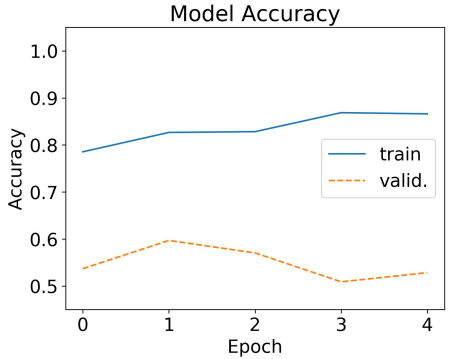
More regularized

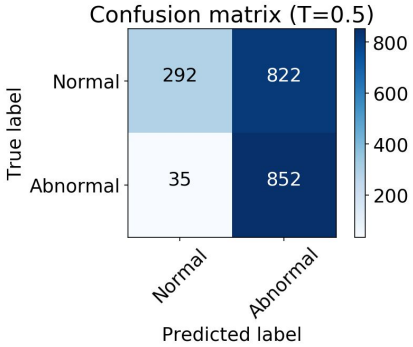




68%

Results



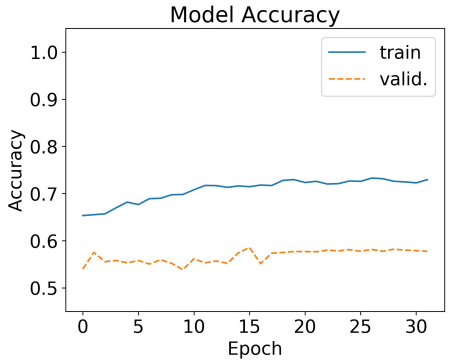


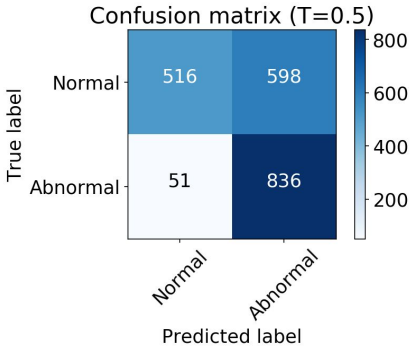
Accuracy

57%

N	Saving time	Wasting time
P	Losing lives	Saving lives
	MN	MP

More regularized

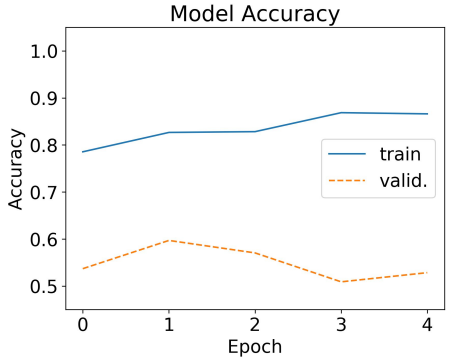


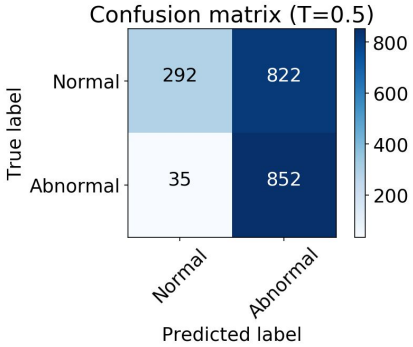


68%

Read time: -14%

Results



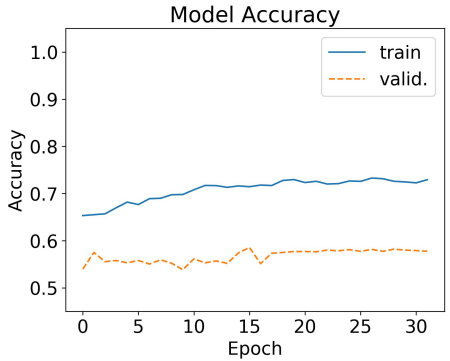


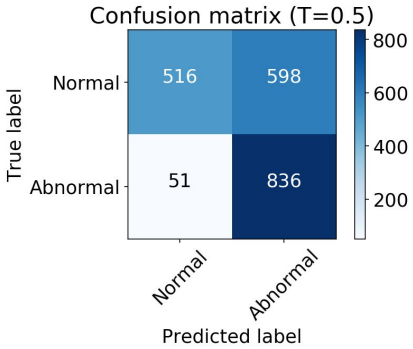
Accuracy

57%

N	Saving time	Wasting time
P	Losing lives	Saving lives
	MN	MP

More regularized





68%

Read time: -14%

Lives lost: +46%

Thank you!

Andrew Forrester

UCLA Physics PhD

AGi3 NLP & AGI

Interests: Philosophy in action
Society and its organization
Alternative education

Appendix

Lessons Learned

- Data
  - Quality of data / mis-labeling
  - Reorganize pipeline for raw data, file/folder name/structure
- Analysis
  - Progress ← verification, evaluation
  - Many experiments ← organization, automation
- Development
  - Balance custom dev. with usage of existing tools
  - Beware of “code creep”, re-evaluate strategy
  - Continuous refactoring, retooling
- Tools
  - Efficiency/automation requires an investment
  - Learning curve
    AWS EC2, S3; git-lfs; tmux

Next Steps

- Data
  - Classify by region, hierarchy of categories
  - Advise / quality control on labeling, raw pipeline
- Analysis
  - Further quantify training improvement
  - Systematize, automate hyperparameter tuning (AutoML, SigOpt?)
- Development
  - Explore existing tools
  - Refactor code, possibly re-tool

Next Steps

- Data
  - Classify by region, hierarchy of categories
  - Advise / quality control on labeling, raw pipeline
- Analysis
  - Further quantify training improvement
  - Systematize, automate hyperparameter tuning (AutoML, SigOpt?)
- Development
  - Explore existing tools
  - Refactor code, possibly re-tool

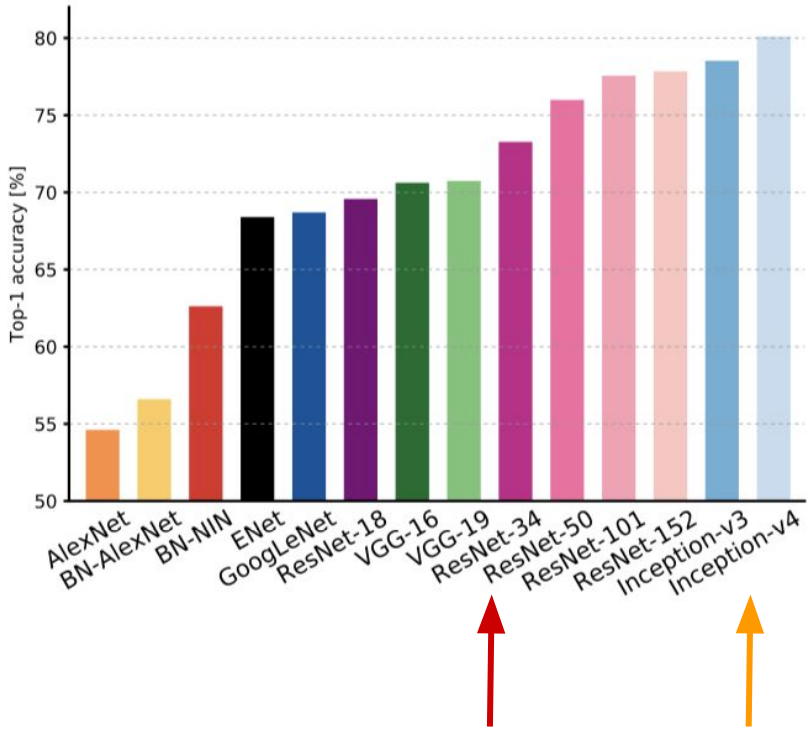
Object Detection / Segmentation / Recognition

