Chest X-Ray Tuberculosis Identification with Deep Learning

CS5242 Group 16

Ye Pingxiao, Xiong Wei, Bai Yaguang, Wang Chen

About The Study

Background

- Tuberculosis (TB) can be identified through inspection of chest X-Ray images.
- For example, on a chest X-ray from someone with TB, the cavitation that the TB bacteria form in the lung tissue can often be observed.

Objective

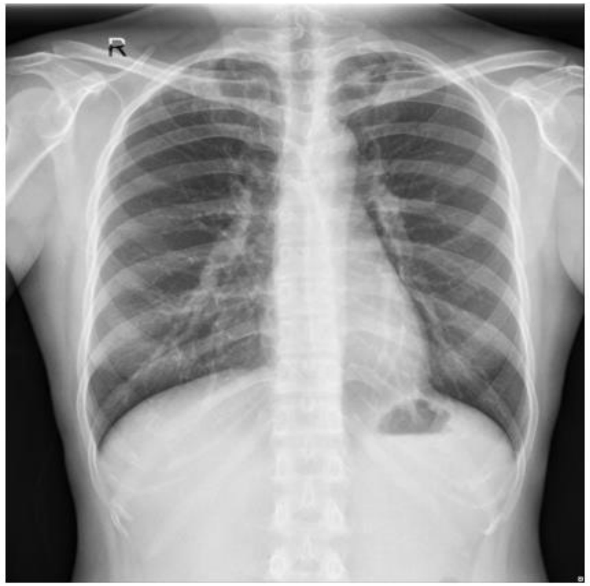
- We want to train a neural network that is capable of analyzing the chest x-ray images and identifying TB just like a radiologist.
- It will take an image as an input and output whether the patient has contracted TB.

Source: https://emedicine.medscape.com/article/358610-overview

About The Dataset

- Chest X-ray images:
  - 662 from Shenzhen
  - 138 from Montgomery
- 800 clinical readings
- 704 masks
- 394 TB positive + 406 negative
- 725 male + 74 female + 1 unclassified

TB Positive

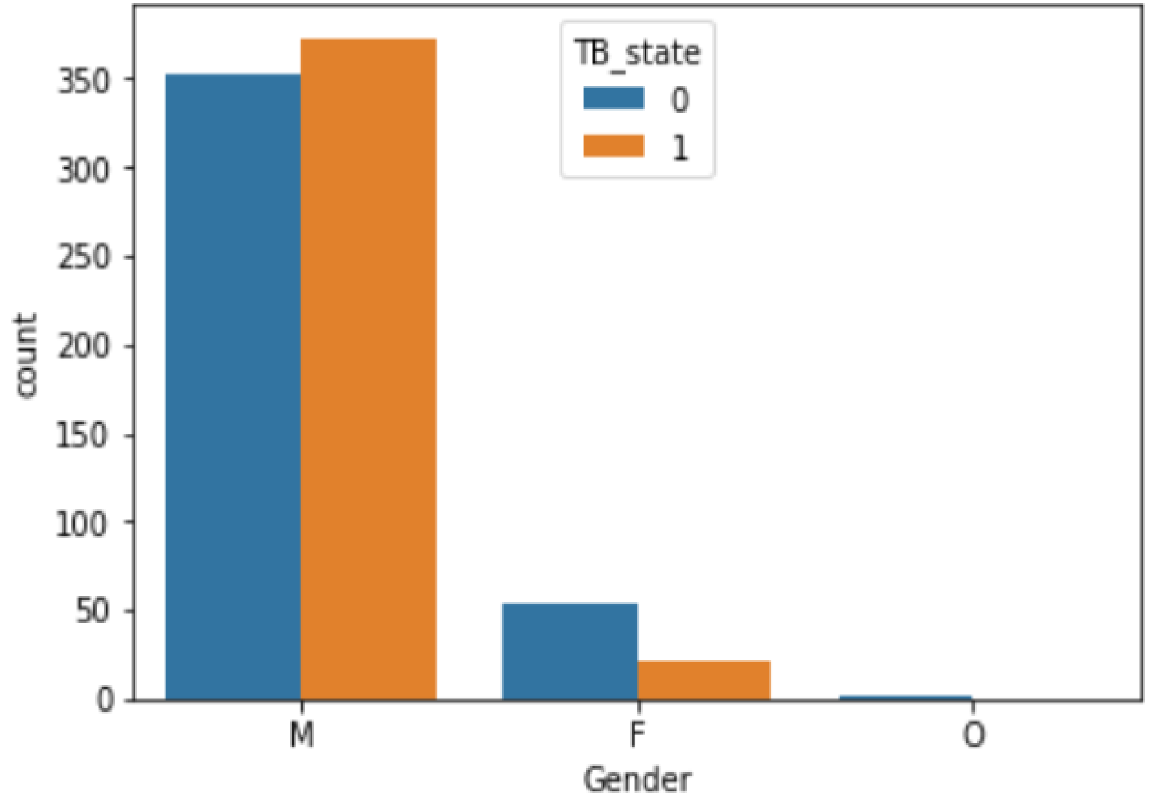


TB Negative

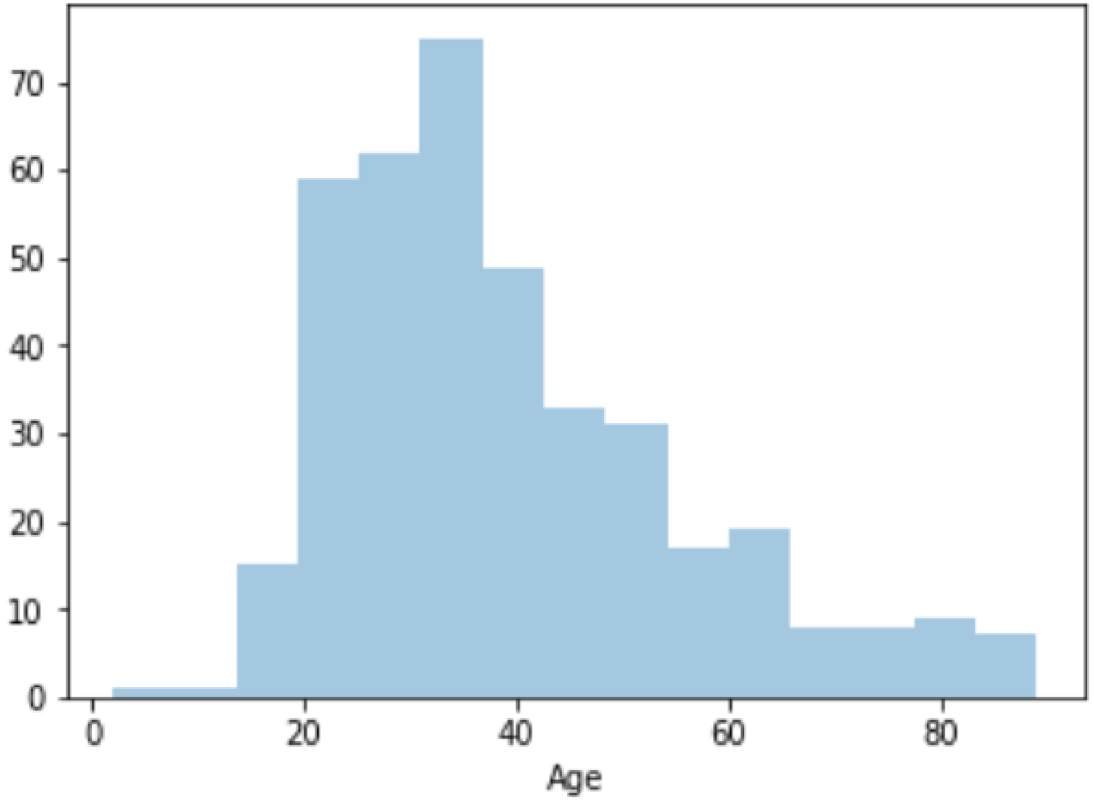
Source: https://www.kaggle.com/nikhilpandey360/chest-xray-masks-and-labels

About The Dataset

TB positive (%): Male 51.4% vs. Female 28.4%



Age distribution of TB positive patients



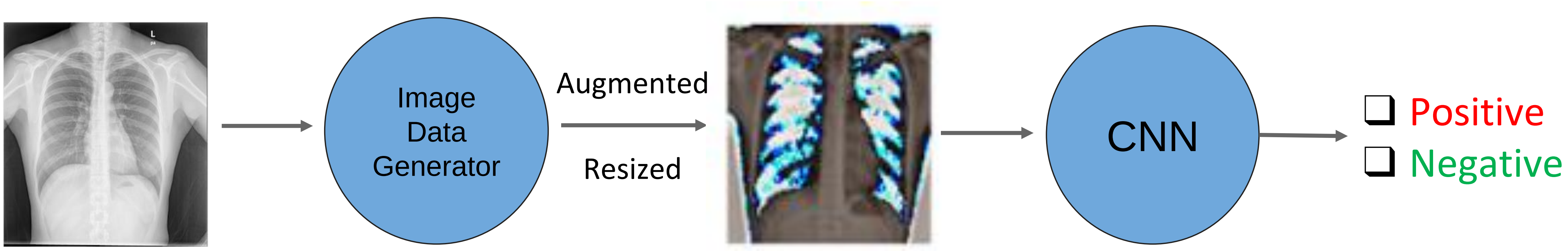
Outline of Our Project

- Transfer Learning of CNN + Hyperparameter Tuning
- Training using masked images
- Image segmentation/mask generation + dilation
- Pretrained Model as feature extractor + SVM
- Combining methods to construct best model

