

# Image Classification with Deep Learning: Detecting Pneumonia in X-ray Images

Aysu Erdemir  
January, 2023



**Stakeholder:**

Be Well Healthcare Center, Radiology Department

---

**Business Problem:**

Looking for an efficient way to screen chest x-ray images for pneumonia detection in pediatric patients.

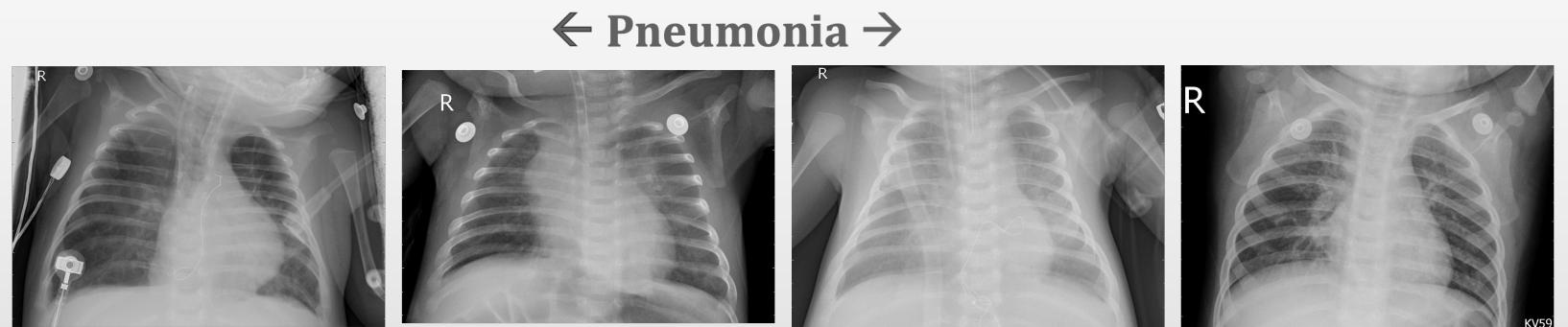
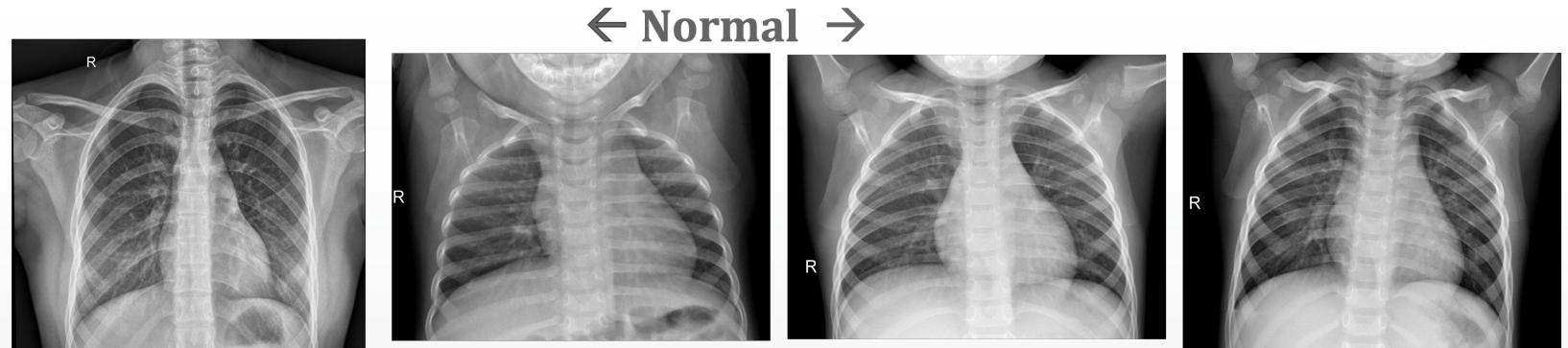
---

