

# PNEUMONIA DETECTION

OpenCampus – TensorFlow – Januar 2020

# Structure

- What is Pneumonia?
- Overview of the dataset
- Collab Presentation
- Outlook

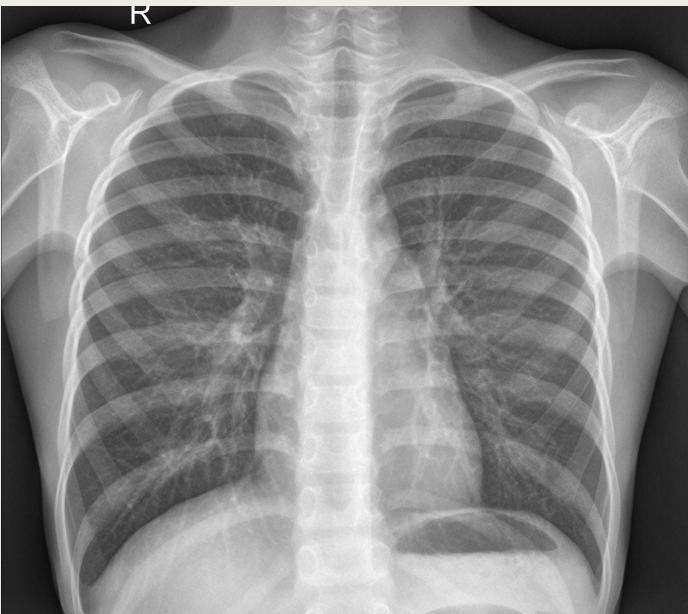
# What is Pneumonia?

- Inflammatory condition of the lung
- Affects 450 mio. people per year
- Usually caused by infection with viruses or bacteria
- Survival improved with antibiotics → in developing countries and among the very old/young still a severe problem

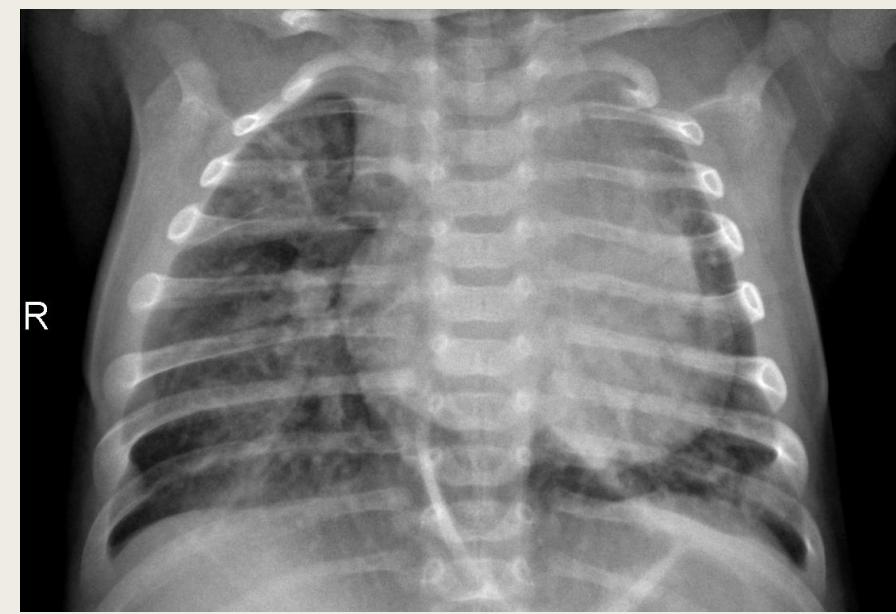
# Diagnosis based on X-ray

- Diagnosis often based on findings on chest X-ray images
- Typical findings:
  - local consolidation of lung tissue
  - hyper-inflated patches
  - badly ventilated appear brighter