Transfer Learning

Using pre-trained CNN architectures +
Hyperparameter Tuning

Process Flow



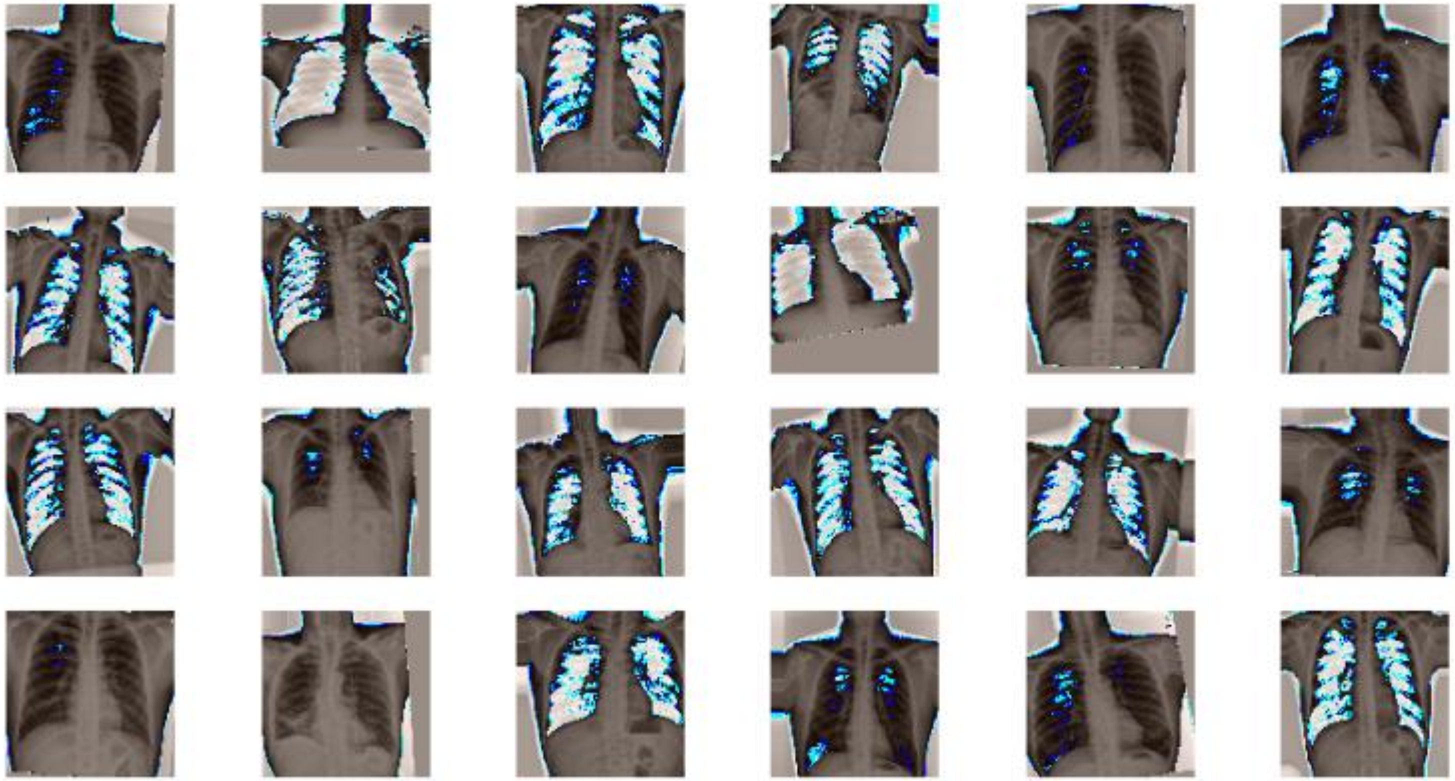
Data Augmentation + Image Processing

- ImageDataGenerator function in Keras:

```
data_generator_with_aug = ImageDataGenerator(preprocessing_function=p
reprocess_input,
                                   rotation_range=10,
                                   width_shift_range = 0.1,
                                   height_shift_range = 0.1,
                                   zoom_range=0.1)
```

- Resize image into size (3, 256, 256)

Data Augmentation + Image Processing



Models to Be Used + Different Ways of Applying Models

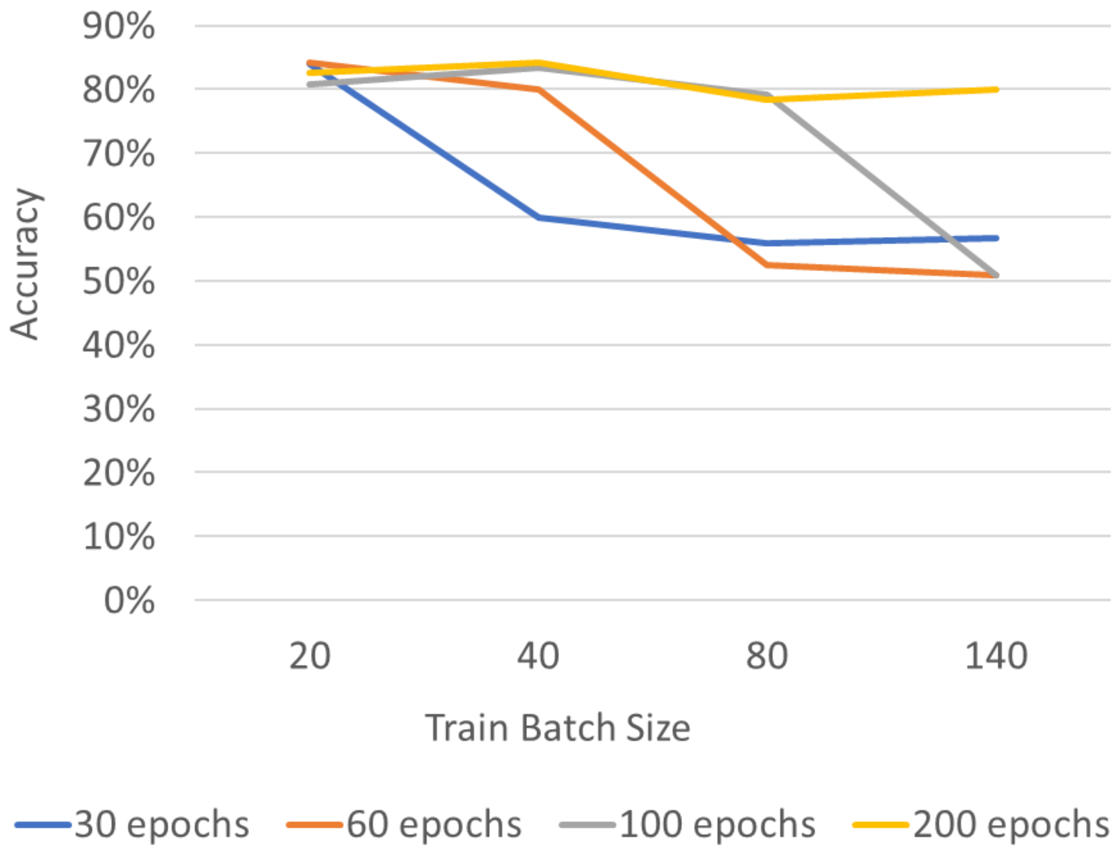
Last layer of CNN is modified for binary classification

- ResNet50
- InceptionV2
- InceptionResNetV2
- Xception
- VGG16
- VGG19

3 different ways of applying models:

- Applying pretrained model directly
- Training from scratch (random initialization)
- Refining with initial weights from ImageNet

Influence of train batch size on prediction



- Running smaller number of epochs, smaller train batch size tends to have higher accuracy
- For larger number of epochs, the above effect is not obvious
- When number of epochs is larger than 100, train batch size of 40 has highest accuracy
- So **Train batch size 40** and **100 epochs** will be used

Training Models

- Train data size 560 (70%) + Validation data size 120 (15%) + Test data size 120 (15%)
- Train batch size 40
- 100 epochs
- Save best checkpoint

