1 Description

Dataset name	Chest X-Ray Images (Pneumonia)
Link	Kaggle
License	CC BY 4.0
Medical discipline	Radiology
Medical procedure	X-ray
Multi-class problem	✗
Multi-label problem	✗

Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care.

For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

The classes that can be found in this dataset are:

- NORMAL
- PNEUMONIA

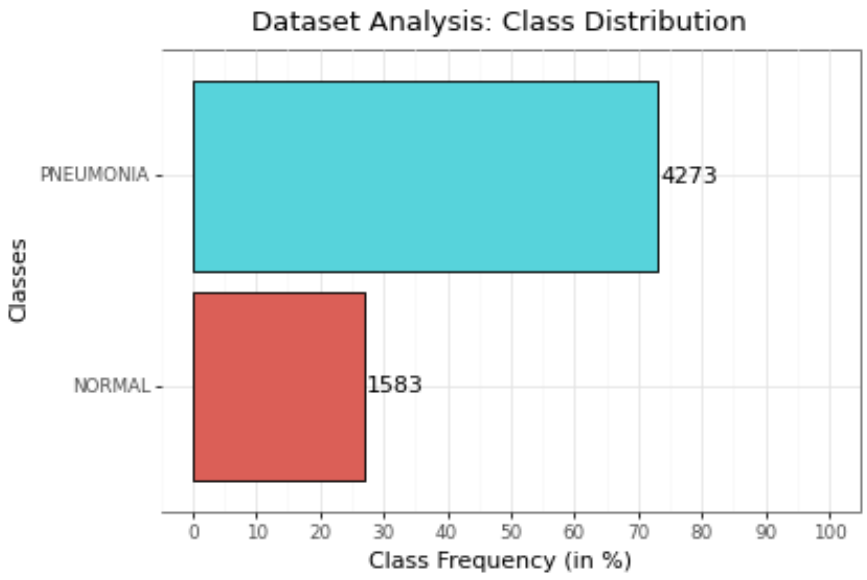


Figure 1: Class distribution: Chest X-Ray Images (Pneumonia)

2 Pre-processing

No pre-processing was done.

3 Training

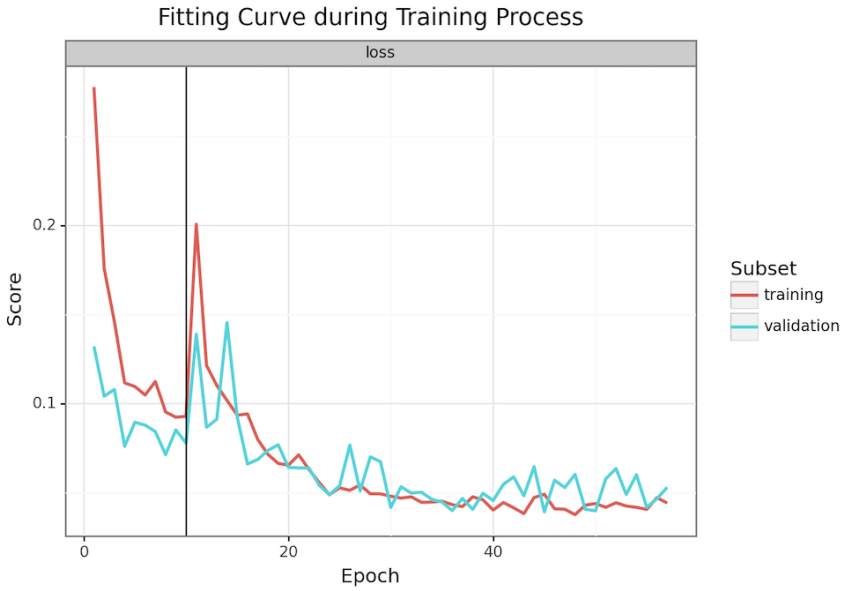


Figure 2: Training: Chest X-Ray Images (Pneumonia)

As expected, the unfreezing of the model's weights in episode 10 results in a peak in both training and validation loss. Further training shows a decrease in both losses. Validation loss shows a slight tendency to increase again, which may indicate overfitting.

4 Results

Figure 3: Metric overview: Chest X-Ray Images (Pneumonia)

All calculated metrics (see figure 3) and the ROC curve in figure 4 attest to the model's great classification performance.