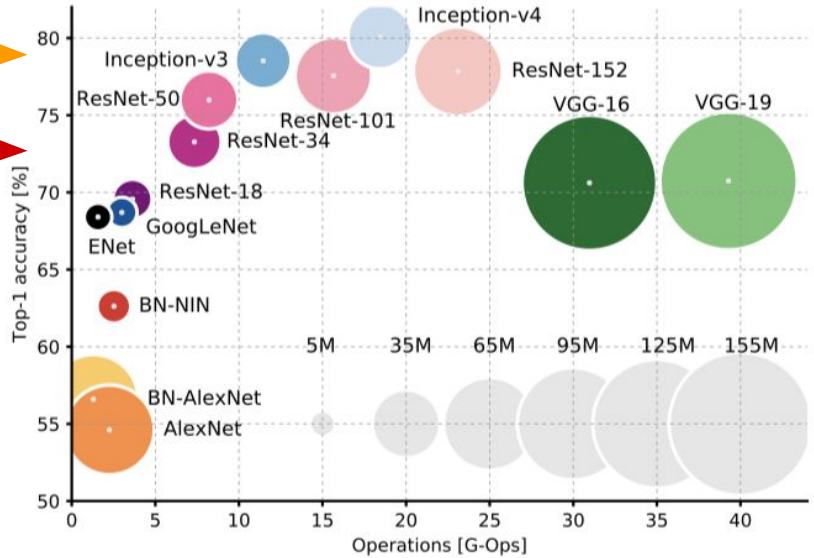
Models

Documentation for individual models

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.713	0.901	138,357,544	23
VGG19	549 MB	0.713	0.900	143,667,240	26
ResNet50	99 MB	0.749	0.921	25,636,712	168
InceptionV3	92 MB	0.779	0.937	23,851,784	159
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
MobileNet	16 MB	0.704	0.895	4,253,864	88
MobileNetV2	14 MB	0.713	0.901	3,538,984	88
DenseNet121	33 MB	0.750	0.923	8,062,504	121
DenseNet169	57 MB	0.762	0.932	14,307,880	169
DenseNet201	80 MB	0.773	0.936	20,242,984	201
NASNetMobile	23 MB	0.744	0.919	5,326,716	-
NASNetLarge	343 MB	0.825	0.960	88,949,818	-

Models





MobileNetV2

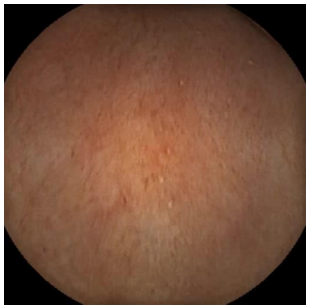
Xception

Data Processing Example

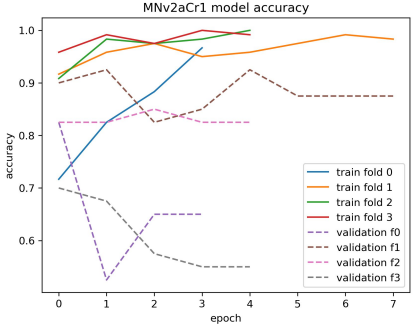
- Anonymize
- Decrease size of inputs / network
- Erase irrelevant noise

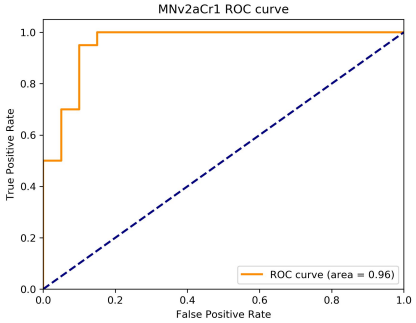
crop

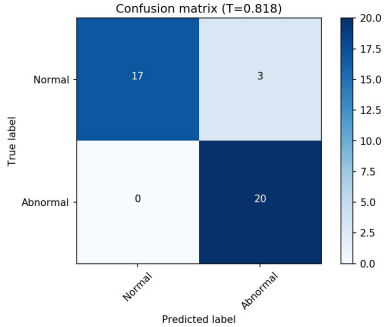
mask (green)

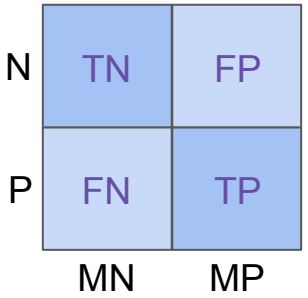


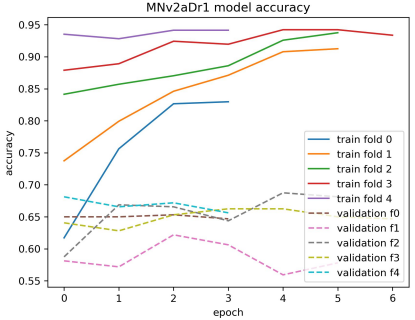
mask (black)

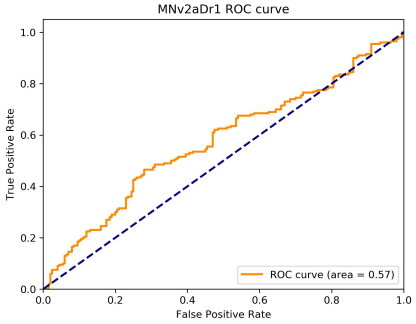

MNv2aCr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
validation f0
validation f1
validation f2
validation f3

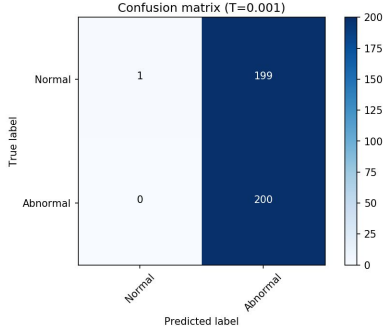

MNv2aCr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.96)