Prediction Accuracy Comparison

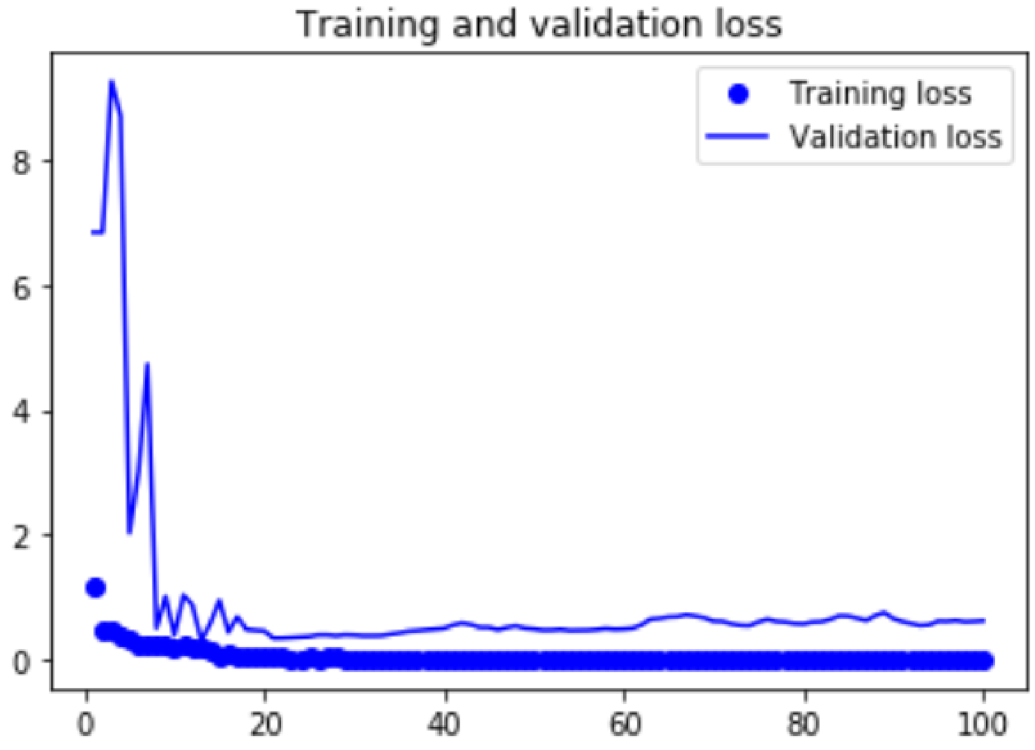
VGG16, VGG19 and Xception perform better with pretrained model

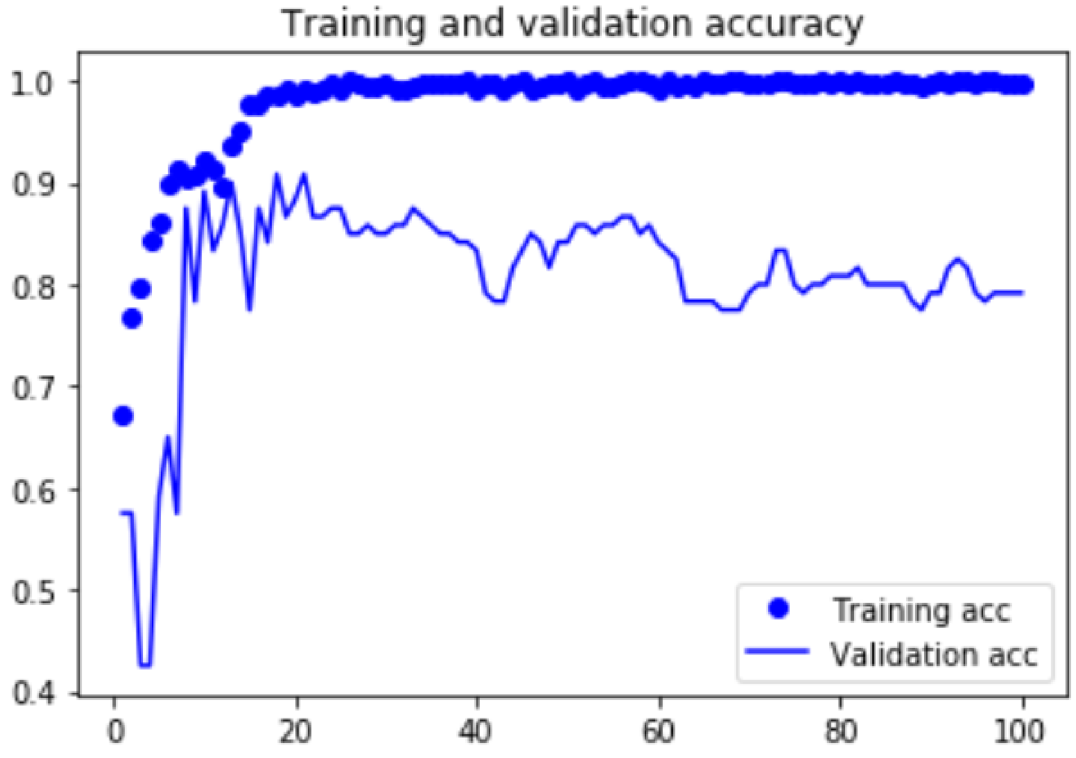
ResNet50, InceptionV3 and InceptionResNetV2 perform better when training with initial weights

Model	Pretrained	Train from Scratch	Train with Initial Weights
VGG16	0.83	0.73	0.43
VGG19	0.84	0.72	0.43
Xception	0.83	0.52	0.59
ResNet50	0.83	0.70	0.86
InceptionV3	0.75	0.81	0.82
InceptionResNetV2	0.78	0.82	0.77

ResNet50 (85.8% Acc.)

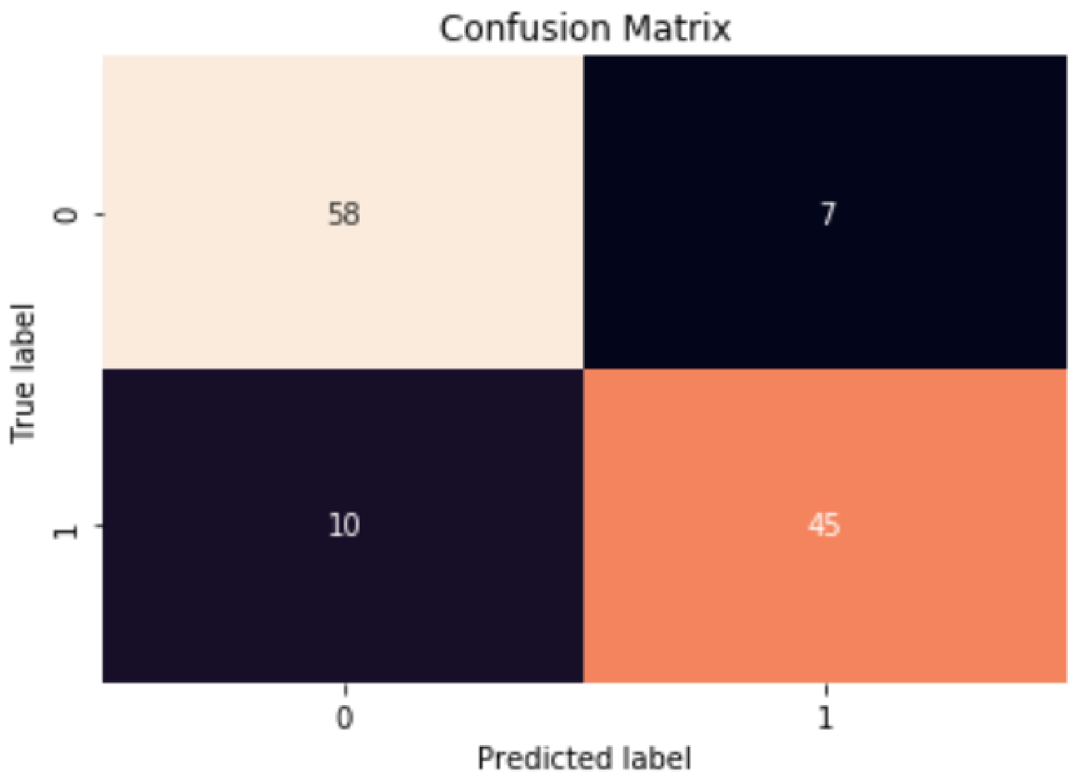
- Validation accuracy reaches the highest around 30 epochs
- Overfitting





ResNet50 (85.8% Acc.)

False negative rate 18.2% vs. False positive rate 10.8%

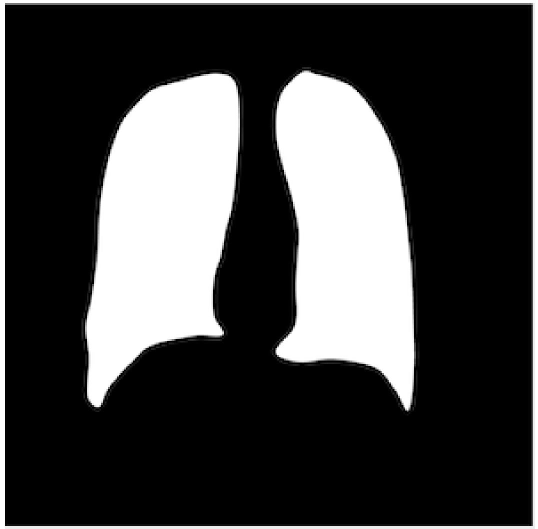


	precision	recall	f1-score	support
0	0.85	0.89	0.87	65
1	0.87	0.82	0.84	55
micro avg	0.86	0.86	0.86	120
macro avg	0.86	0.86	0.86	120
weighted avg	0.86	0.86	0.86	120