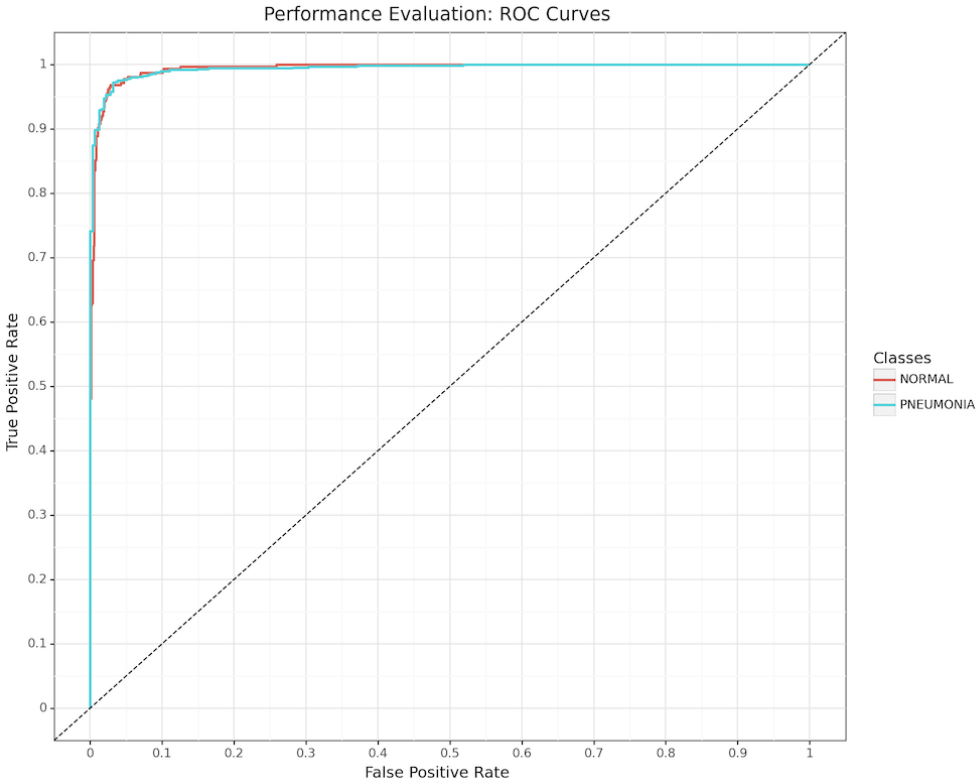


Figure 4: ROC curve: Chest X-Ray Images (Pneumonia)

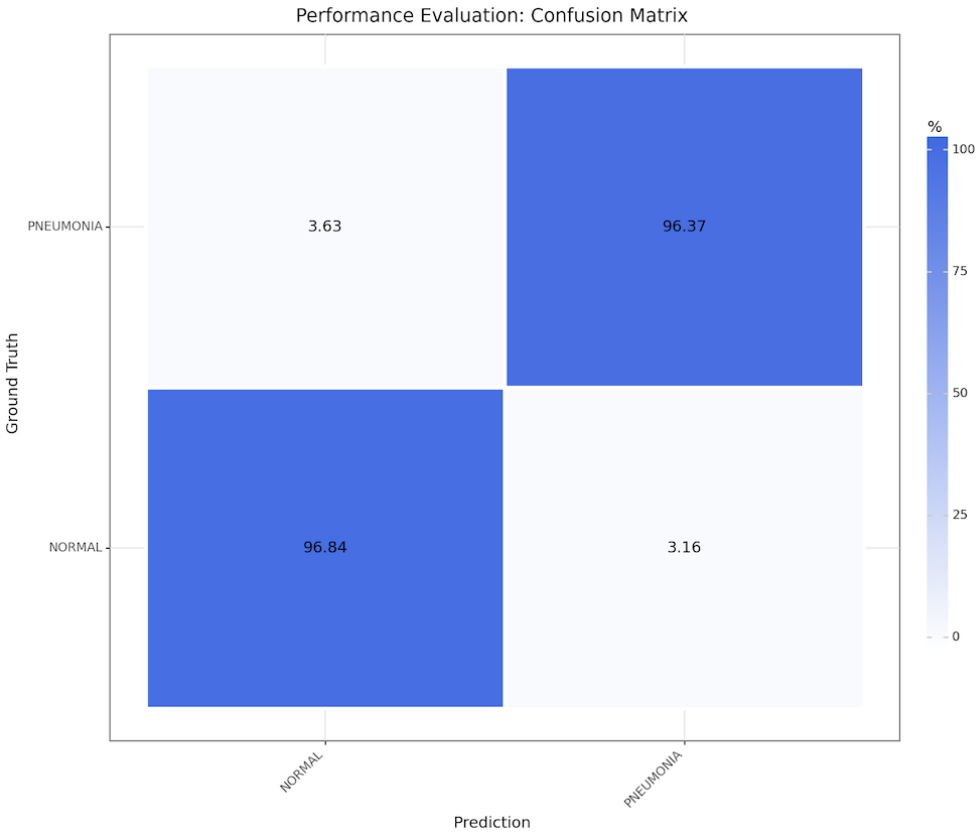


Figure 5: Confusion matrix: Chest X-Ray Images (Pneumonia)

5 XAI

Both samples were correctly classified with high confidence. However, the Grad-CAM image for sample **IM-0023-0001** shows, that not the lung, but the patient's clavicle seems to be highly relevant for our model.

Looking at sample **person1345 bacteria 3425**, we notice a similar phenomenon: Relevant to a classification as pneumonia is not the lung tissue. Instead, our model seems to recognize, that the patient has some sort of drainage on his or her left side. That may be an indicator of a patient's bad constitution, but is not, what we want the model zu learn. The second segment, that is marked in the Grad-CAM image is a region on the rib cage, close to the patient's (presumably ECG) sensor.

With that in mind, the highlighted clavicle in Grad-CAM image **IM-0023-0001** might point to the lack of an ECG sensor, hence suggesting a healthy patient. It seems our model was basing its decision on information, that is not a medically reliable predictor.

Nevertheless, the clearly visible sensors are not precisely marked in the Grad-CAM image **person1345 bacteria 3425**, which supports the argument, that our model does not entirely base the decision on medical instrumentation.

Image: IM-0023-0001

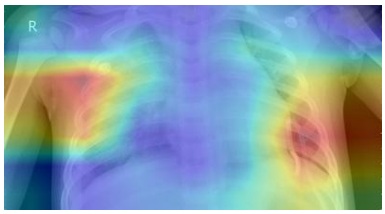
Class: NORMAL

Classified as: NORMAL (99.1 %)

Image: person1345_bacteria_3425

Class: PNEUMONIA

Classified as: PNEUMONIA (99.0 %)



6 Summary

The distinction between a healthy and sick patient is a valuable use case for machine learning. All calculated metrics attest to the model's great classification performance for this task.

However, Grad-CAM images for our two samples reveal, that the model may not be using medically relevant features for its decisions. Of course, more Grad-CAM images should be inspected to solidify that suspicion.