

---

---

# Pneumonia Binary Classifier

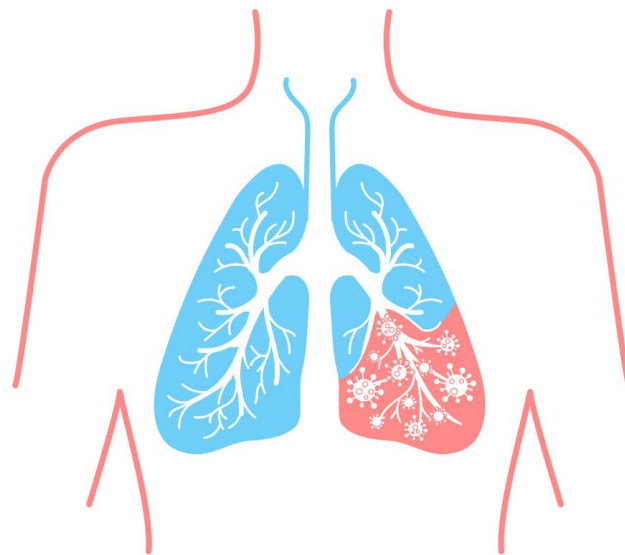
— Adriana Ramirez Franco —  
November 20th, 2024

---

---

# Summary

- Model that **detects** pneumonia from X-rays.
- Addresses a **real-world** healthcare challenge.
- Faster, **reliable diagnosis**.
- Ensure patient **safety**.

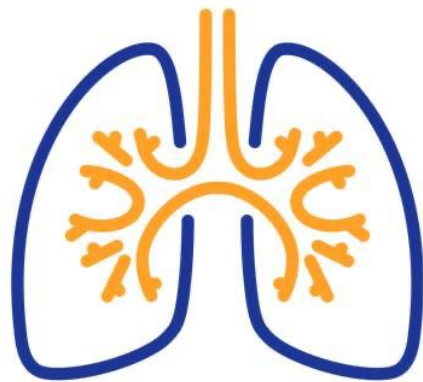


# Outline

- Goals
- Data
- Methods
- Conclusions

# Goals

- **Classify** X-ray images of patients with pneumonia vs. healthy individuals.
- Select the **best-performing** model that optimizes a high recall.
- The model **generalizes** well to new data for real-world application.



# Data

- **Kaggle** dataset from Guangzhou Women and Children's Medical Center.
- **Labeled** grayscale chest X-rays.
- **Split** training Validation, and Test sets.
- Presents **Imbalance** with ~3x more pneumonia cases than normal.

PNEUMONIA

person808\_bacteria\_2716.jpeg person1648\_bacteria\_4376.jpeg person1079\_bacteria\_3019.jpeg



NORMAL

NORMAL2-IM-1322-0001.jpeg



IM-0294-0001.jpeg



NORMAL2-IM-0839-0001.jpeg

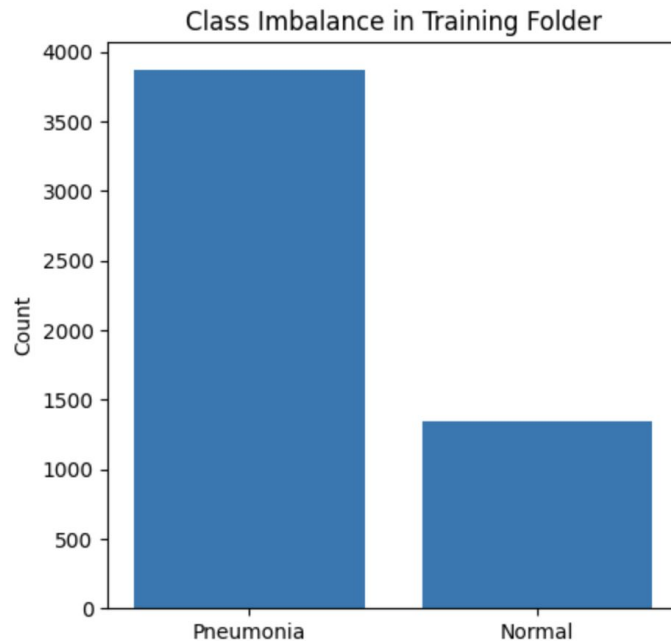


# Methods

## Class Imbalance

- 74% pneumonia images and 26% normal lung images.
- Leads to challenges in training.