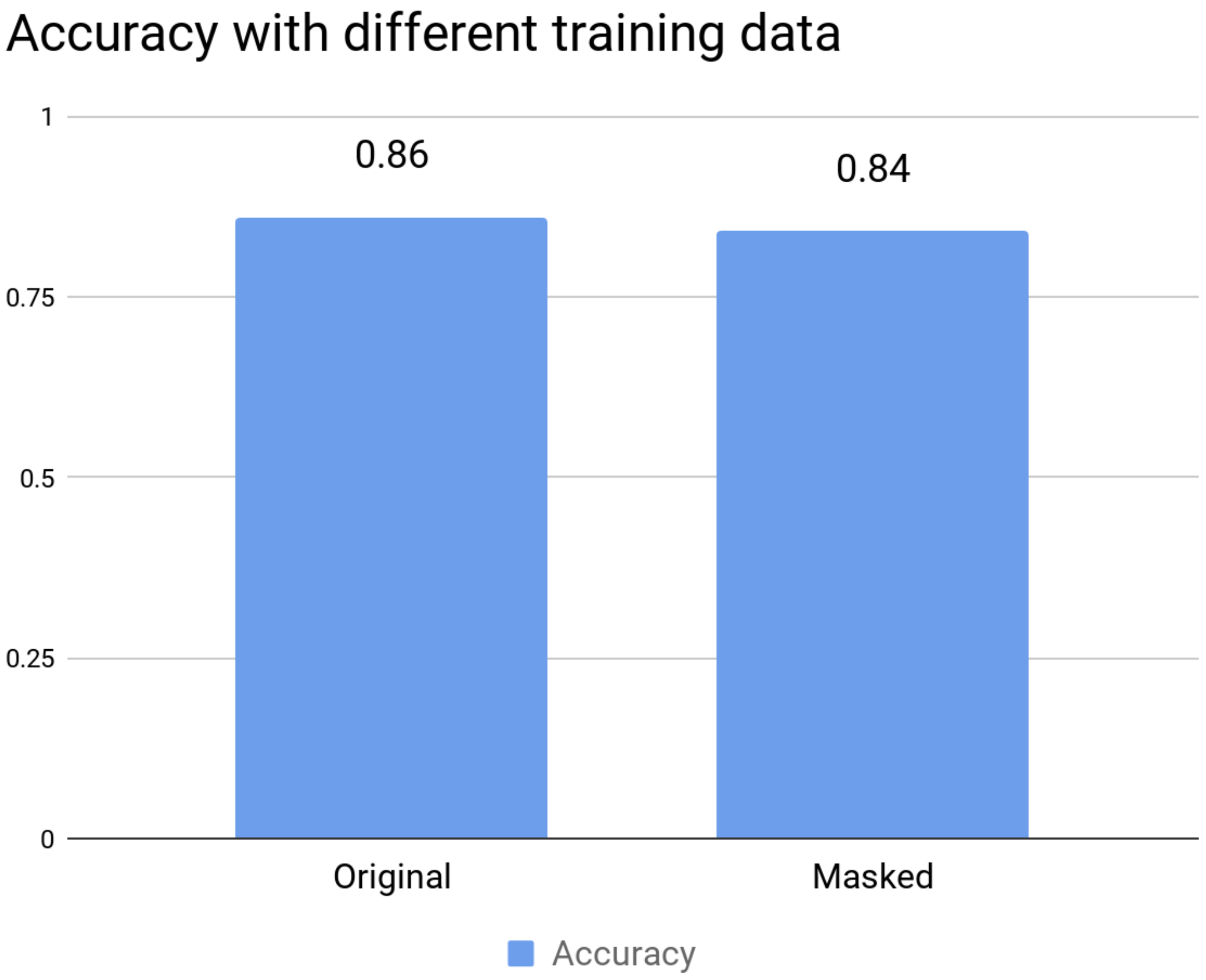
Apply mask

To improve transfer learning

- There are 704 masks outlining the lung’s shape manually marked by medical professionals.
- Masking the x-ray image removes noise outside the region of interest (i.e. the lung)
- We suspect the model will learn better with a cleaner data after masking.



- We do the same ResNet50 transfer learning.
- Instead of feeding the original x ray images, we feed the corresponding masked x-ray images as training input.



The result isn't better, why?

- The dataset is missing ~100 masks. (704 masks vs 800 images)
- Thus we have considerably less training data. (12% less)
- Need to get the leftover 96 images masked somehow.

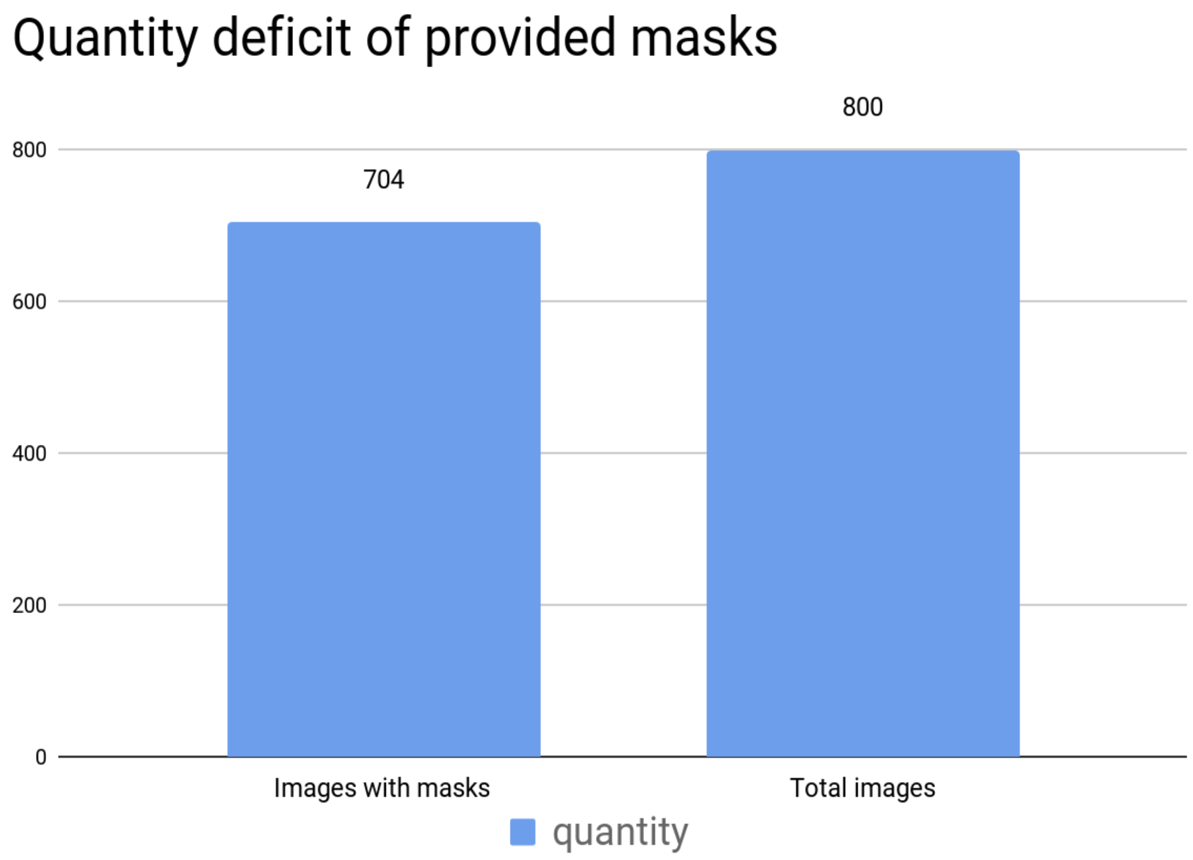


Image Segmentation

To generate masks

Challenge + Approach

Challenge: Missing masks for 96 images

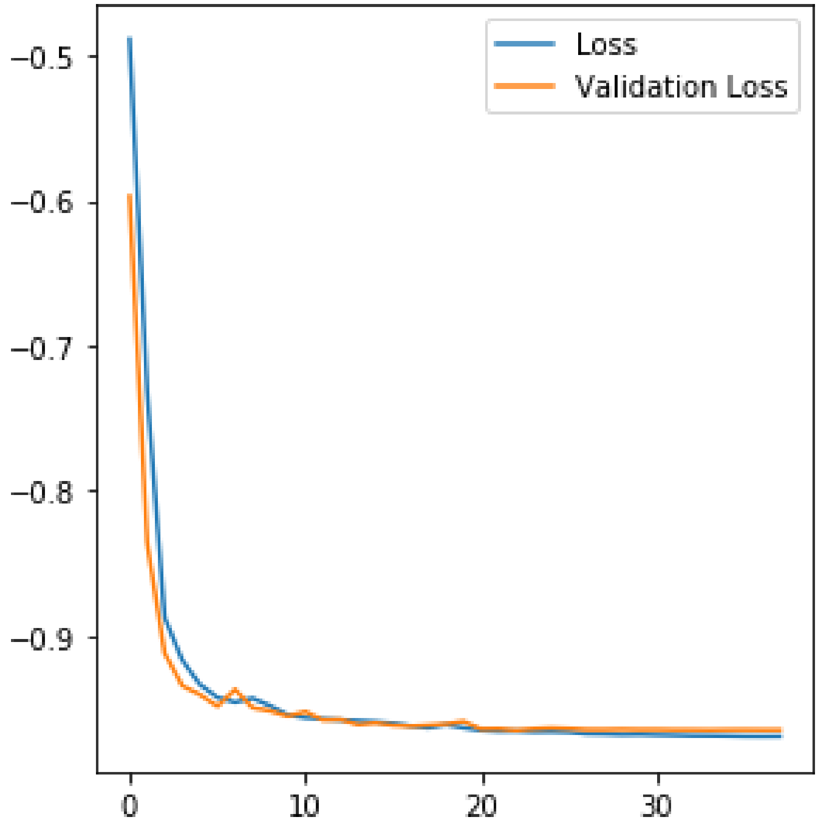
Solution: Train a U-Net model to generate masks