# Comparison of the X-rays



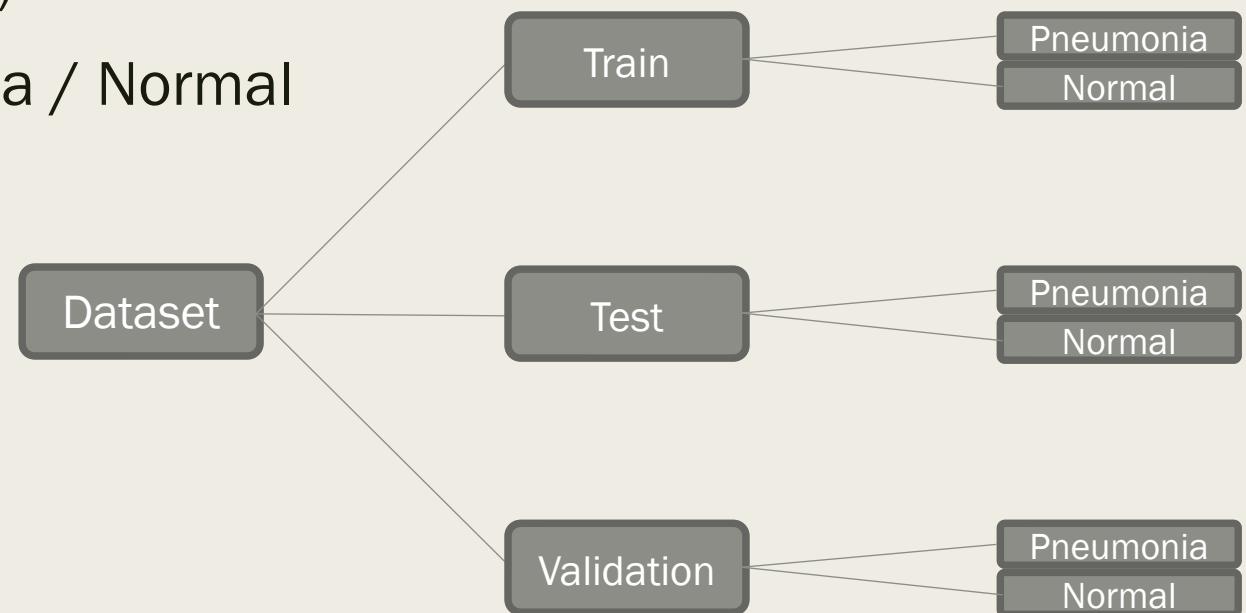
healthy



Pneumonia

# Overview of the dataset

- Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), “Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification”, Mendeley Data, v2
- 5863 X-Ray images (JPEG)
- Two categories: Pneumonia / Normal
- 4678 for training
- 554 for testing
- 624 for validation



# Goal of our estimations

- Automated detection of potential Pneumonia cases
- Support, not replace pulmonologists

# Technical environment

- Data repository: Google Drive
- Coding environment: Google Colab
- Python version: 3.6
- Tensorflow version: 2.0.0

# Train data-generator

- Image Augmentation: rotation, zoom, width and height shift
- Convert images to 150x150 pixels in grayscale
- Batch size: 20

# The Model

- 6 convolutional layers (16, 32, 48, 64, 80, 96) which are separated by max-pooling and batch normalization layers
- A global-average pooling layer flattens the output from the last convolutional layer for the first dense layer
- 3 dense layers (64, 32 and 1 neuron), each separated by a dropout layer
- Optimizer: Adam with lr=0.00009
- Loss function: binary cross entropy
- We trained until validation accuracy exceeded 95%

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 148, 148, 16)	160
batch_normalization (BatchNormal)	(None, 148, 148, 16)	64
max_pooling2d (MaxPooling2D)	(None, 74, 74, 16)	0
conv2d_1 (Conv2D)	(None, 72, 72, 32)	4640
batch_normalization_1 (BatchNormal)	(None, 72, 72, 32)	128
max_pooling2d_1 (MaxPooling2D)	(None, 36, 36, 32)	0
conv2d_2 (Conv2D)	(None, 34, 34, 48)	13872
batch_normalization_2 (BatchNormal)	(None, 34, 34, 48)	192
max_pooling2d_2 (MaxPooling2D)	(None, 17, 17, 48)	0
conv2d_3 (Conv2D)	(None, 15, 15, 64)	27712
batch_normalization_3 (BatchNormal)	(None, 15, 15, 64)	256
max_pooling2d_3 (MaxPooling2D)	(None, 7, 7, 64)	0
conv2d_4 (Conv2D)	(None, 5, 5, 80)	46160
batch_normalization_4 (BatchNormal)	(None, 5, 5, 80)	320
conv2d_5 (Conv2D)	(None, 3, 3, 96)	69216
batch_normalization_5 (BatchNormal)	(None, 3, 3, 96)	384
global_average_pooling2d (GlobalAveragePooling2D)	(None, 96)	0
dense (Dense)	(None, 64)	6208
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 32)	2080
dropout_1 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 1)	33
<hr/>		
Total params: 171,425		
Trainable params: 170,753		
Non-trainable params: 672		