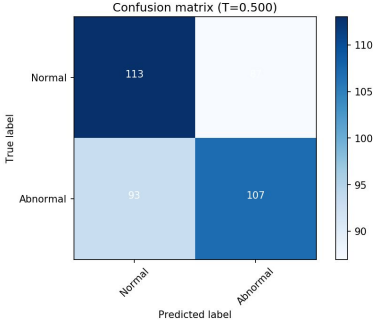

Confusion matrix (T=0.818)
Normal
Abnormal
True label
Predicted label
17
3
0
20

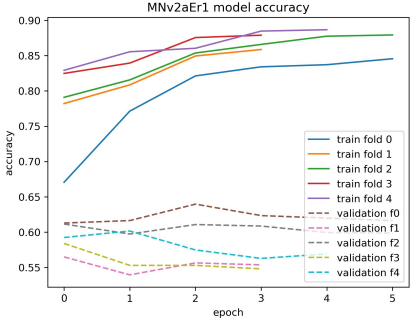

N
TN
FP
P
FN
TP
MN
MP

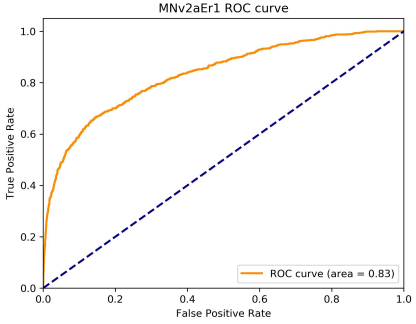

MNv2aDr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
train fold 4
validation f0
validation f1
validation f2
validation f3
validation f4

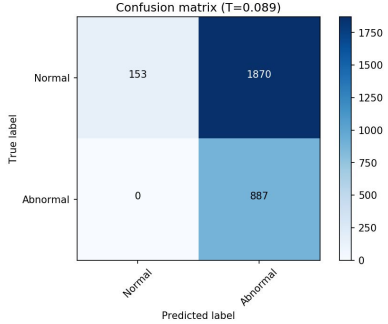

MNv2aDr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.57)

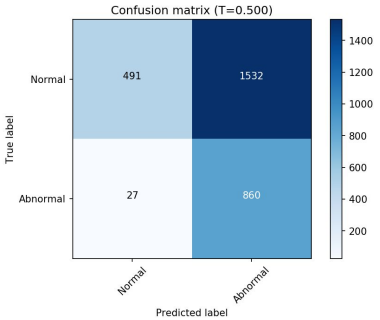

Confusion matrix (T=0.001)
Normal
Abnormal
True label
Predicted label
1
199
0
200

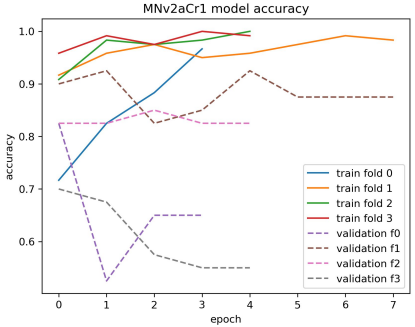

Confusion matrix (T=0.500)
Normal
Abnormal
True label
Predicted label
113
93
107

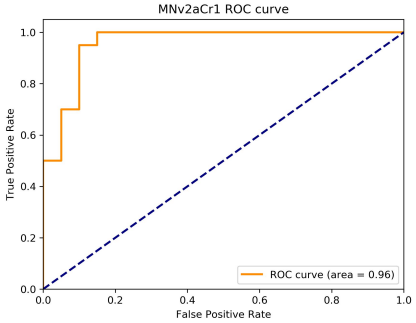

MNv2aEr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
train fold 4
validation f0
validation f1
validation f2
validation f3
validation f4

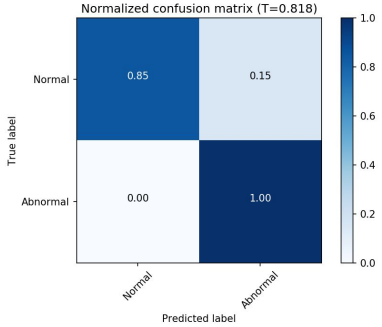

MNv2aEr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.83)

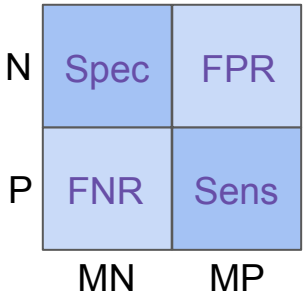

Confusion matrix (T=0.089)
Normal
Abnormal
True label
Predicted label
153
1870
0
887

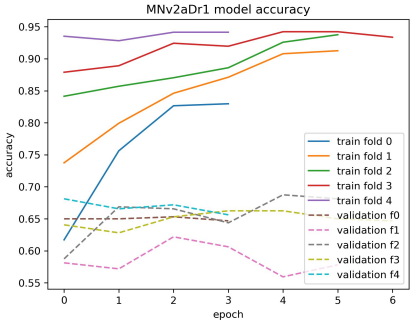

Confusion matrix (T=0.500)
Normal
Abnormal
True label
Predicted label
491
1532
27
860

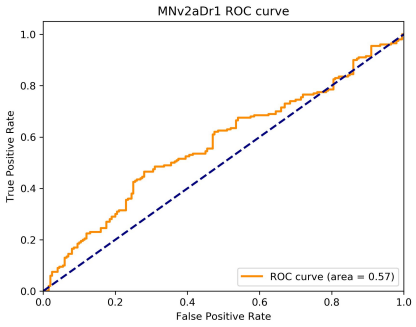

MNv2aCr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
validation f0
validation f1
validation f2
validation f3

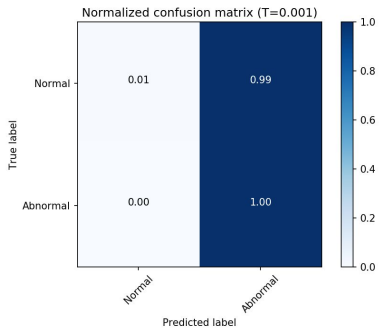

MNv2aCr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.96)

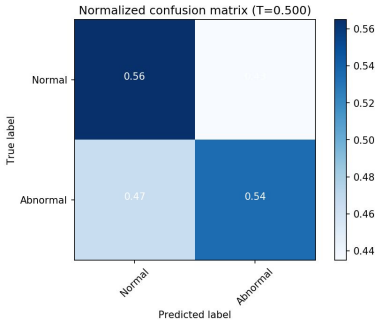

Normalized confusion matrix (T=0.818)
Normal
Abnormal
True label
Predicted label
0.85
0.15
0.00
1.00