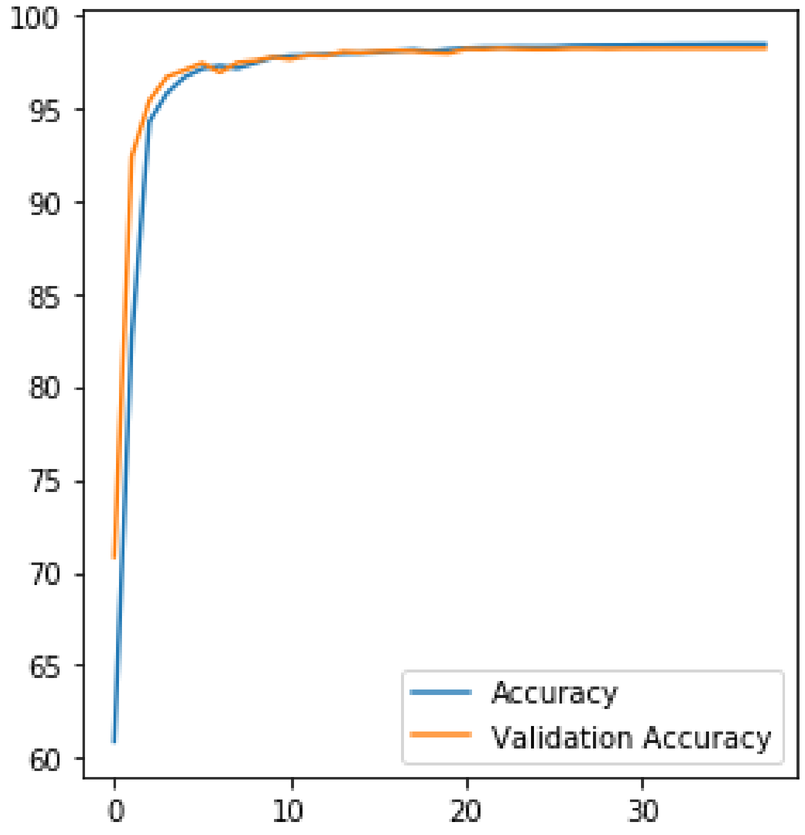
Training Specifics:

Preprocessing masks to binary encoding
Scale images to center around 127
Save best only
150 epochs + 4 batch size + early stopping patience = 15
Reduce learning rate on plateau