**Project Value:**

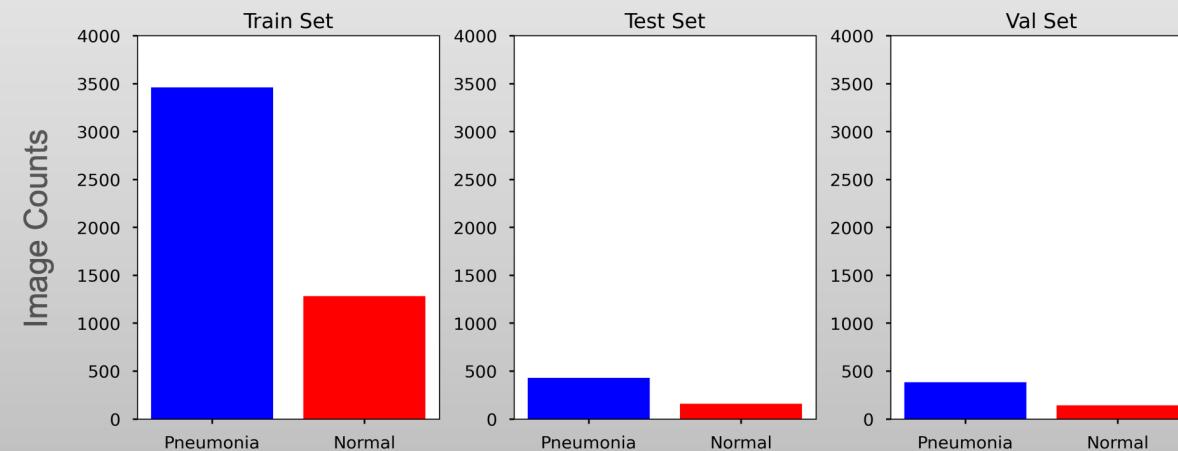
- Decrease the workload of its radiologists.
  - Minimize diagnosis time and allows for faster treatment.
  - Lower the risk of human errors and improve patient safety.
-

# Data:

Kaggle Dataset of **Chest X-ray images** from **5,863** pediatric patients.

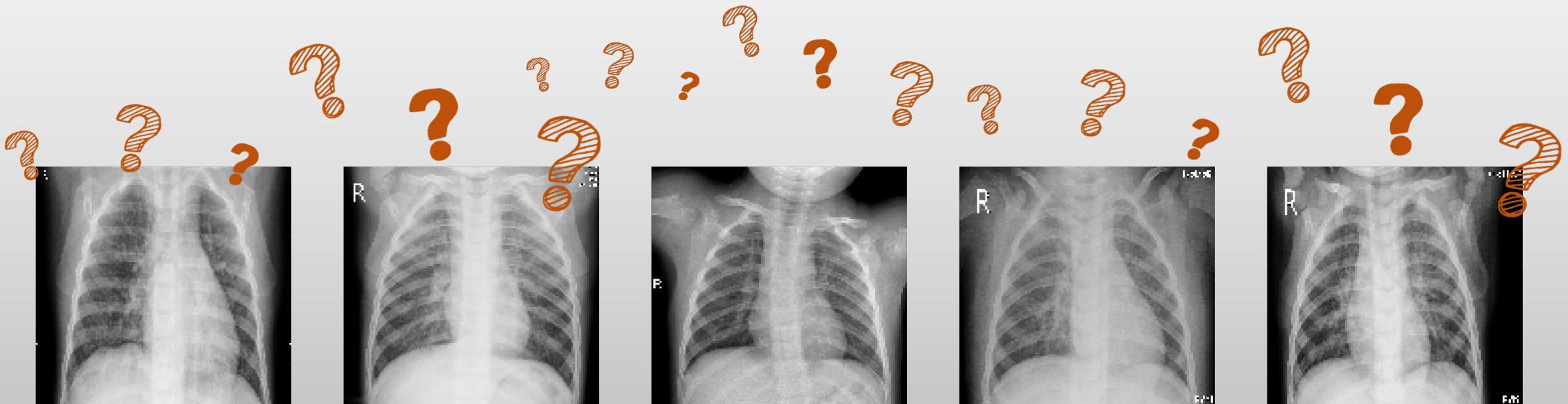


Data is separated by:  
**train** (80%), **test** (10%),  
**validation** (10%) sets and the  
categories they belong to  
(Pneumonia/Normal).

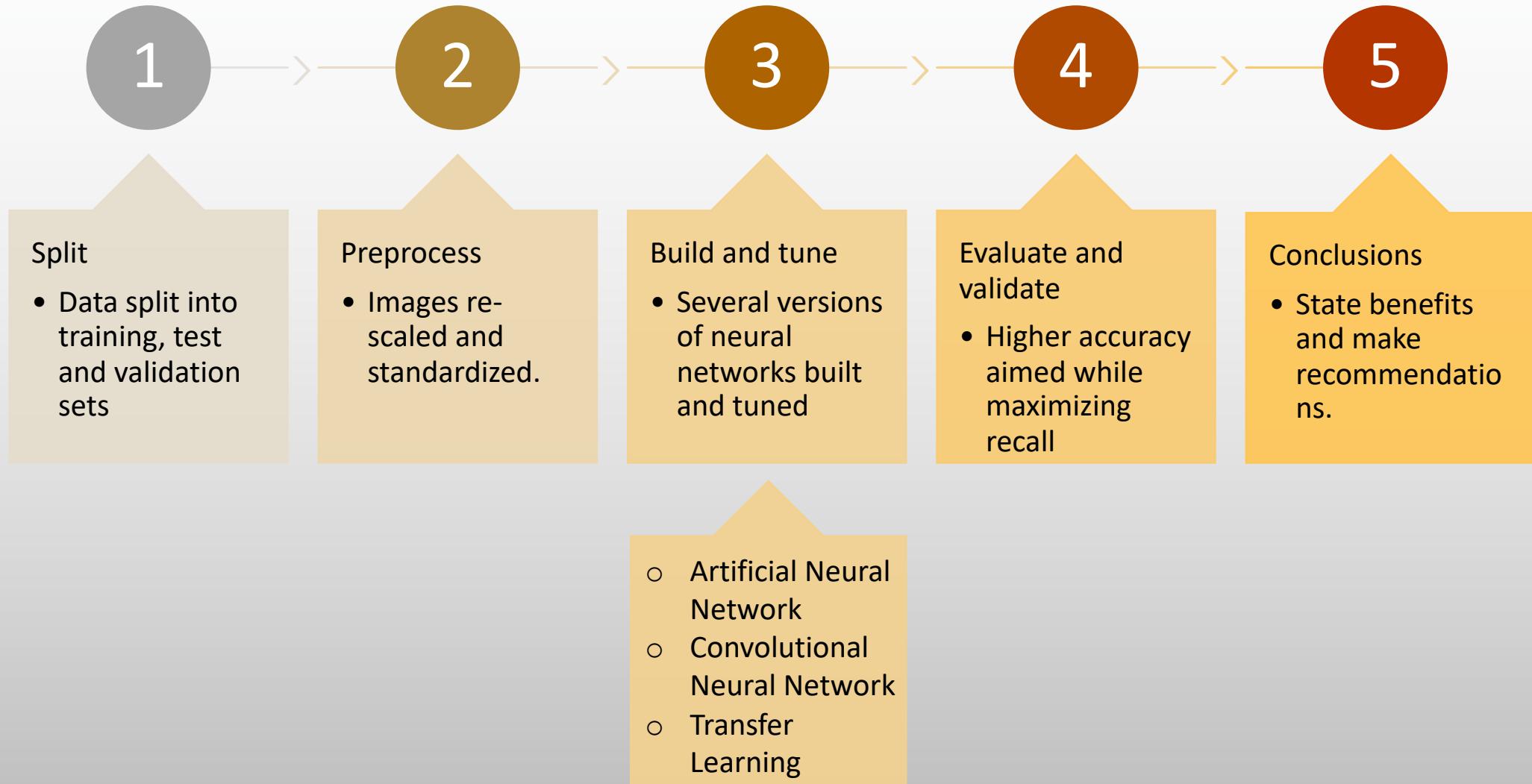


# Goal:

- Generate a **Neural Network Model** that classifies the X-ray images by detecting pneumonia.
- **Target Accuracy:** Predict the status of the lungs (**Normal vs pneumonia**) as accurately as possible.
- **Maximize Recall:** Identify majority of the **True Positive** cases so that we catch as many kids with pneumonia as possible.

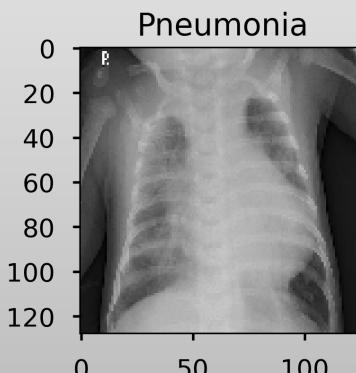
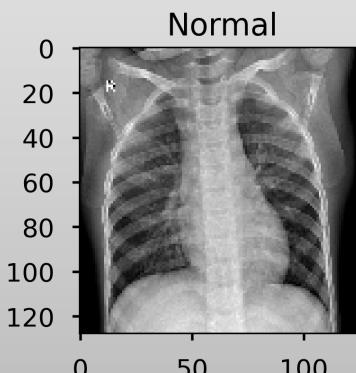
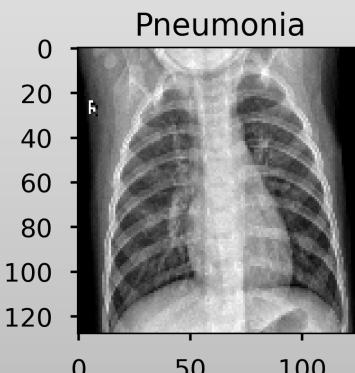
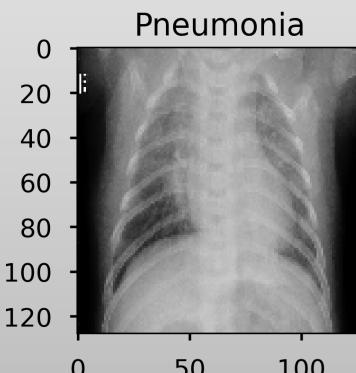
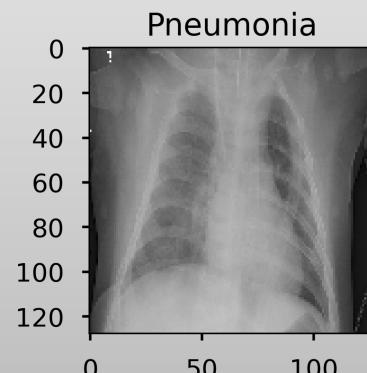
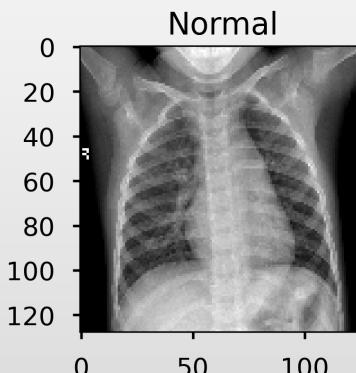
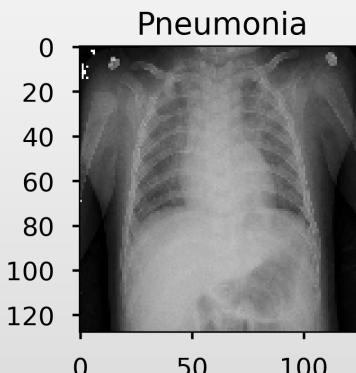
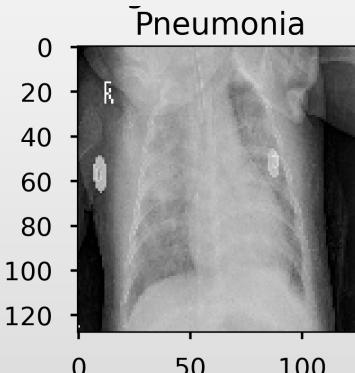
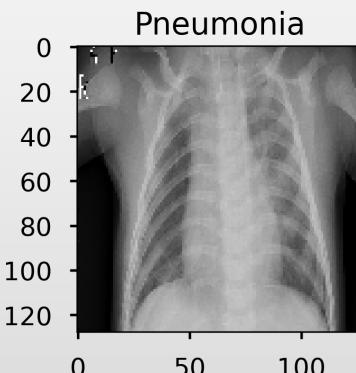
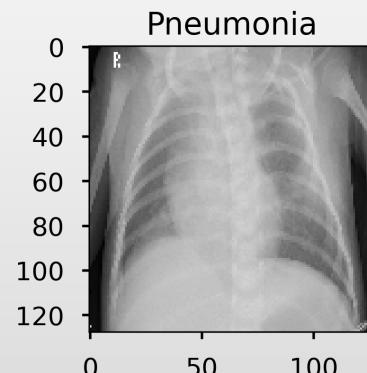


# Methodology:

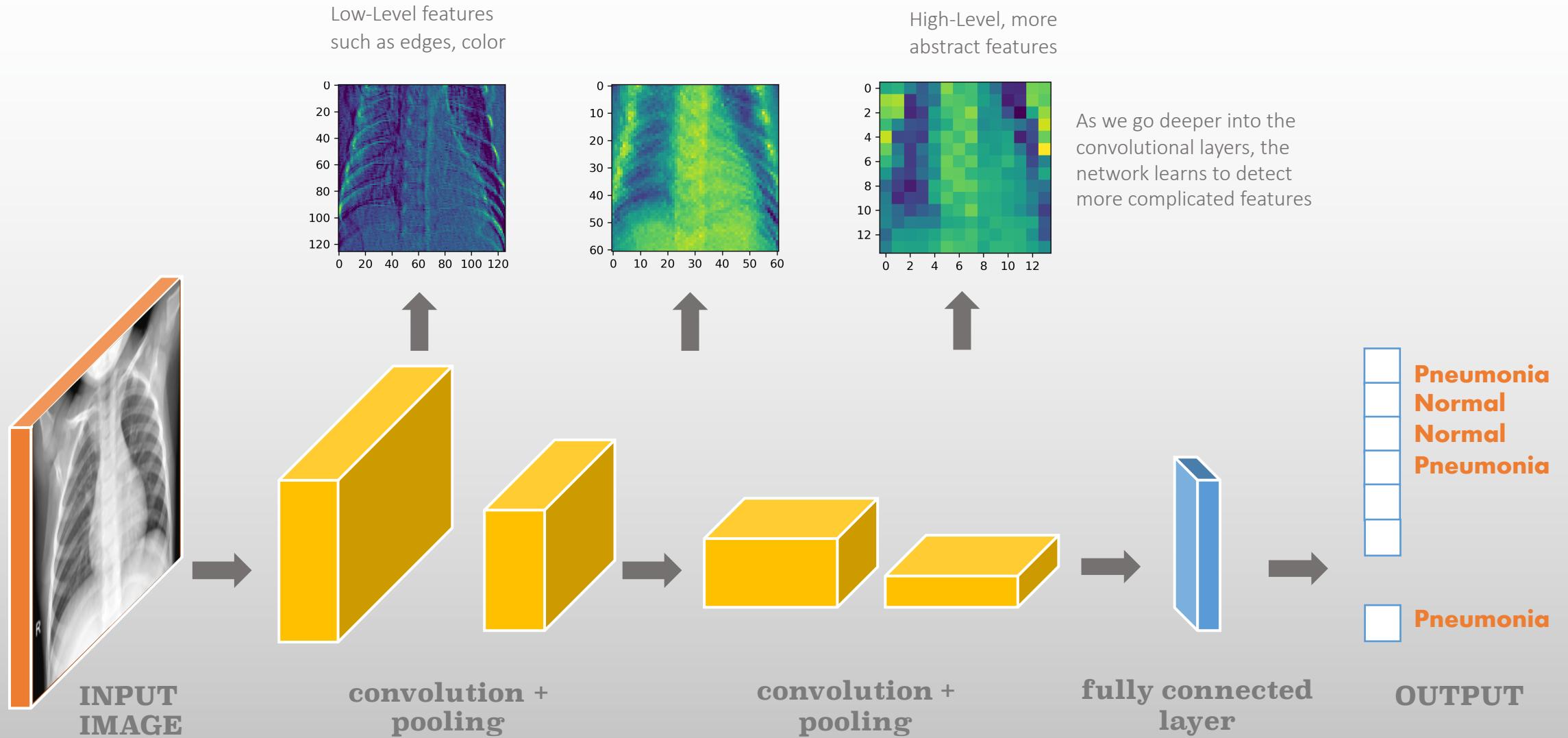


# Preprocess:

- ❑ All images are downsized to a size of 128 x 128 pixels
- ❑ Pixels values (0–255) normalized to 0–1



# Build and Tune:

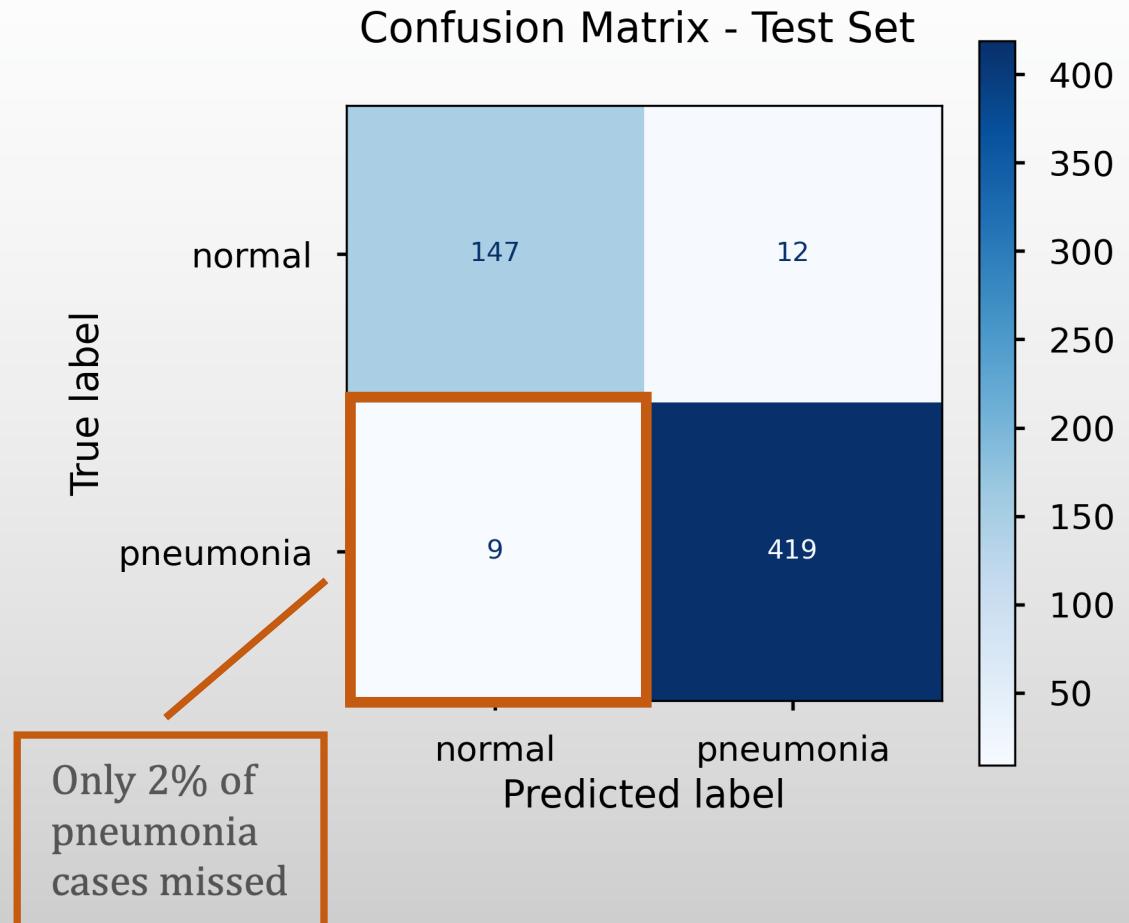


# Validate and Evaluate:

## Model Performance

	Pneumonia	Normal
Accuracy		96%
Precision	97%	94%
Recall <i>(True positive rate)</i>	98%	92%
f1-score	98%	93%

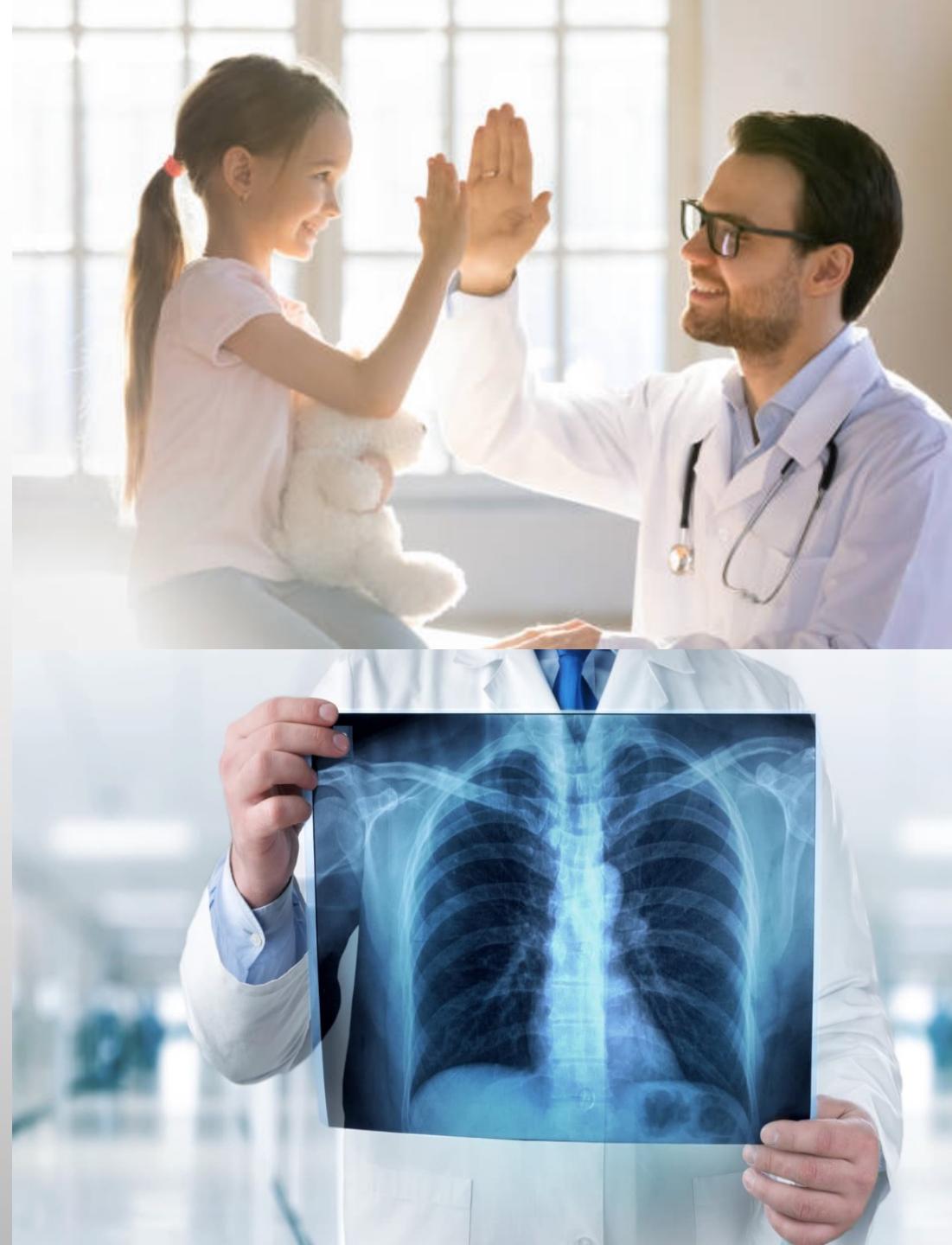
“Very Good” performance in all metrics.



# Recommendations:

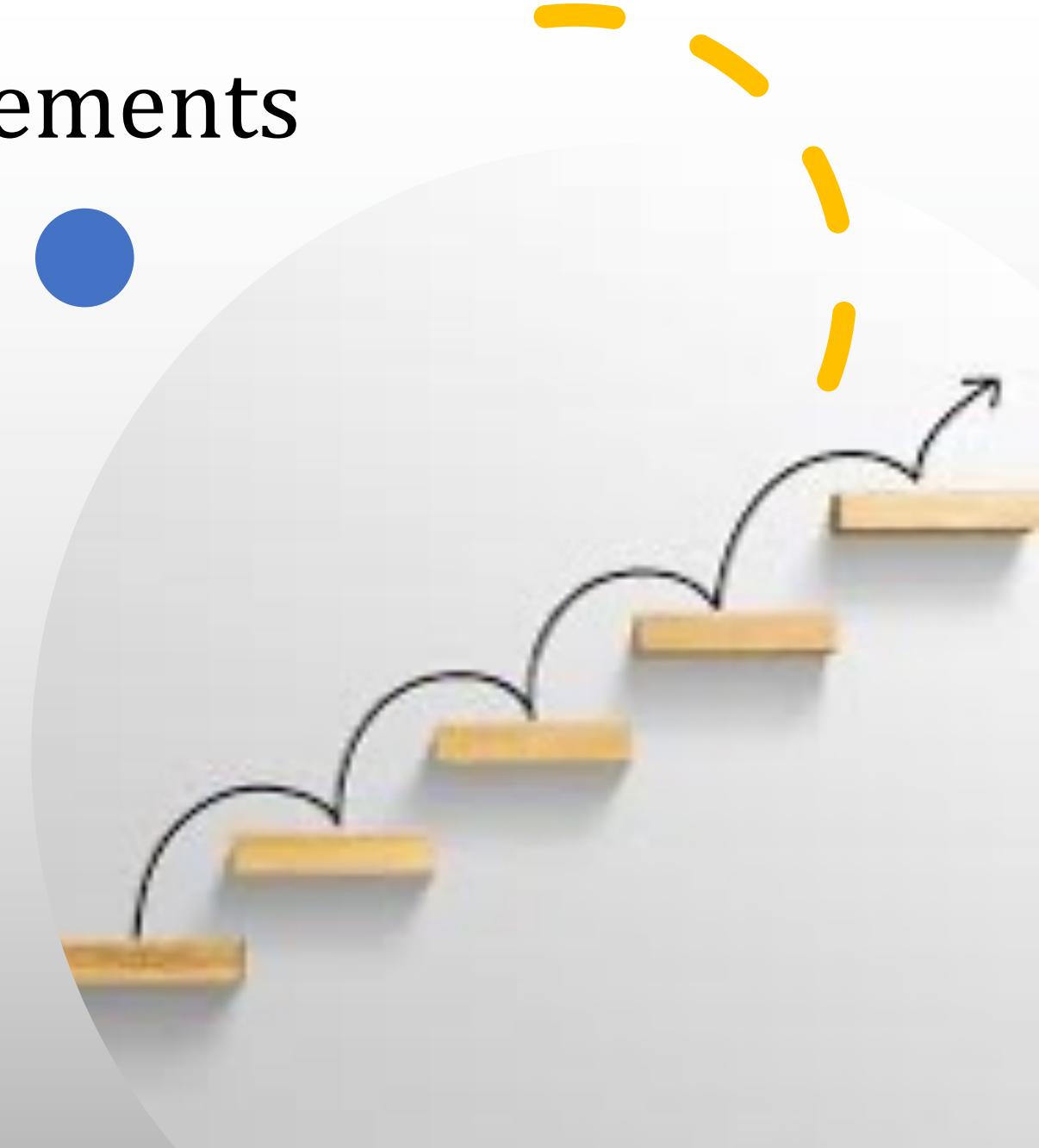
Use of CNN to classify x-ray images

- Decreased work-load by stream-lining the diagnosing process → *quicker return time and greater patient satisfaction.*
- Faster initiation of treatment for patients classified as high-risk.
- Doctors / radiologists can allocate **more time** to go over the images in the grey zone more rigorously or **for more demanding and complex procedures** in general.



# Limitations and Improvements

- We can use **data augmentation** methods to increase the size of the training set.
- We can **crop the images** to exclude the electrodes and letter R from the display which might be negatively affecting the image processing algorithm.
- We could address the **class imbalance** issue using **oversampling** techniques.



---

Thank You

Email: erdemiraysu@gmail.com

GitHub: @erdemiraysu

LinkedIn:

linkedin.com/in/aylsruerdemir/