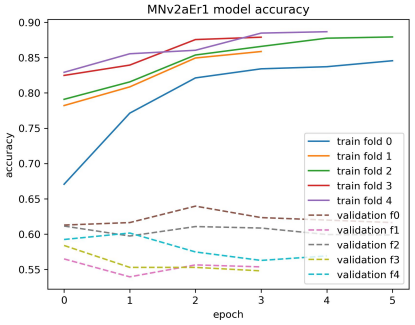

N
Spec
FPR
P
FNR
Sens
MN
MP

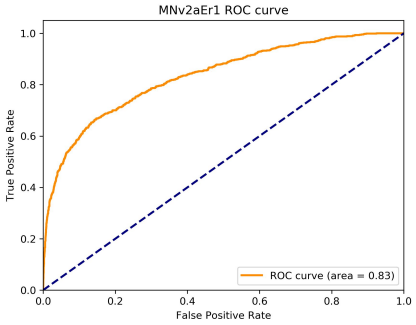

MNv2aDr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
train fold 4
validation f0
validation f1
validation f2
validation f3
validation f4

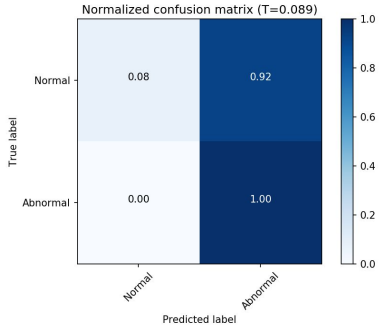

MNv2aDr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.57)

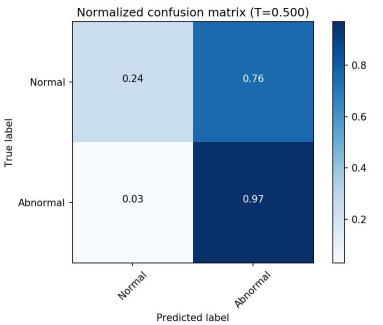

Normalized confusion matrix (T=0.001)
Normal
Abnormal
True label
Predicted label
0.01
0.99
0.00
1.00

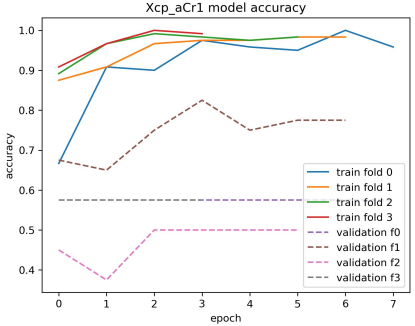

Normalized confusion matrix (T=0.500)
Normal
Abnormal
True label
Predicted label
0.56
0.47
0.54

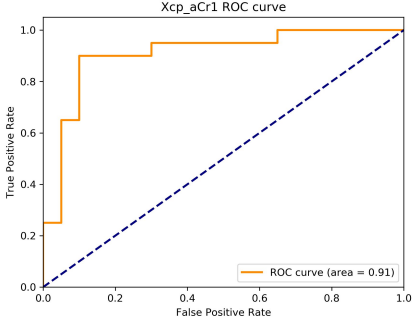

MNv2aEr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
train fold 4
validation f0
validation f1
validation f2
validation f3
validation f4

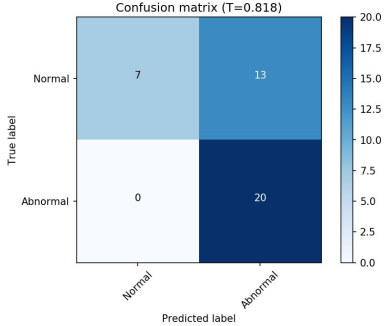

MNv2aEr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.83)

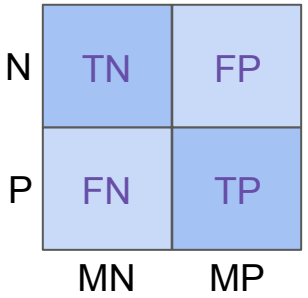

Normalized confusion matrix (T=0.089)
Normal
Abnormal
True label
Predicted label
0.08
0.92
0.00
1.00

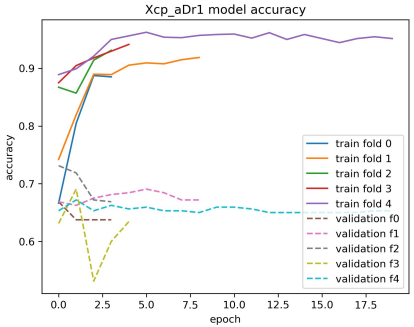

Normalized confusion matrix (T=0.500)
Normal
Abnormal
True label
Predicted label
0.24
0.76
0.03
0.97

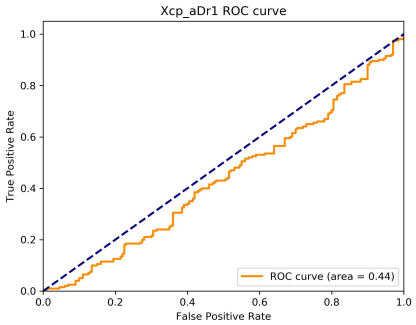

Xcp_aCr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
validation f0
validation f1
validation f2
validation f3

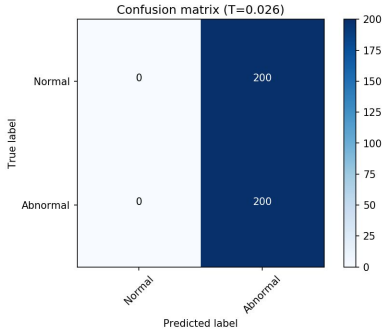

Xcp_aCr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.91)

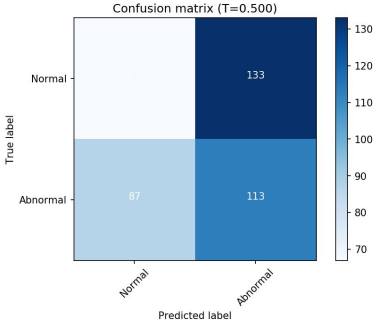

Confusion matrix (T=0.818)
Normal
Abnormal
True label
Predicted label
7
13
0
20

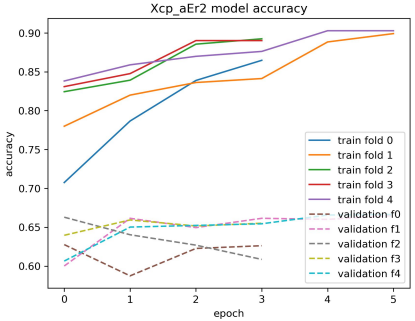

N
TN
FP
P
FN
TP
MN
MP

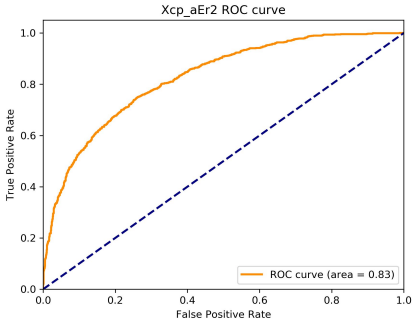

Xcp_aDr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
train fold 4
validation f0
validation f1
validation f2
validation f3
validation f4