Source: https://arxiv.org/abs/1505.04597

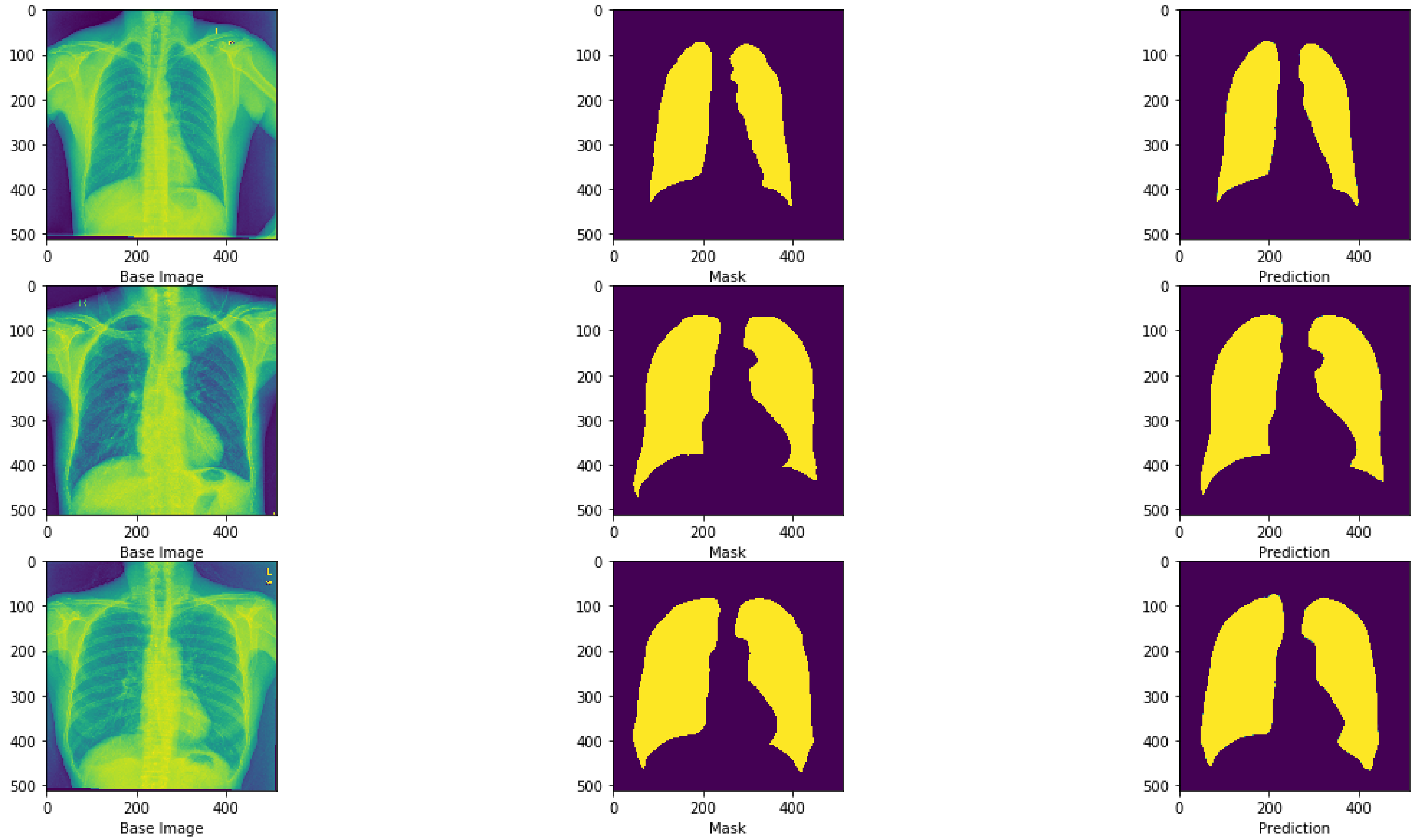
Training Result





Training Accuracy: 98.3%

Sampled Testing Output



Testing Accuracy: 97.7%

Generating masks

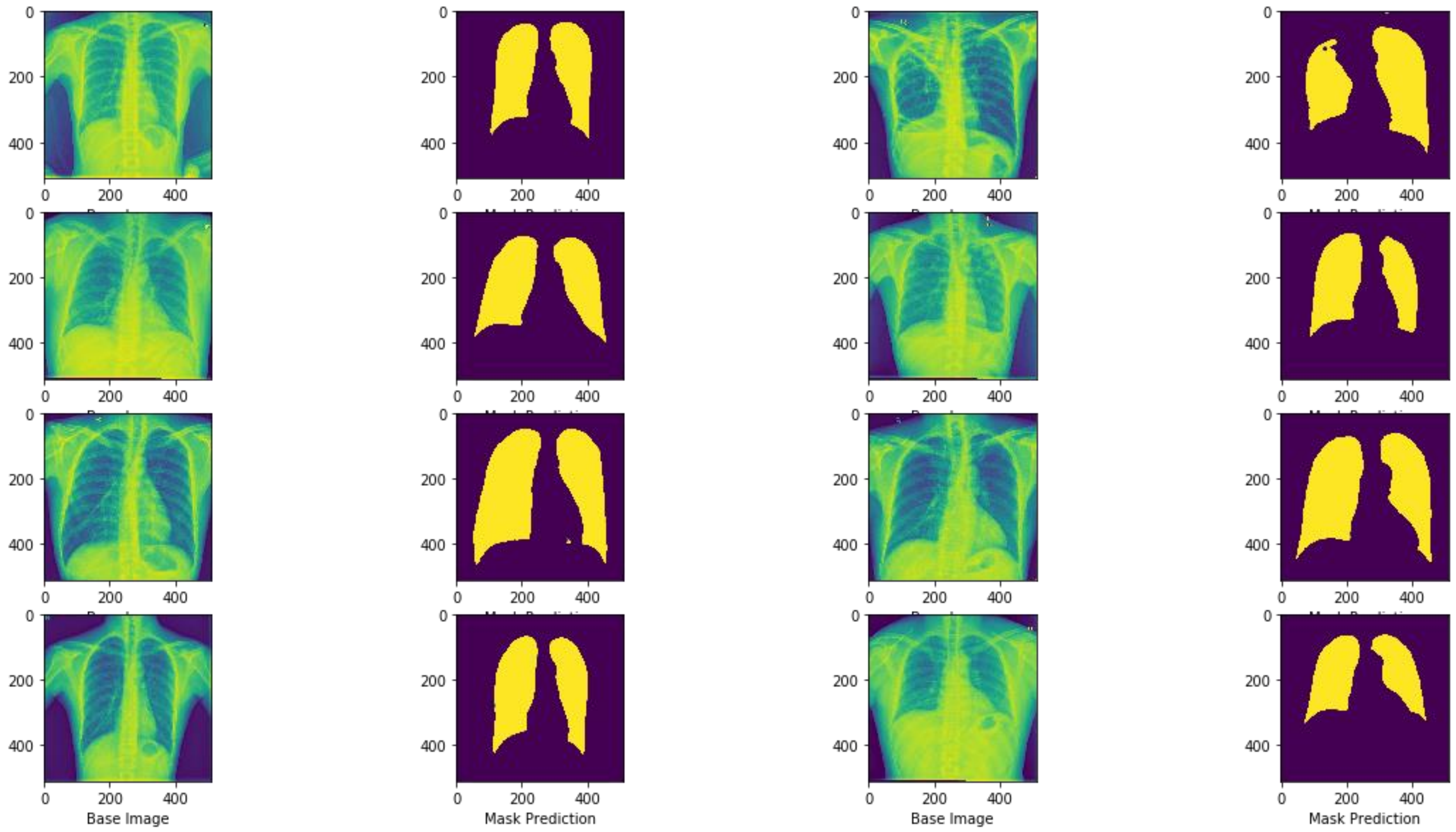
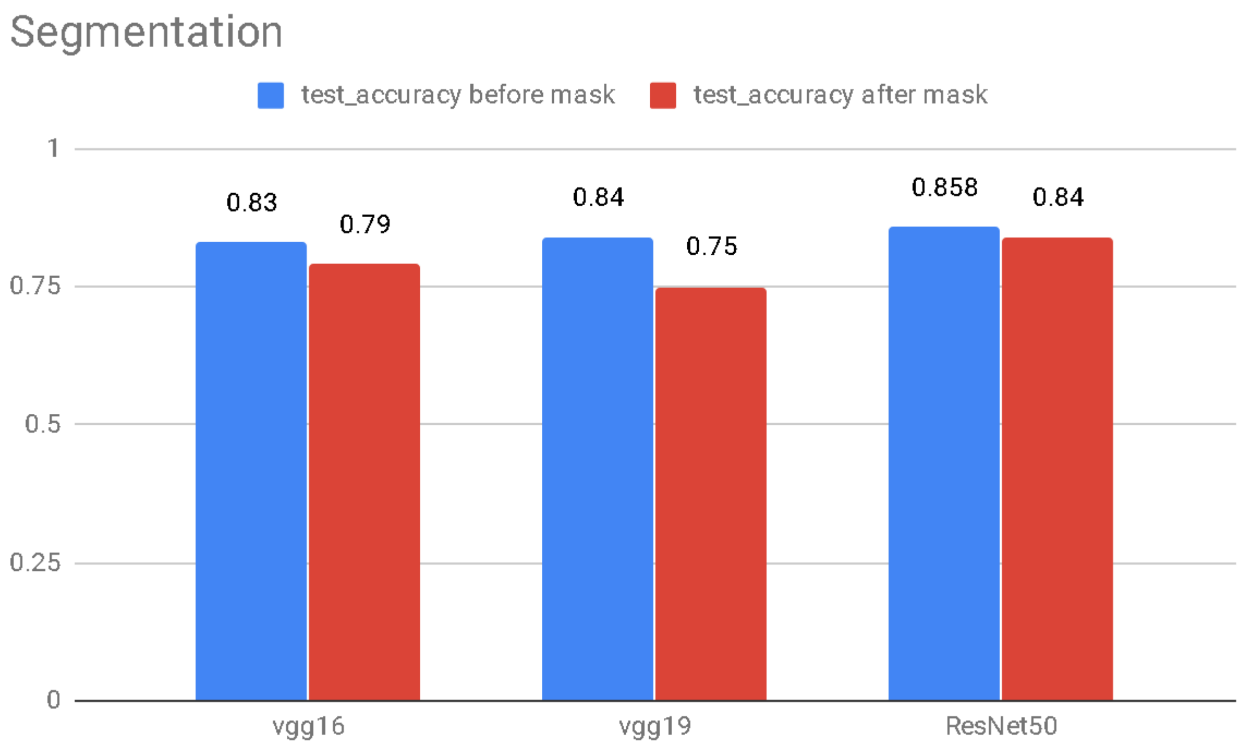


Image Segmentation



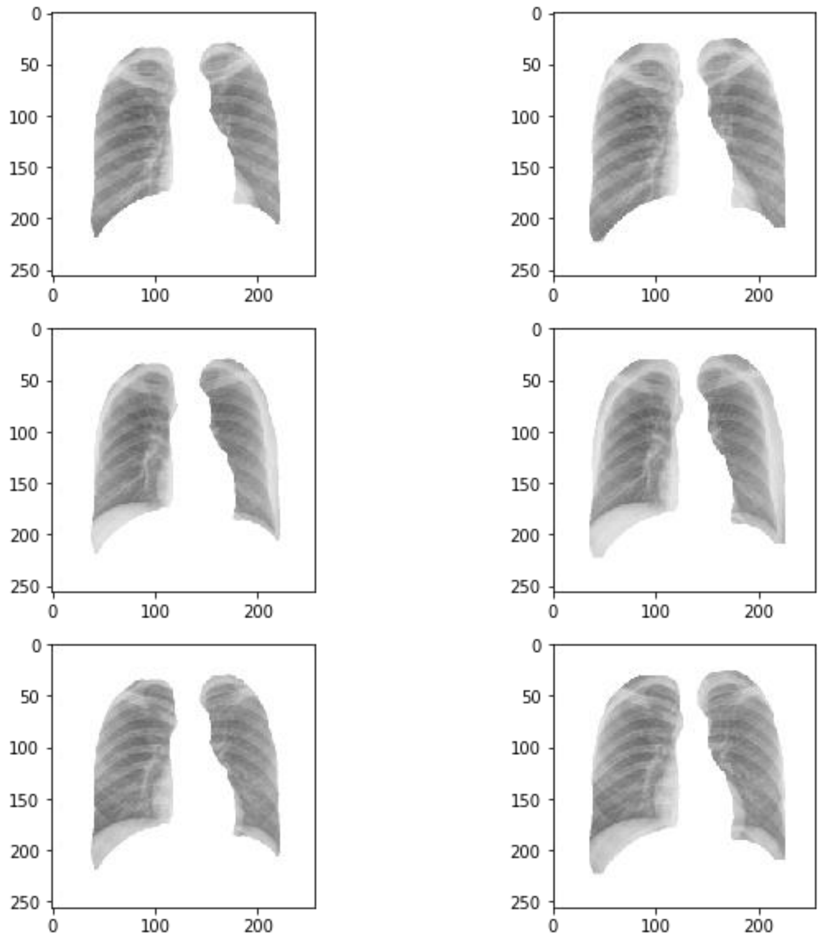
- Segment Image with masks to focus training on lung areas
- Finding: performance deteriorated after mask application

New Challenge + Approach

Challenge: Performance decreases after applying masks to image data

Our Assumption: Masked images might lose information along the fuzzy boundaries of lungs

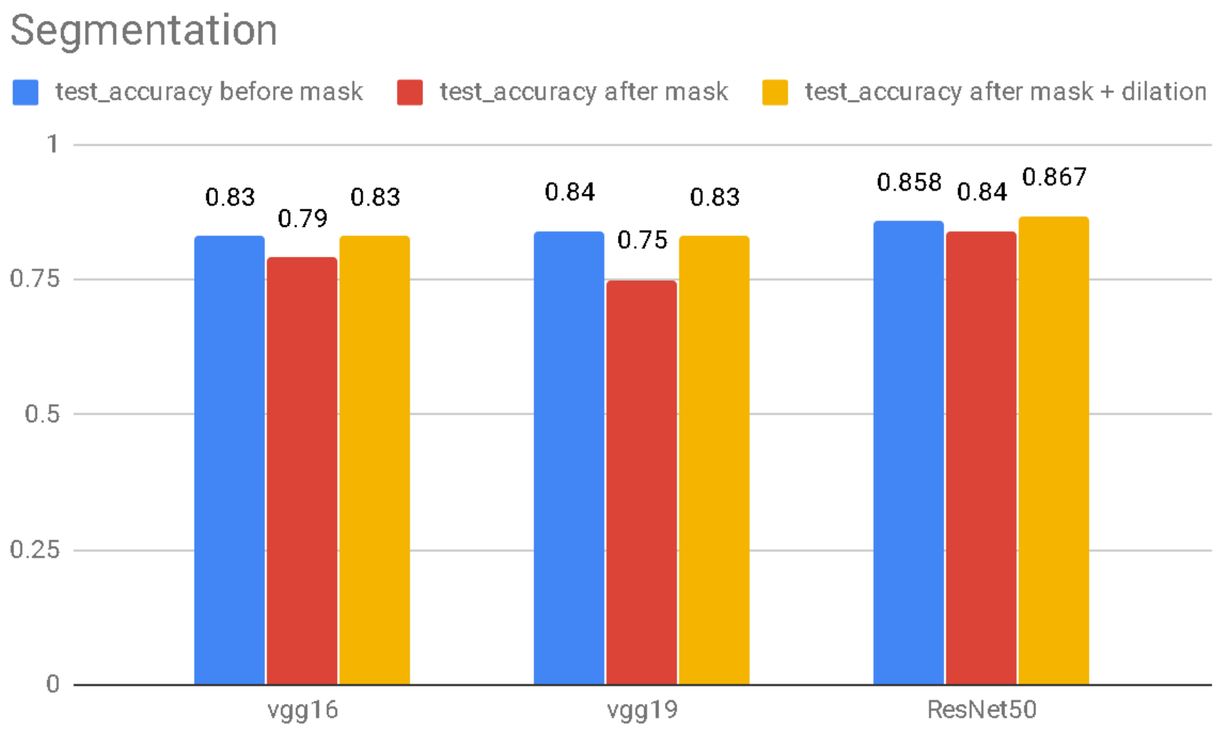
Solution: Apply dilation to enlarge the boundaries of regions of foreground pixels in order to incorporate more information