# Performance measured on validation data

Confusion matrix

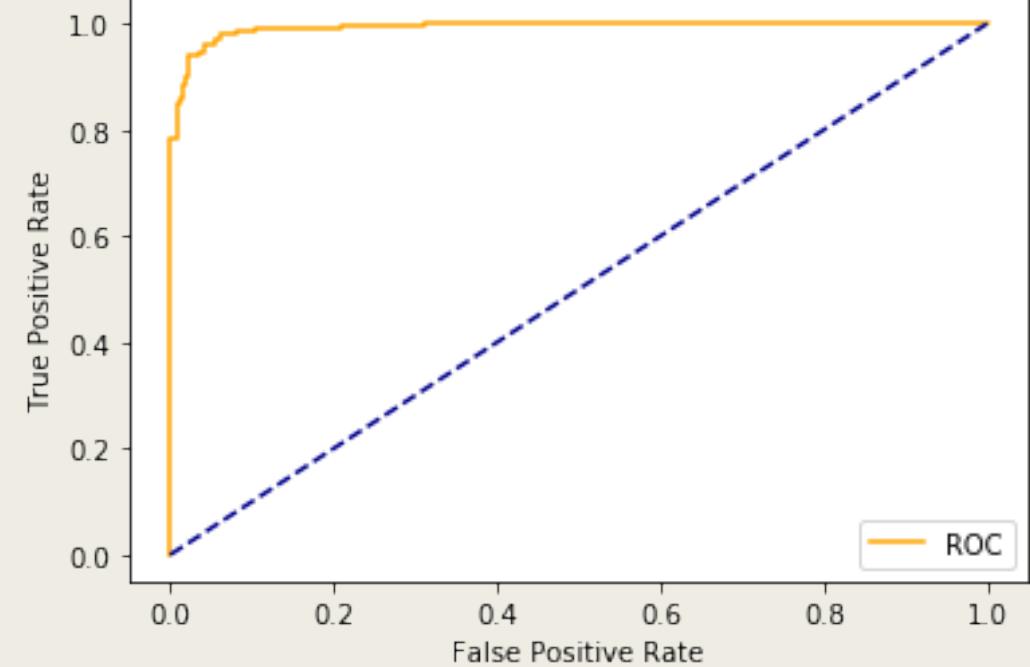
		Predicted	
		Normal	Pneomonia
True	Normal	245	16
	Pneumonia	9	284

Recall: 0.97

Precision: 0.95

Accuracy: 0.95

Receiver Operating Characteristic (ROC) Curve



AUC of the model is 0.99

# Performance on the unseen test data

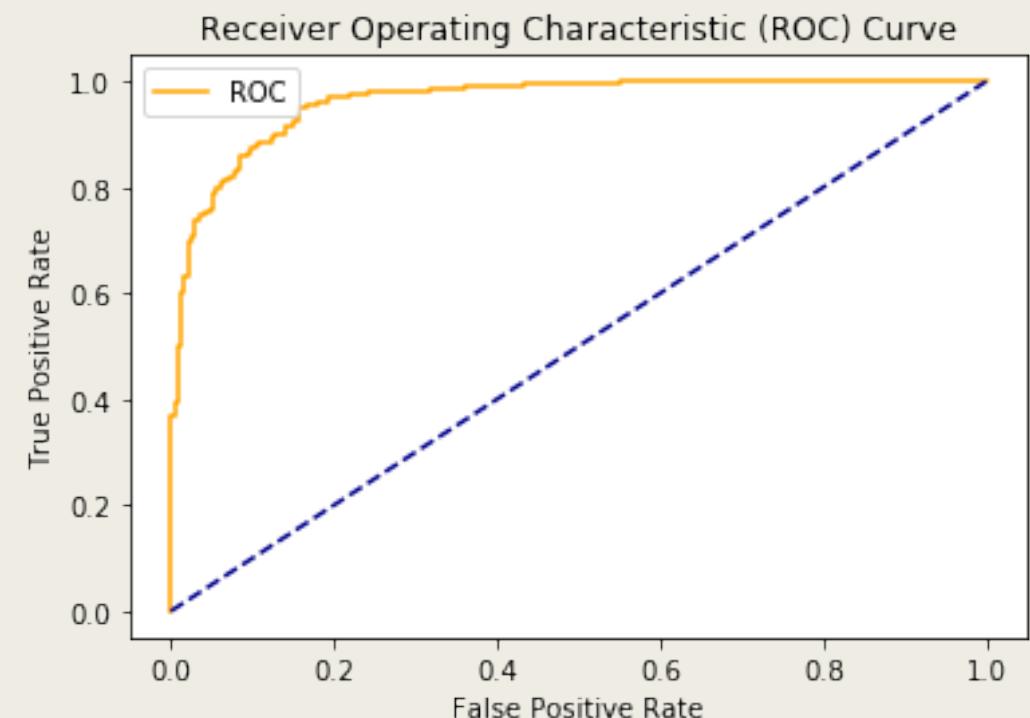
Confusion matrix

		Predicted	
		Normal	Pneomonia
True	Normal	164	70
	Pneumonia	4	386

Recall: 0.98

Precision: 0.82

Accuracy: 0.88



AUC of the model is 0.96

# Outlook and future challenges

- Implement other diseases
- Maintenance of medical databases
- Share network paramters
- Responsible handling of data