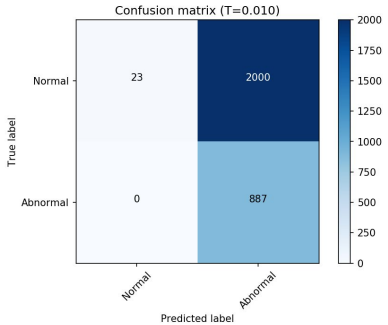

Xcp_aDr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.44)

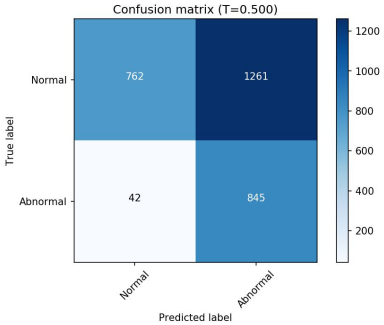

Confusion matrix (T=0.026)
Normal
Abnormal
True label
Predicted label
0
200
0
200

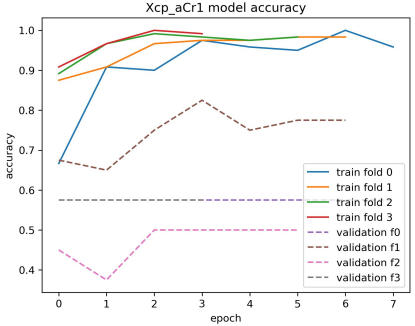

Confusion matrix (T=0.500)
Normal
Abnormal
True label
Predicted label
133
87
113

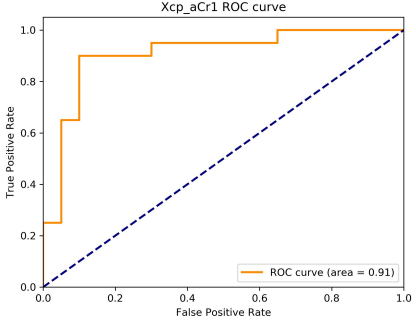

Xcp_aEr2 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
train fold 4
validation f0
validation f1
validation f2
validation f3
validation f4

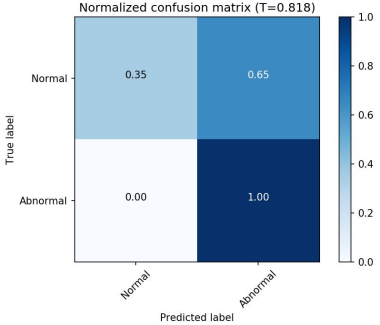

Xcp_aEr2 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.83)

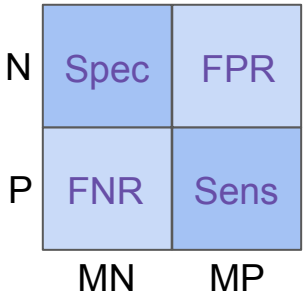

Confusion matrix (T=0.010)
Normal
Abnormal
True label
Predicted label
23
2000
0
887

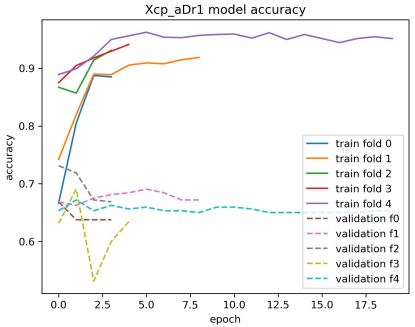

Confusion matrix (T=0.500)
Normal
Abnormal
True label
Predicted label
762
1261
42
845

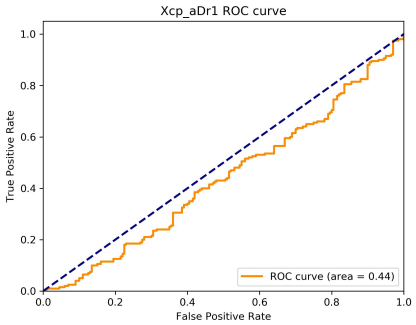

Xcp_aCr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
validation f0
validation f1
validation f2
validation f3

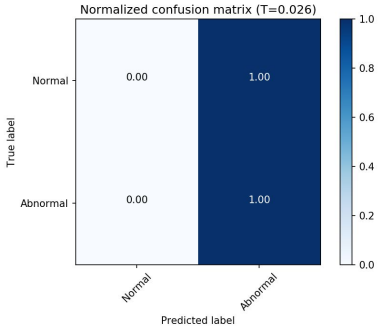

Xcp_aCr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.91)

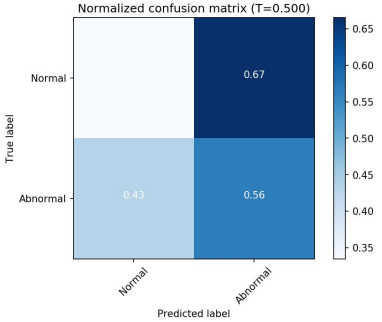

Normalized confusion matrix (T=0.818)
True label
Predicted label
Normal
Abnormal
0.35
0.65
0.00
1.00

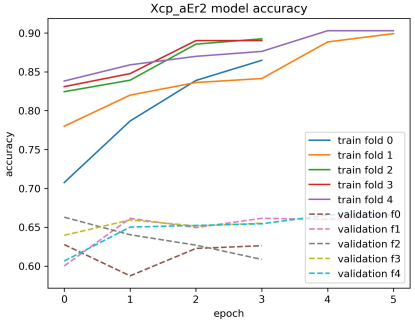

N
P
Spec
FPR
FNR
Sens
MN
MP

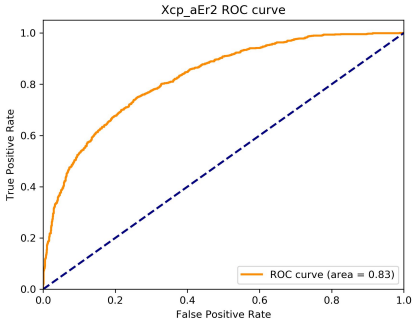

Xcp_aDr1 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
train fold 4
validation f0
validation f1
validation f2
validation f3
validation f4

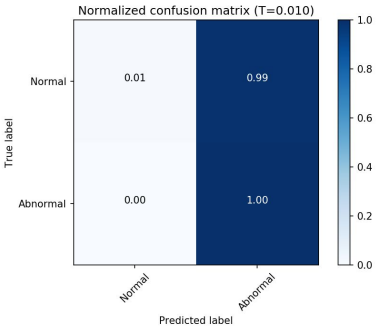

Xcp_aDr1 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.44)

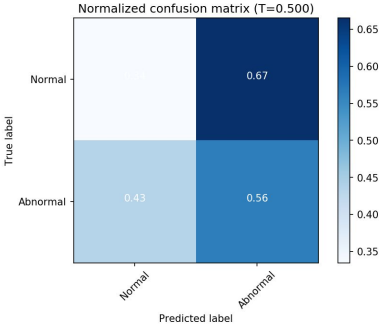

Normalized confusion matrix (T=0.026)
True label
Predicted label
Normal
Abnormal
0.00
1.00
0.00
1.00


Normalized confusion matrix (T=0.500)
True label
Predicted label
Normal
Abnormal
0.67
0.43
0.56


Xcp_aEr2 model accuracy
accuracy
epoch
train fold 0
train fold 1
train fold 2
train fold 3
train fold 4
validation f0
validation f1
validation f2
validation f3
validation f4

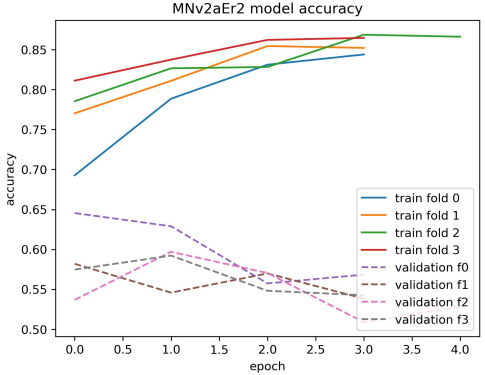

Xcp_aEr2 ROC curve
True Positive Rate
False Positive Rate
ROC curve (area = 0.83)

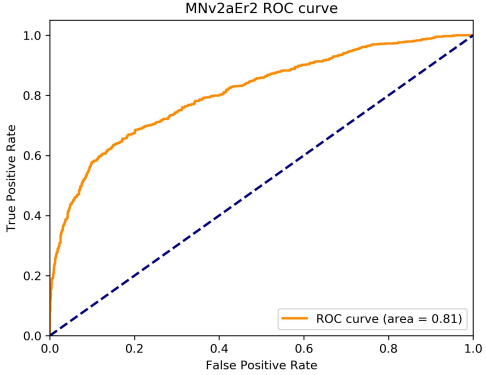

Normalized confusion matrix (T=0.010)
True label
Predicted label
Normal
Abnormal
0.01
0.99
0.00
1.00

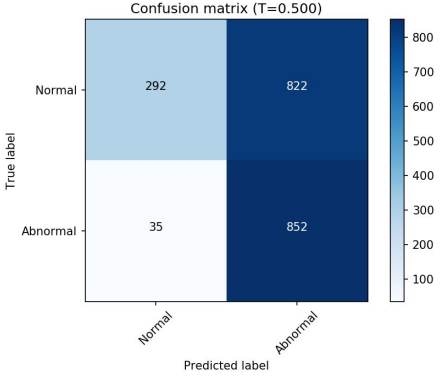

Normalized confusion matrix (T=0.500)
True label
Predicted label
Normal
Abnormal
0.67
0.43
0.56

MobileNetV2

a)





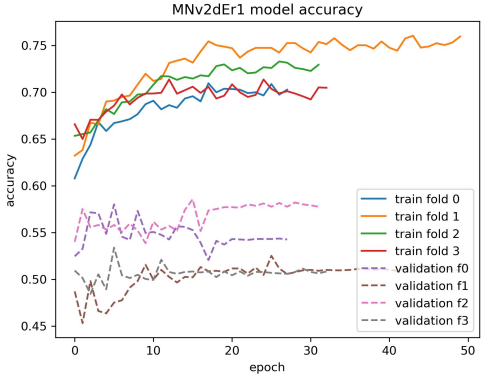


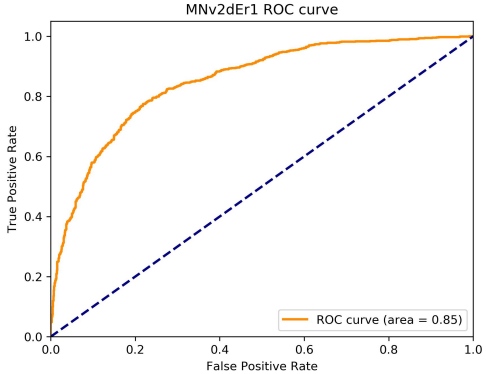
Accuracy

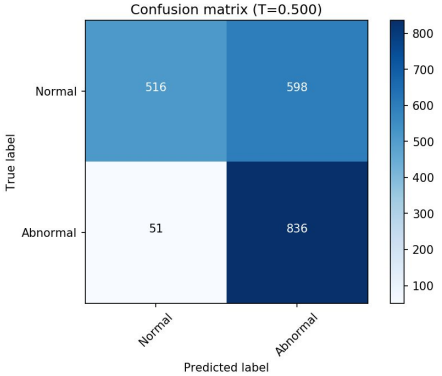
57%

More regularization

d)







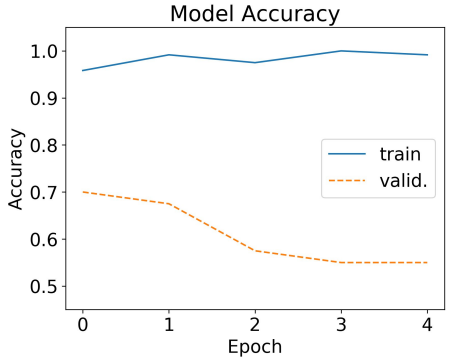
68%

		TRAIN						
MobileNet		Speed (imgs/sec)	Acc		Sens		Spec	
Xception		tot time	thresh1	thresh2	thresh1	thresh2	thresh1	thresh2
	C	0.9	93	80	100	100	85	60
cpu:		0.4	68	55	100	100	35	10
45 min		3 min						
2 hr		10 min						
	D	1.2	50	55	100	54	0.5	57
		0.2	50	45	100	57	0	34
		30 min						
		2 hr 15 min						
	E	1.4	36	46	100	97	8	24
		0.5	31	55	100	95	1	38
		2 hr						
		6 hr						
		INFER						
	C	13	95	70	100	100	90	40
		8	78	55	100	100	55	10
15 s		3 s						
45 s		5 s						
	D	40	50	50.5	100	52	0	49
		15	50	49	100	78	0	20
2 min 15 s		10 s						
7 min 30 s		30 s						
	E	59	30	45	100	96	0	22
		17	30	60	100	91	0	46
		50 s						
1 hr 50 min		3 min						

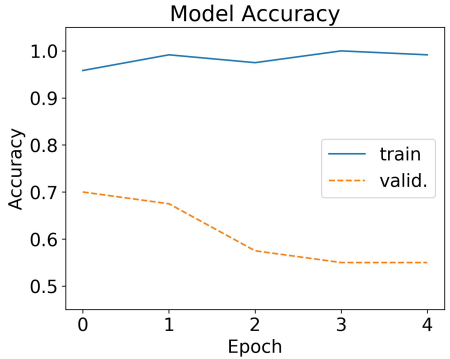
Results

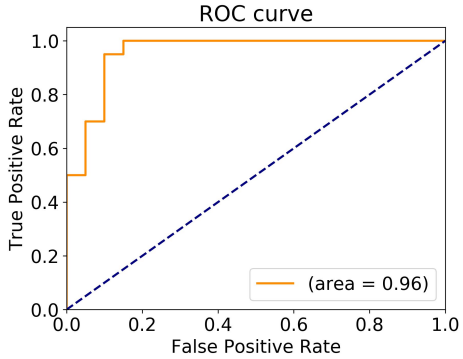
- MobileNetV2
- 200 images
- Colon: Normal vs Abnormal (blood)

Results

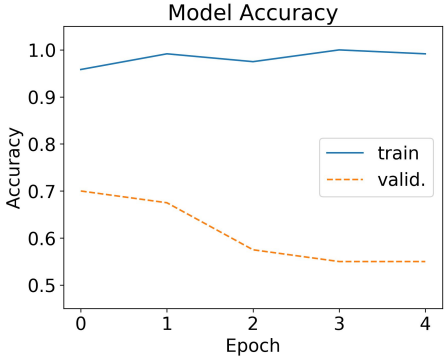

Model Accuracy
Accuracy
Epoch
train
valid.
1.0
0.9
0.8
0.7
0.6
0.5
0
1
2
3
4

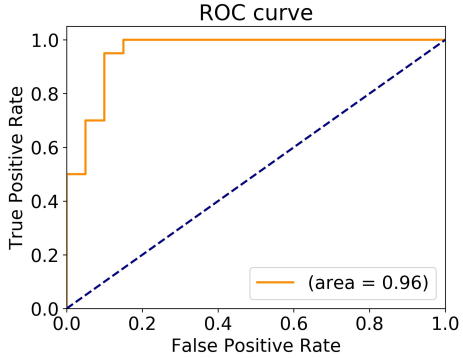
Results

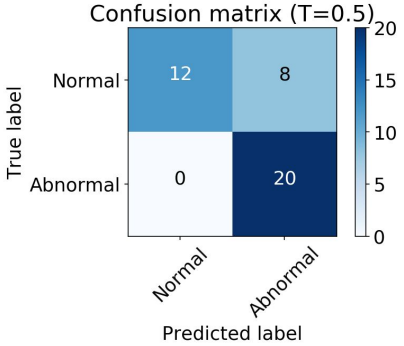

Model Accuracy
Accuracy
Epoch
train
valid.
1.0
0.9
0.8
0.7
0.6
0.5
0
1
2
3
4


ROC curve
True Positive Rate
False Positive Rate
(area = 0.96)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Results



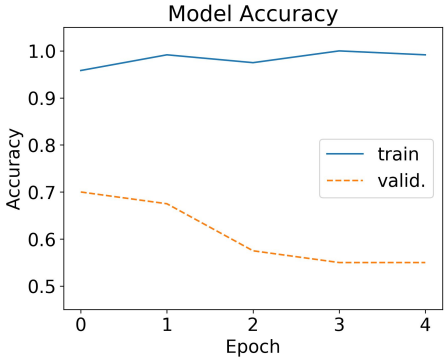


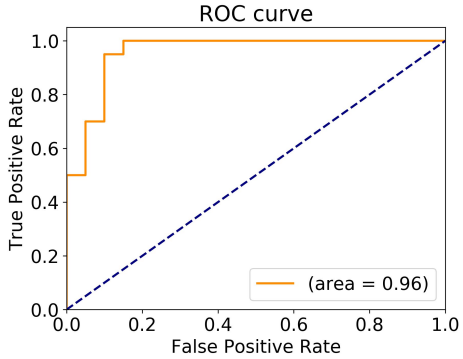


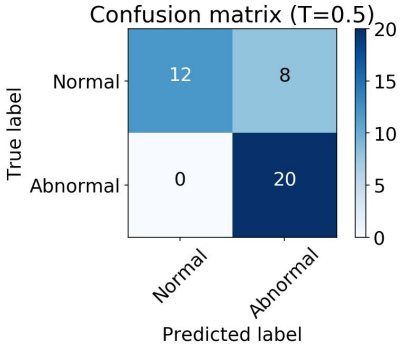
Accuracy

80%

Results

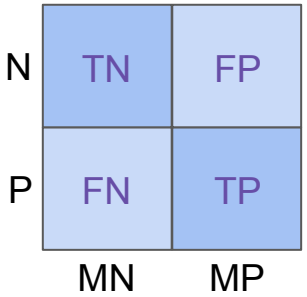




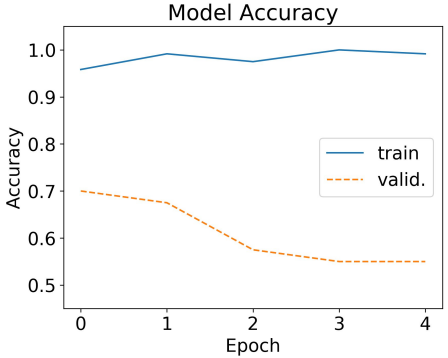


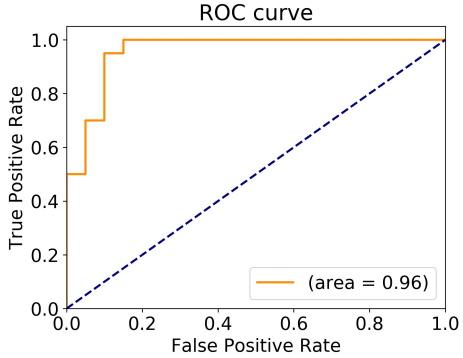
Accuracy

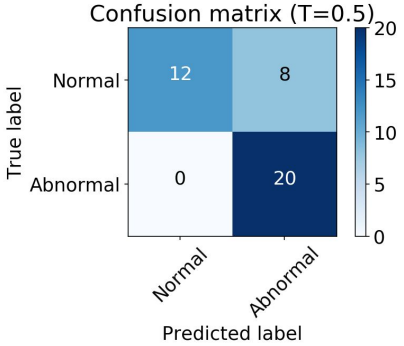
80%



Results



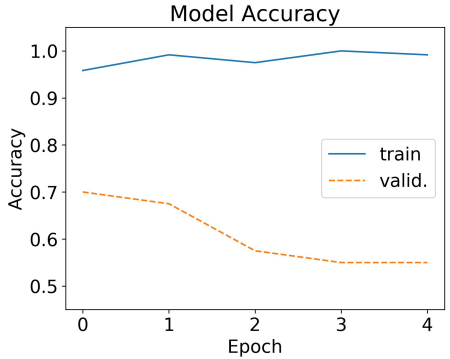


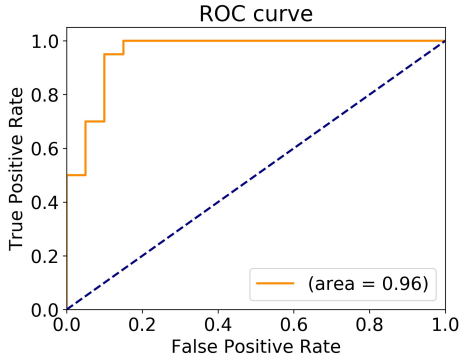


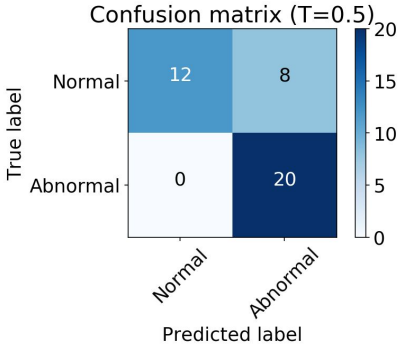
Accuracy

80%

Results







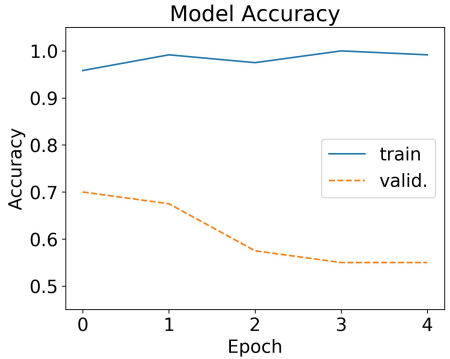
Accuracy

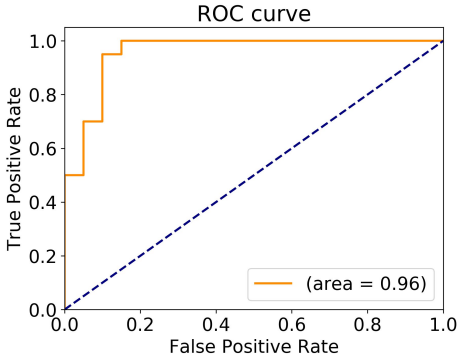
80%

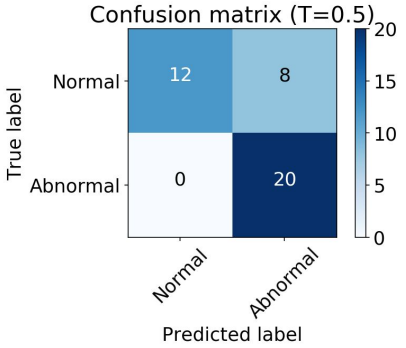
More regularized

Results

Accuracy

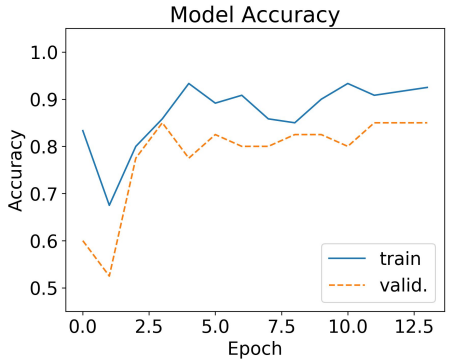






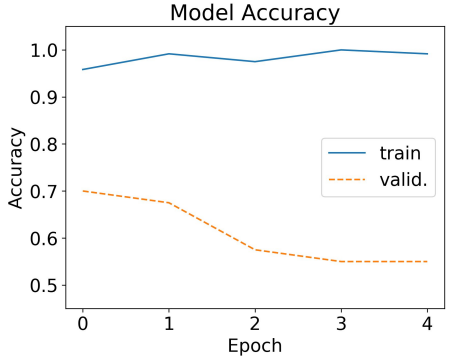
80%

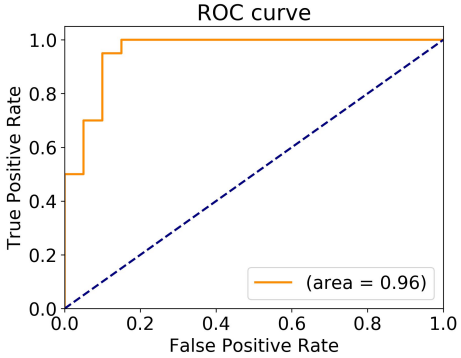
More regularized

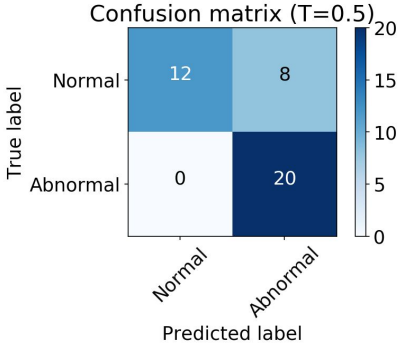


Results

Accuracy

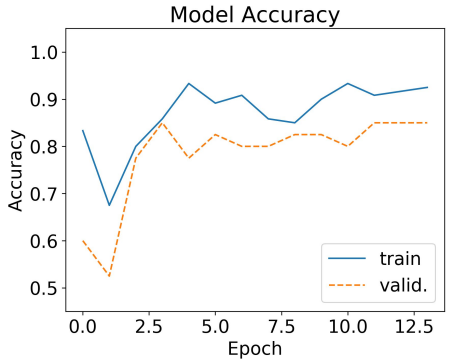


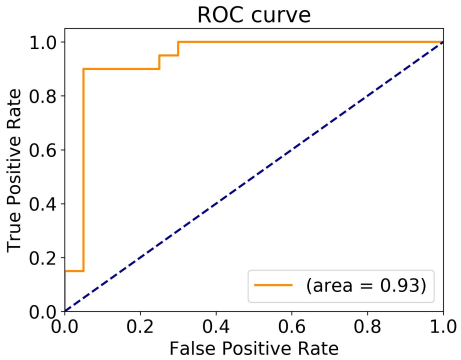


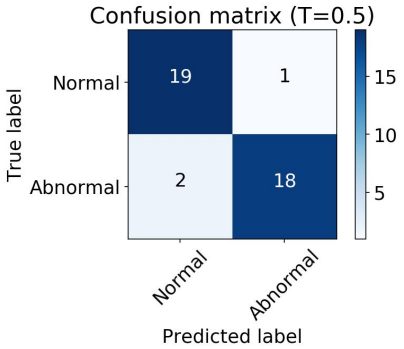


80%

More regularized







85%

Goal

Python package:

- abnormality classifier
- region classifier

Performance:

- higher accuracy (> 0.96 AUC)
- need high sensitivity (few false negatives)
- want high specificity (few false positives)
- want high speed