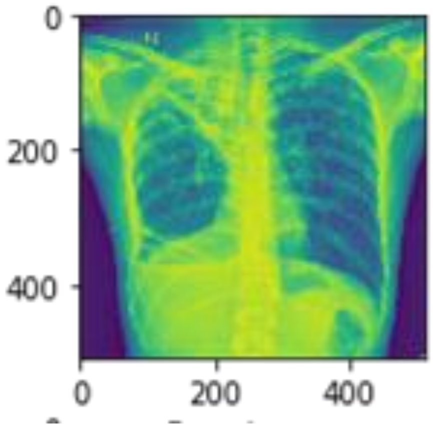


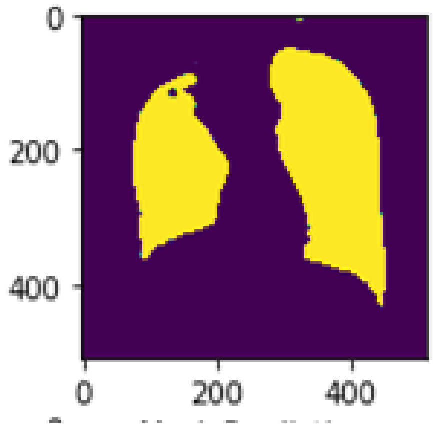
Source: https://homepages.inf.ed.ac.uk/rbf/HIPR2/dilate.htm

https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_imgproc/py_morphological_ops/py_morphological_ops.html

Image Segmentation + Dilation





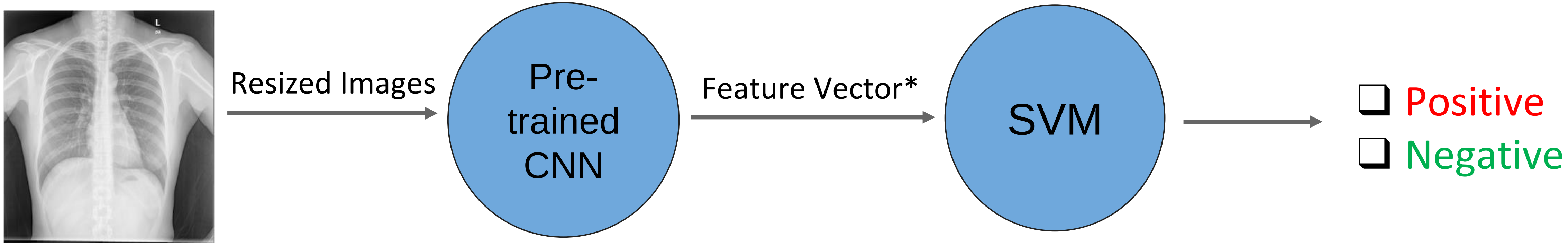


- Improved accuracy, similar to before mask application
- Validated our assumption

Feature Extraction

Using pretrained model to extract features and feeding them into traditional machine learning model

Approach



*Feature vector: from the layer before the classification layer

Model for Feature Extraction	Image Size	Feature Shape
ResNet50	(224, 224)	(1, 2048)
InceptionV3	(299, 299)	(1, 2048)
InceptionResNetV2	(299, 299)	(1, 1536)

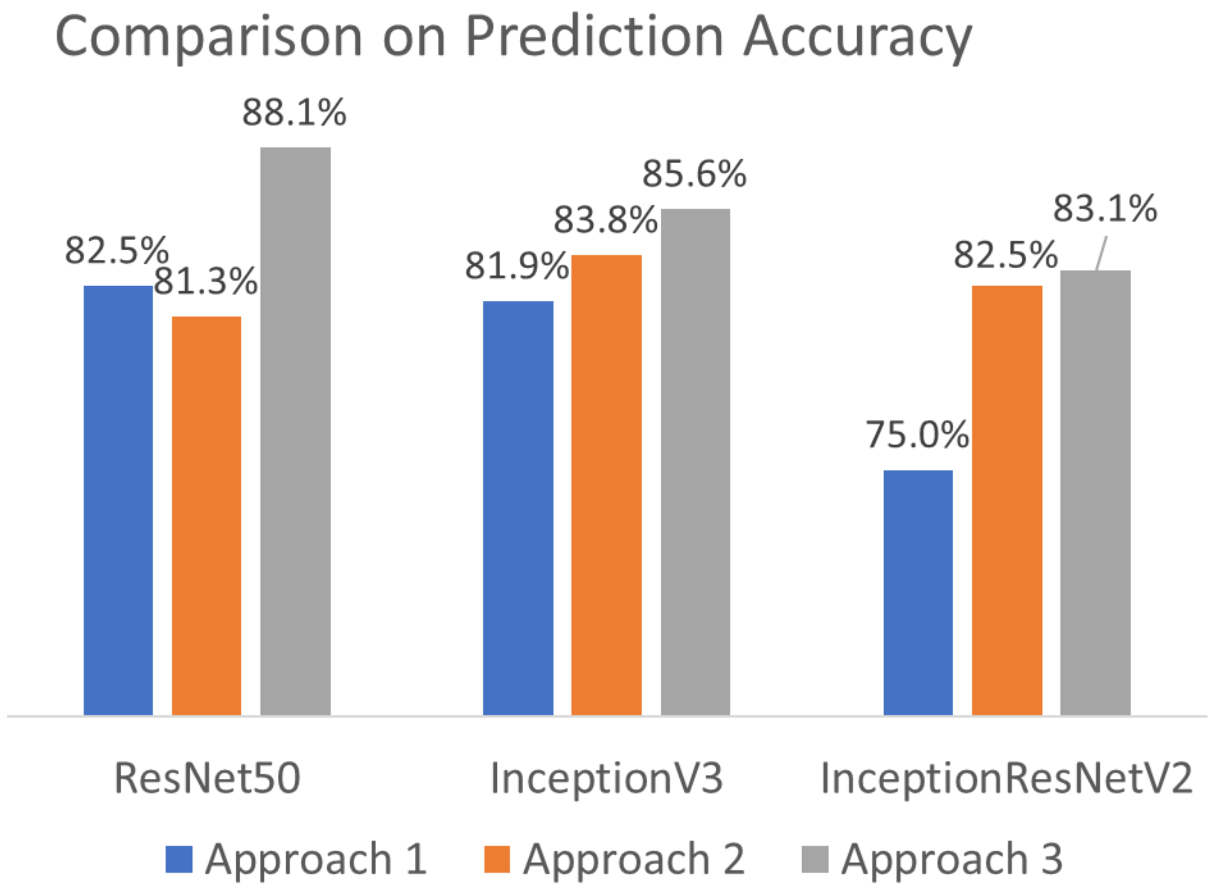
Source: https://arxiv.org/abs/1403.6382
https://www.sciencedirect.com/science/article/pii/S0010482517302548?via%3Dihub

Fine Tune SVM

Approach	Cross Validation	Grid Search Parameters	Feature Vector Normalization	Scoring Methods
Approach 1	✗	✗	✗	Accuracy
Approach 2	✓	✓	✗	AUC & Accuracy
Approach 3	✓	✓	✓	AUC & Accuracy

- Cross validation: 5 folds
- Grid search parameters: C (1, 10, 100, 1000), gamma (0.0001, 0.001, 0.01, 0.1, 1)
- Train data size 512 (64%) + Validation data size 128 (16%) + Test data size 160 (20%)

Prediction Performance Comparison

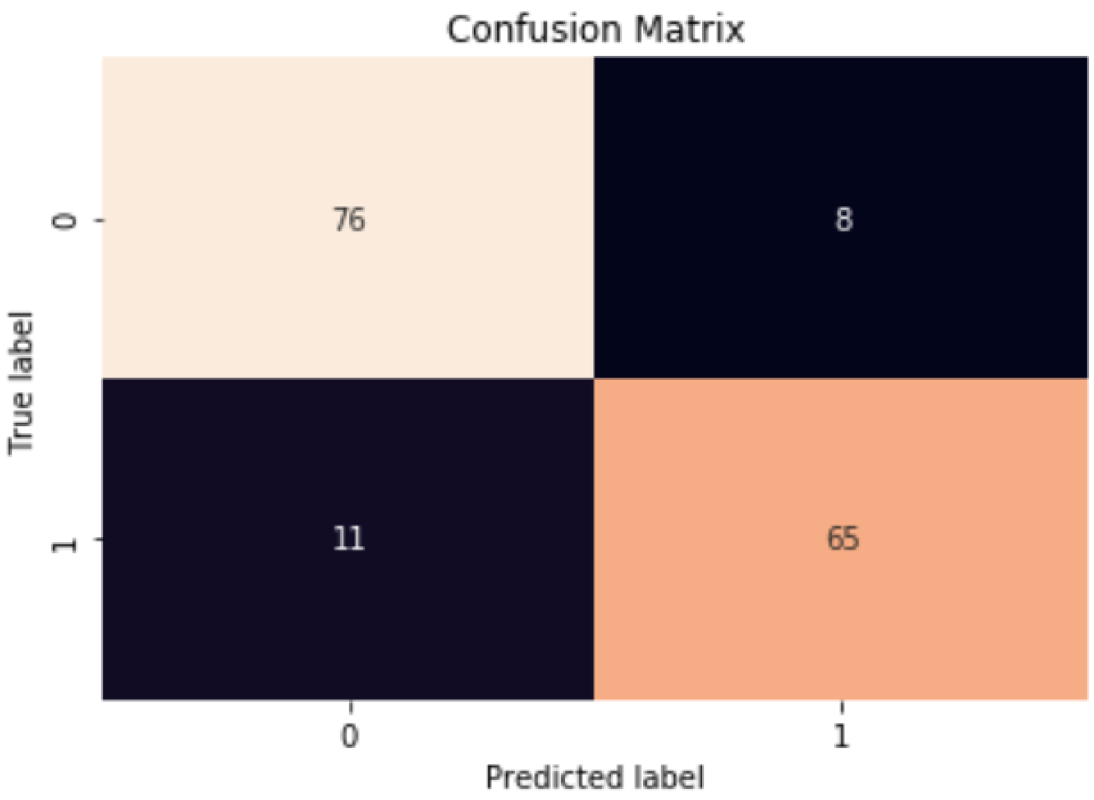


Observations:

- Best performance 88.1%:
  ResNet50 with Approach 3
- Approach 3 > Approach 2 > Approach 1
- Normalization of feature vectors improves performance

ResNet50 in Approach 3 (88.1% Acc.)