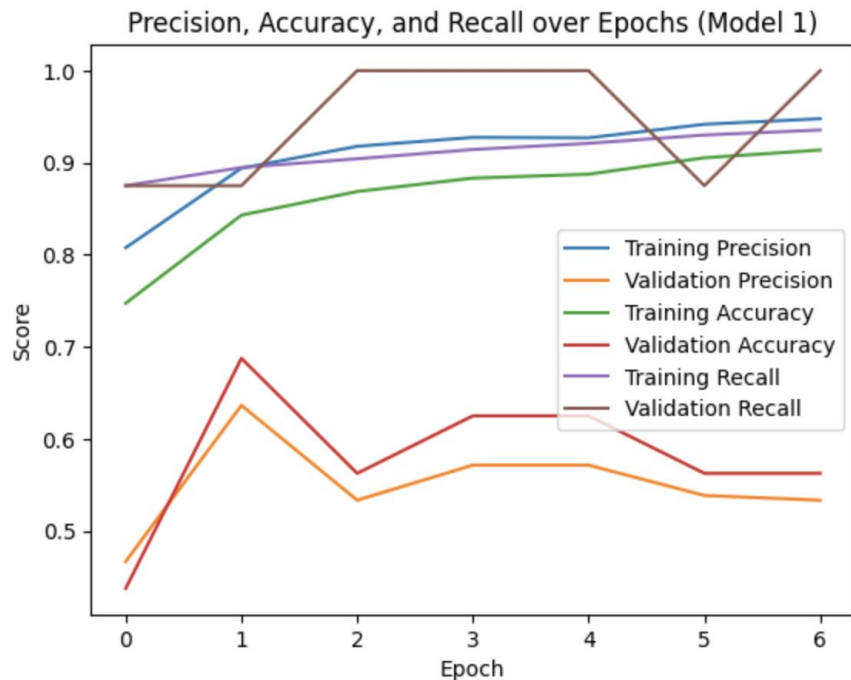
**Solution:** Image data generator to augment the training set.



# Methods

## Model 1:

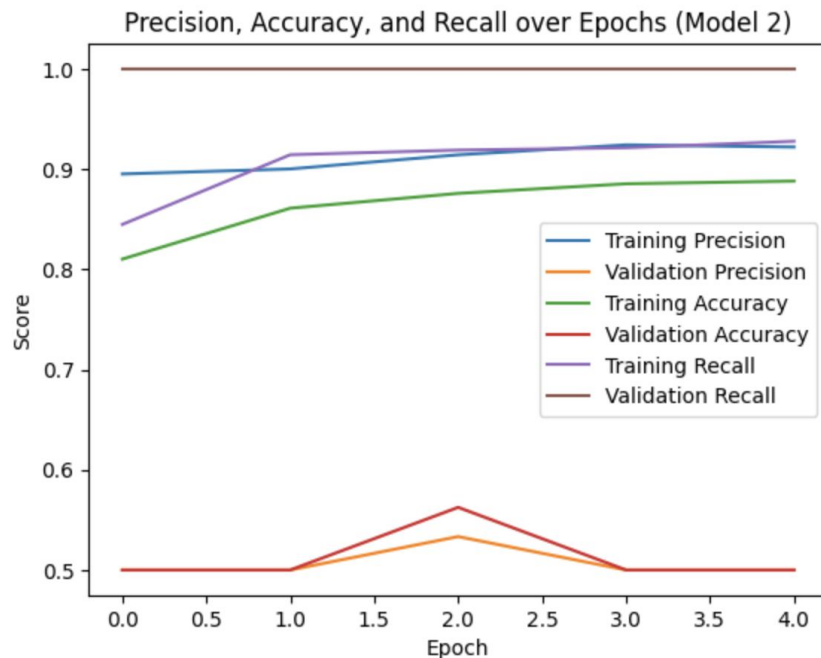
- Basic CNN Architecture.
  - **Precision:** 0.90
  - **Recall:** 0.86
- **Recall** was suboptimal, indicating that some pneumonia cases were missed.



# Methods

## Model 2:

- Enhanced Architecture.
  - **Recall:** 0.99 (maximized)
  - **Precision:** 0.69
- **Low precision, resulting in poor balance between identifying pneumonia and avoiding false positives.**

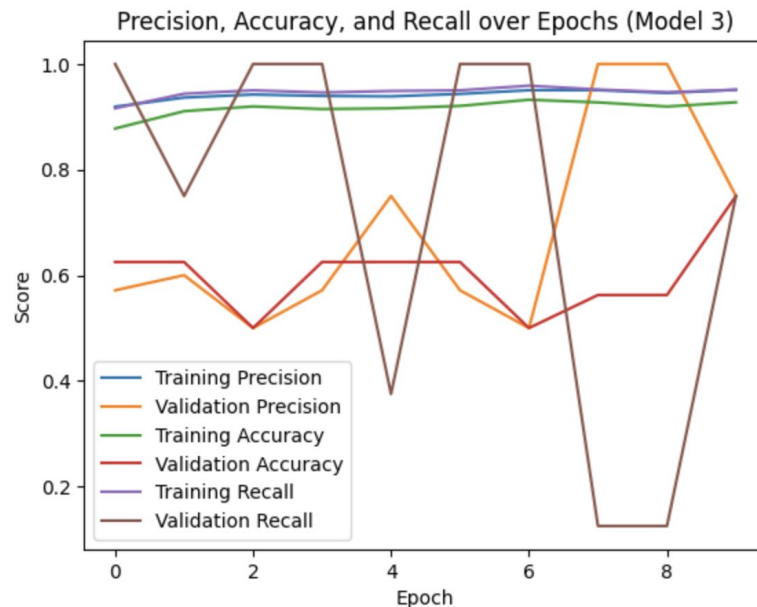




# Methods

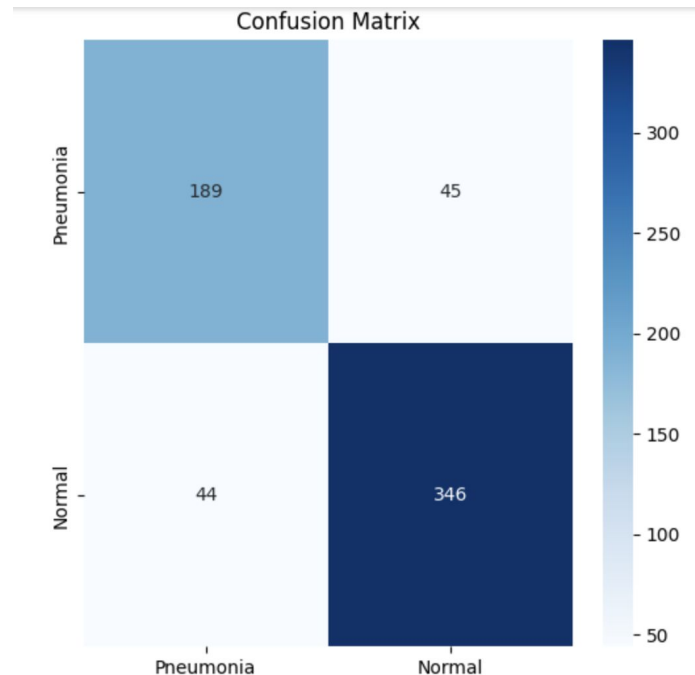
## Model 3:

- Optimized CNN.
  - **Precision:** 0.87
  - **Recall:** 0.94
  - **F1 Score:** 0.90
- Model 3 achieved the best balance.
- Most suitable for potential clinical applications.



# Conclusions

- Model 3 **minimized** false negatives while maintaining accuracy.
- The model achieves an **accuracy** of approximately 85.7%.
  - 189 true positives
  - 346 true negatives.
- **44** false negatives highlights an area for improvement.



# Thank you!

Email: [aramirezfr20@gmail.com](mailto:aramirezfr20@gmail.com)

Github: [@aramirezfr](https://github.com/aramirezfr)