False negative rate 14.5% vs. False positive rate 9.5%



Classification Report

	precision	recall	f1-score	support
0	0.87	0.90	0.89	84
1	0.89	0.86	0.87	76
micro avg	0.88	0.88	0.88	160
macro avg	0.88	0.88	0.88	160
weighted avg	0.88	0.88	0.88	160

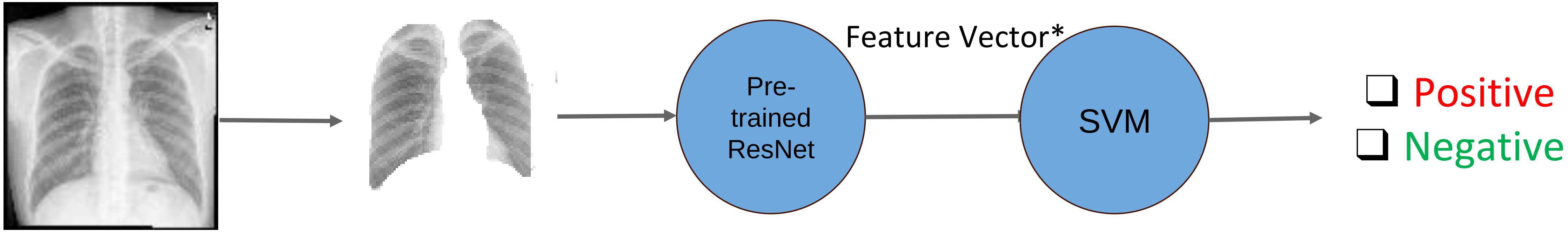
Feature Extraction + SVM vs. Pure CNN

- Better prediction accuracy:
  - Feature extraction + SVM: 88.1%
  - Pure CNN: 85.8%
- Shorter running time of model:
  - Feature extraction + SVM: ~ 5mins (2 mins on feature extraction + 3 mins on Grid Search)
  - Pure CNN: > 20 mins for 100 epochs

Combined

Segmentation/Dilation + ResNet + Feature Extraction

Running the final model



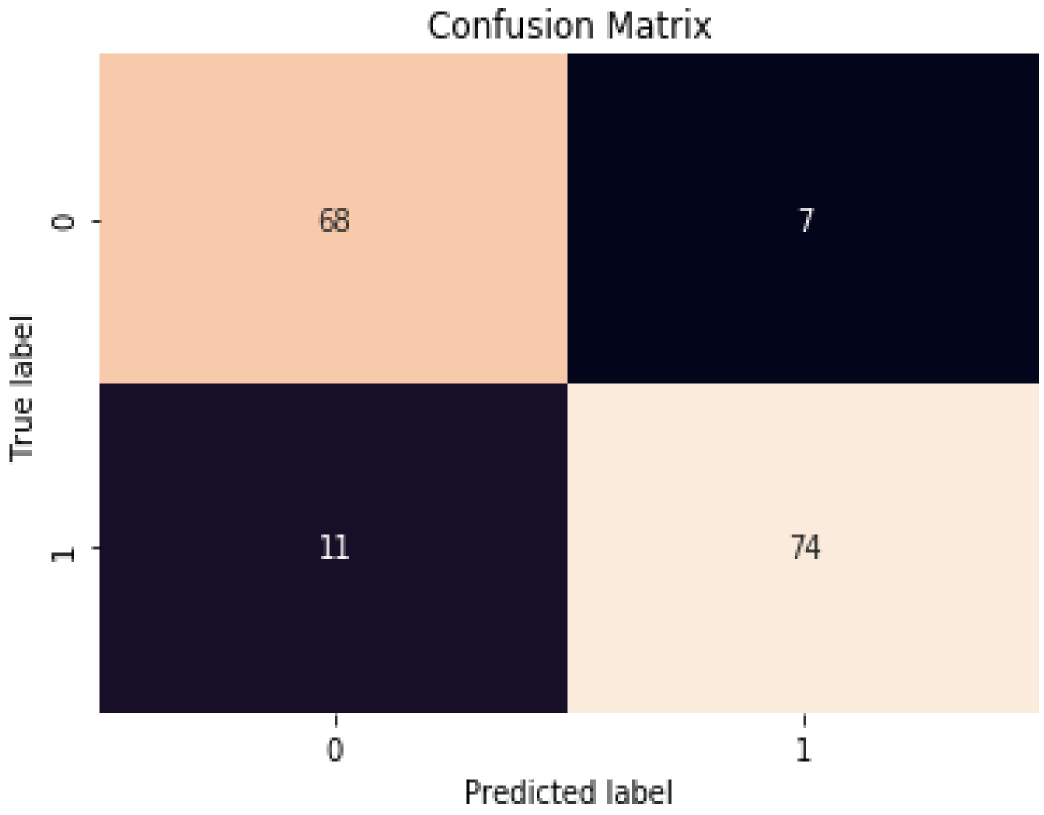
*Feature vector: from the layer before the classification layer

Segmentation/Dilation + ResNet + Feature Extraction (Approach 3)

Test accuracy: 88.75%, best so far

Final model result

False negative rate 12.9% vs. False positive rate 9.3%



	precision	recall	f1-score	support
0	0.86	0.91	0.88	75
1	0.91	0.87	0.89	85
micro avg	0.89	0.89	0.89	160
macro avg	0.89	0.89	0.89	160
weighted avg	0.89	0.89	0.89	160

Project Summary

Key findings

1. For small datasets, Pre-trained models performs much better than train from scratch models.
2. Choice of train batch size for efficiency: with lower epochs, train batch size of 40 is optimal. Higher batch size will compromise the accuracy, or more epochs are needed to achieve the same accuracy level.
3. Mask segmentation without dilation performance deteriorated; After dilation on segmentation is applied, ResNet50 results slightly improved, yielding the best results.
4. Finally CNN Feature vector extraction and normalization + SVM Classifier improved model accuracy by 2.9%

Future works

1. For diagnosis purpose, we should introduce higher penalty for false negative prediction, which will reduce accuracy.
2. If we have more data, we want to examine the effect of mask application with and without dilation.
3. Image annotation techniques shall be applied to pinpoint the most probable regions of disease contraction.
4. Combining point 2 and 3, we can hopefully determine if the reduced performance with undilated masks is cause by the loss of critical region discarded by segmentation.

Acknowledgement

The two datasets were published together in an analysis here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4256233/. The datasets come from Shenzhen and Montgomery respectively.

China Set - The Shenzhen set - Chest X-ray Database

The standard digital image database for Tuberculosis is created by the National Library of Medicine, Maryland, USA in collaboration with Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China. The Chest X-rays are from out-patient clinics, and were captured as part of the daily routine using Philips DR Digital Diagnose systems.

It is requested that publications resulting from the use of this data attribute the source (National Library of Medicine, National Institutes of Health, Bethesda, MD, USA and Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China) and cite the following publications:

Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Xue Z, Palaniappan K, Singh RK, Antani S, Thoma G, Wang YX, Lu PX, McDonald CJ. Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging. 2014 Feb;33(2):233-45. doi: 10.1109/TMI.2013.2284099. PMID: 24108713

Candemir S, Jaeger S, Palaniappan K, Musco JP, Singh RK, Xue Z, Karargyris A, Antani S, Thoma G, McDonald CJ. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans Med Imaging. 2014 Feb;33(2):577-90. doi: 10.1109/TMI.2013.2290491. PMID: 24239990

For masks:

Yu. Gordienko, Yu. Kochura, O. Alienin, O. Rokovyi, S. Stirenko, Peng Gang, Wei Zeng, Chest X-Ray Analysis of Tuberculosis by Deep Learning with Segmentation and Augmentation, arXiv preprint arXiv:1803.01199 (2018).

Acknowledgement (continued)

Montgomery County X-ray Set

X-ray images in this data set have been acquired from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA. This set contains 138 posterior-anterior x-rays, of which 80 x-rays are normal and 58 x-rays are abnormal with manifestations of tuberculosis. All images are de-identified and available in DICOM format. The set covers a wide range of abnormalities, including effusions and miliary patterns. The data set includes radiology readings available as a text file.

National Library of Medicine, National Institutes of Health, Bethesda, MD, USA;

Computer Engineering Department, Faculty of Informatics and Computer Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine;