IMAGE CLASSIFICATION OF X-RAY IMAGES FOR PNEUMONIA DETECTION

PHASE 4 GROUP 5 MEMBERS

Lisa Mwikali

Nicole Bosibori

Maureen Muriithi

Charles Egambi

Ivan Wawire

Anne Njoroge

INTRODUCTION

Pneumonia is a common and potentially fatal lung infection. Accurate diagnosis is essential for effective treatment and patient management. The "Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images" is a comprehensive collection of medical images, aiming to support advancements in medical diagnosis through deep learning.

BUSINESS PROBLEM

In healthcare, accurate and efficient disease diagnosis is vital. Traditional methods of diagnosing pneumonia involve lengthy exams and lab tests, often requiring multiple doctor visits. This project aims to streamline this process using a deep learning model to detect pneumonia from chest x-ray images, providing faster and more precise diagnoses.

OBJECTIVES

Main Objective

To develop a deep learning model to classify chest x-ray images for pneumonia detection.

Specific Objectives

- To determine the most effective architecture for pneumonia detection.
- To train the deep learning model.
- Assess the trained model's performance to validate its effectiveness in detecting pneumonia.

STAKEHOLDERS

- Health Professionals
- Patients
- Radiologists
- Medical Researchers

DATA OVERVIEW

Source: Mendley Data

Data Organization: The images are categorized into "train", "test" and "val" folders

Categories: The images can be classified into two ; "Pneumonia" and "Normal"

Patient Demographics: Ages 1-5years, Guangzhou Women and Children's Medical Center

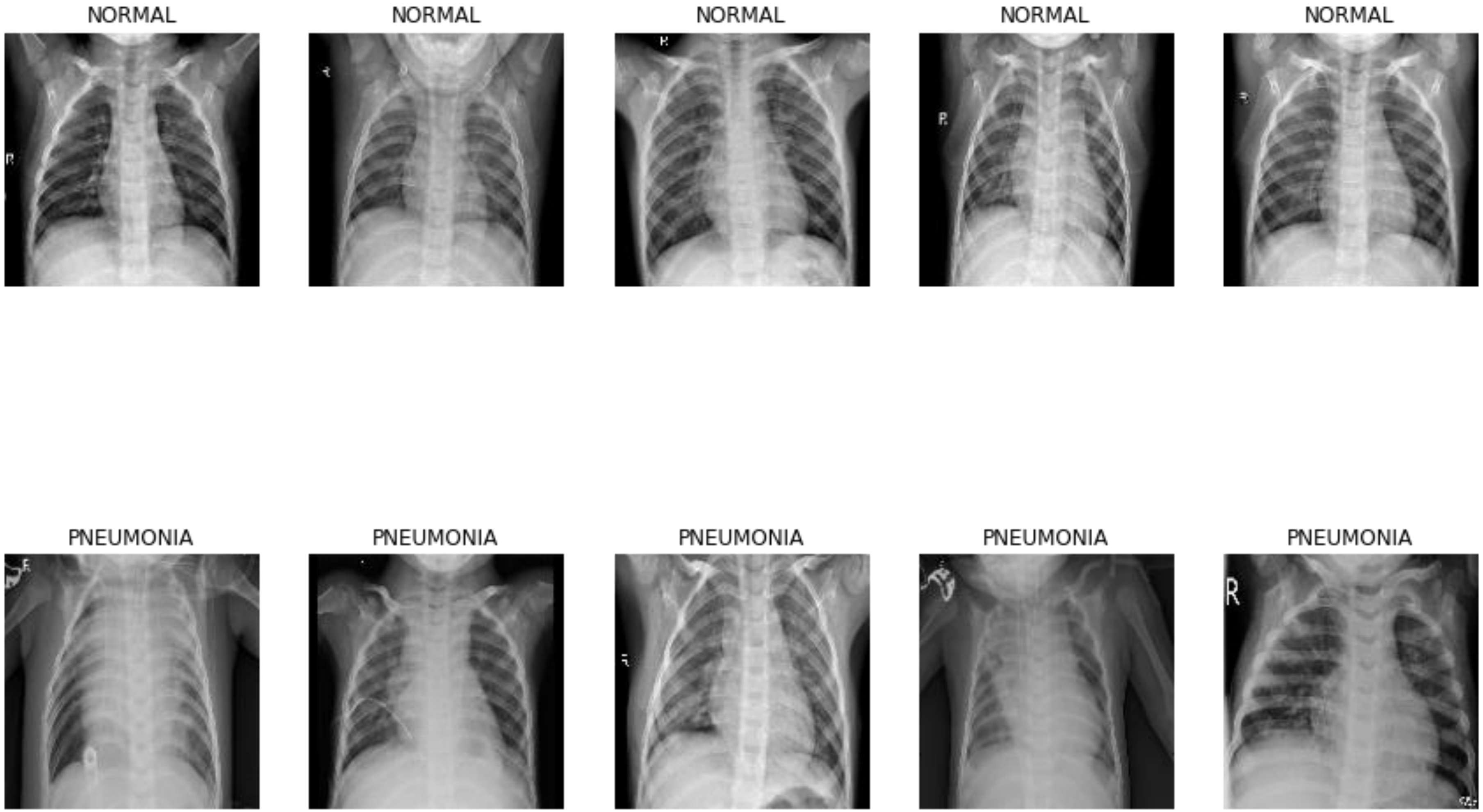
Image Details: Anterior – posterior chest X-rays, part of routine care

METRIC OF SUCCESS

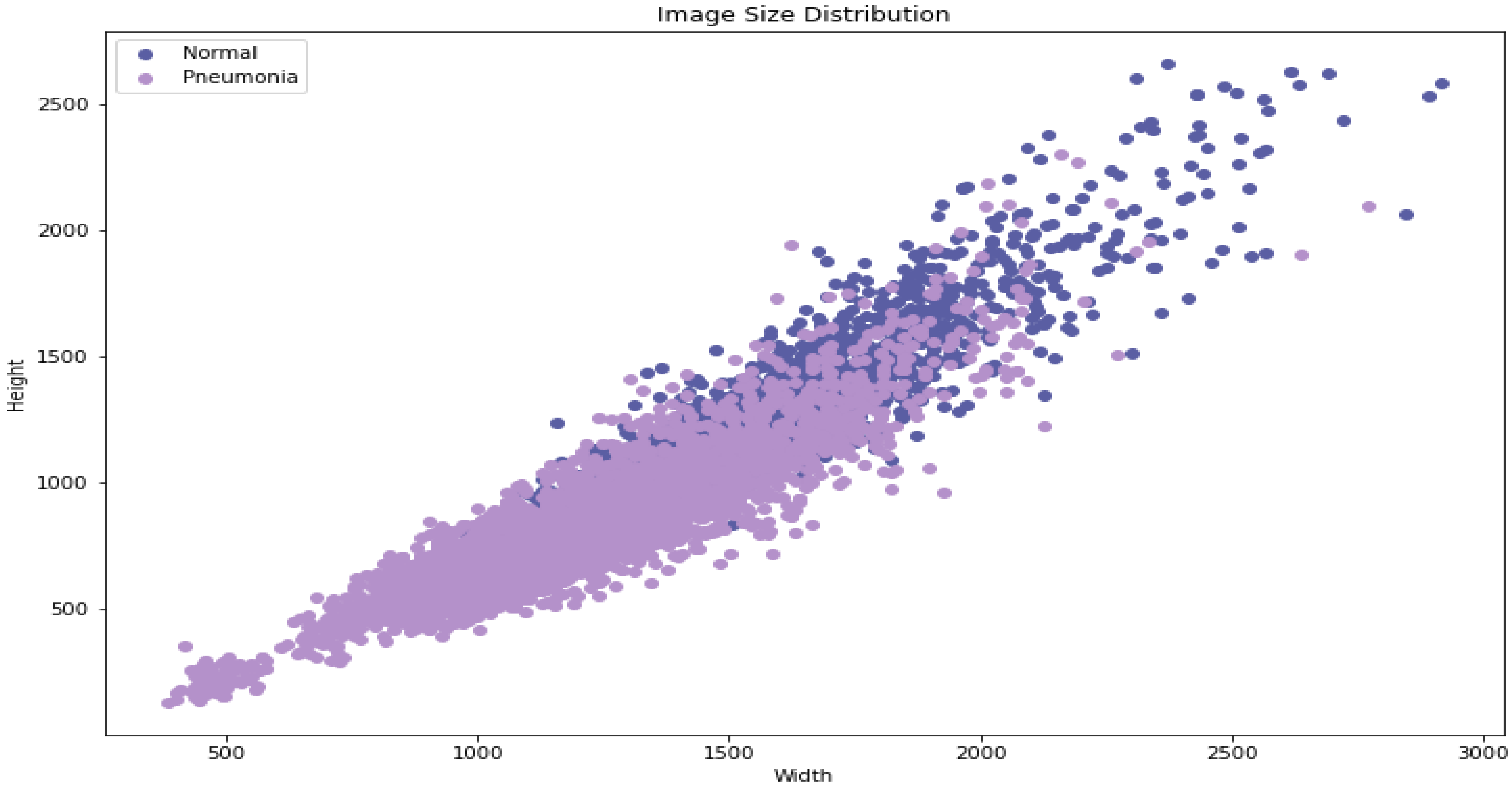
The performance of the models will be evaluated using the following metrics:

- Loss: This metric measures the error between the predicted values and the actual values. Lower loss indicates a better fitting model. We will use the test loss to evaluate the model's ability to generalize to new data.
- Accuracy: This metric indicates the proportion of correctly classified instances out of the total instances. Higher accuracy indicates better model performance in terms of classification correctness.

SAMPLE IMAGES TO DISTINGUISH BETWEEN NORMAL & PNEUMONIA CASES

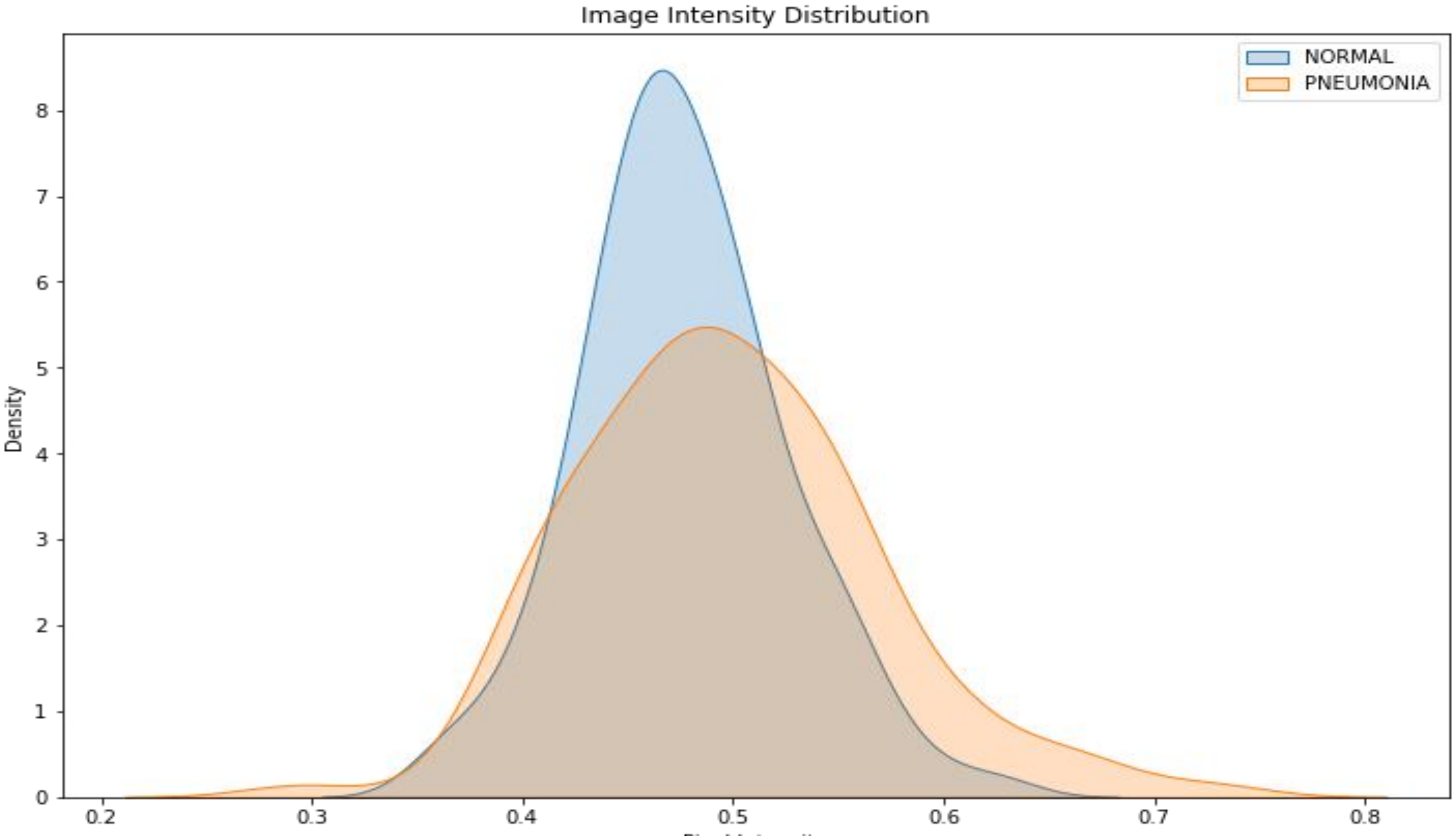


SCATTER PLOT DISPLAYING IMAGE SIZE DISTRIBUTION



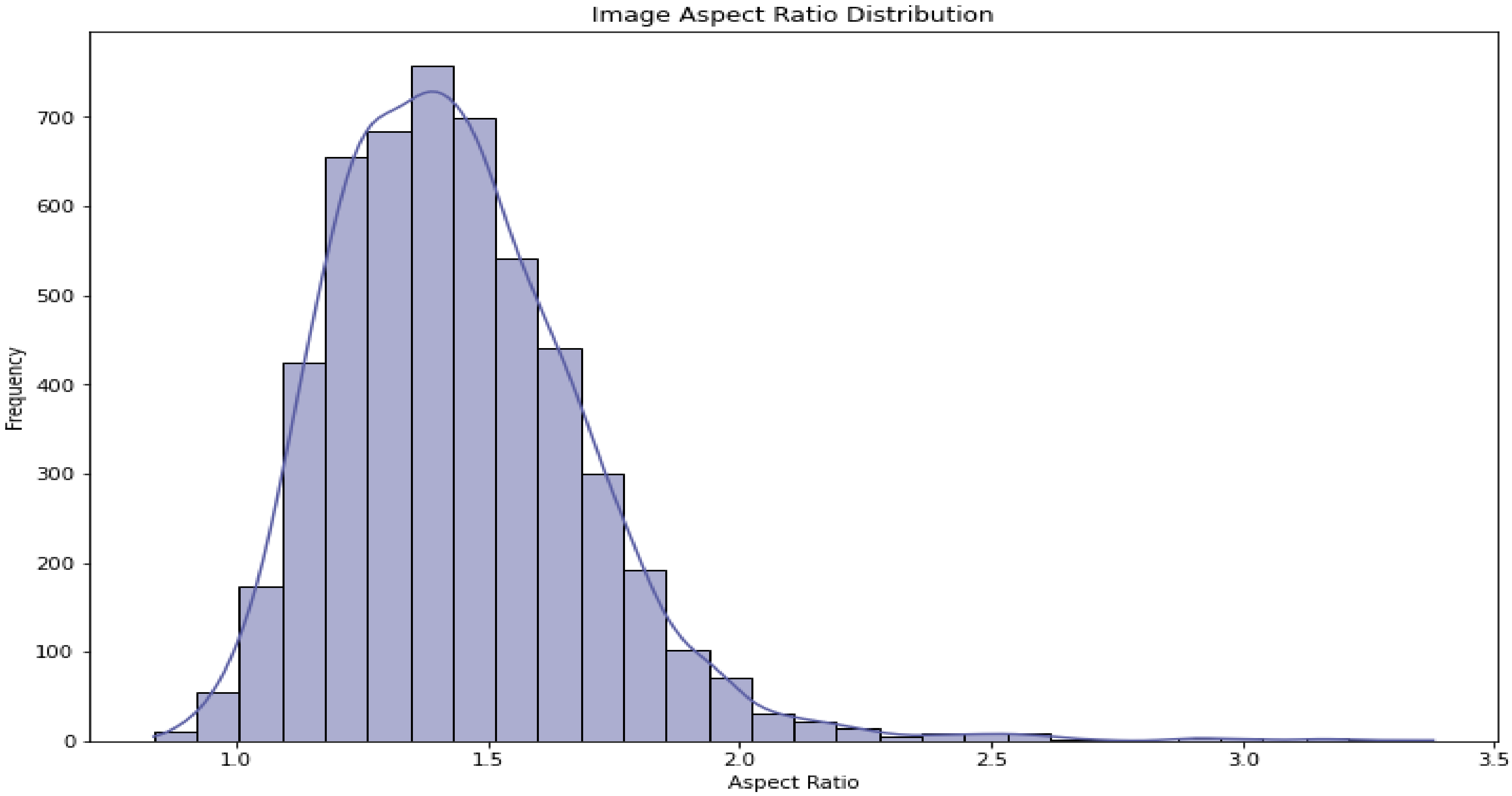
Shows the variation in dimensions of the x-ray images, which was important for preprocessing steps

IMAGE INTENSITY DISTRIBUTION



Normal and pneumonia images have similar peak pixel intensities with significant overlap, but pneumonia images exhibit a broader distribution, indicating greater variability in pixel values possibly due to varying degrees of infection or different visual characteristics of the disease.

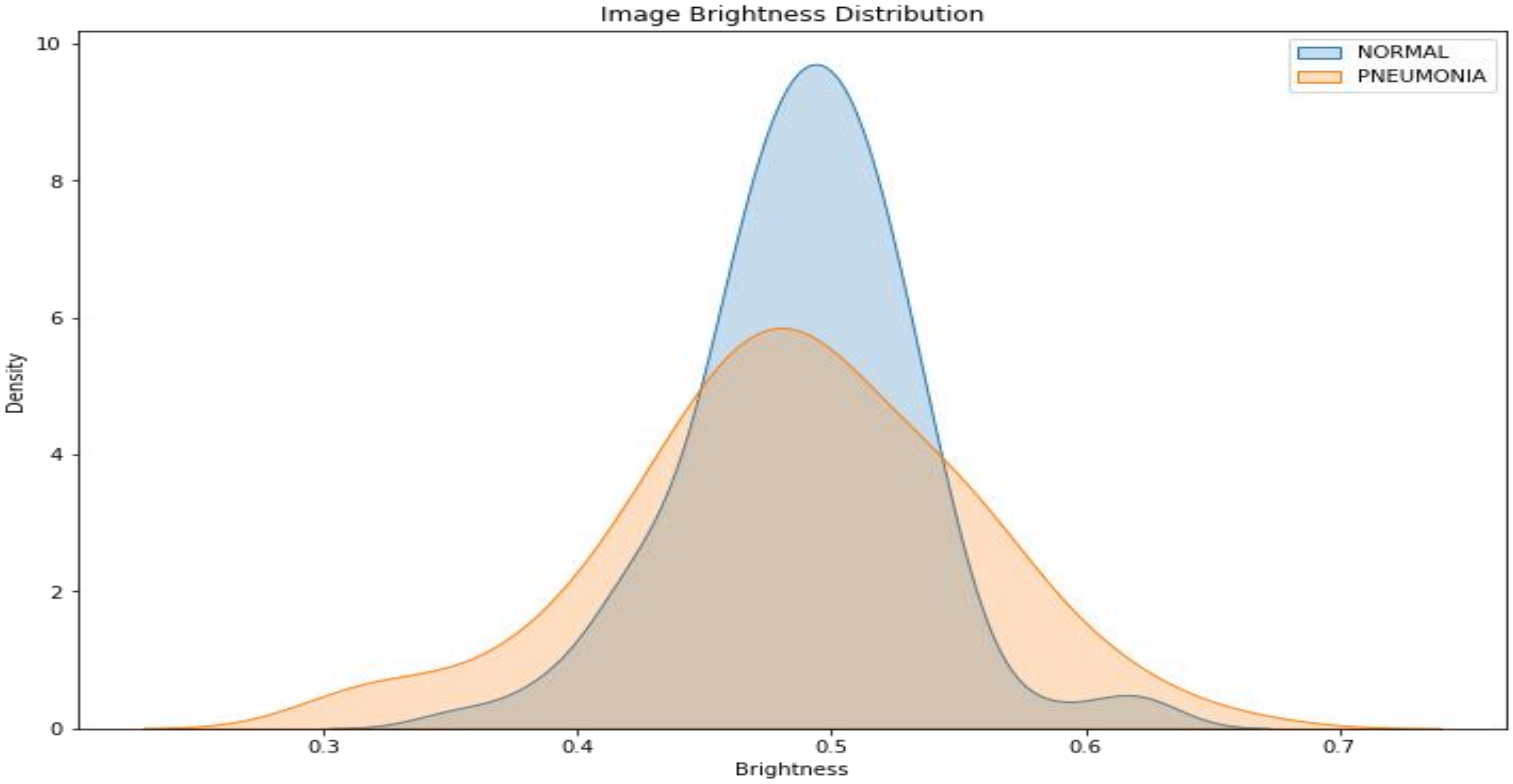
HISTOGRAM SHOWING IMAGE ASPECT RATIO DISTRIBUTION



Shows the shape and proportion of the images.

The aspect ratio distribution indicates that most images have an aspect ratio between 1.0 and 1.5, with a peak around 1.3.

IMAGE BRIGHTNESS DISTRIBUTION



Normal and pneumonia images have overlapping brightness distributions, with pneumonia images showing a wider spread and a lower peak brightness, indicating that normal images are generally brighter on average.

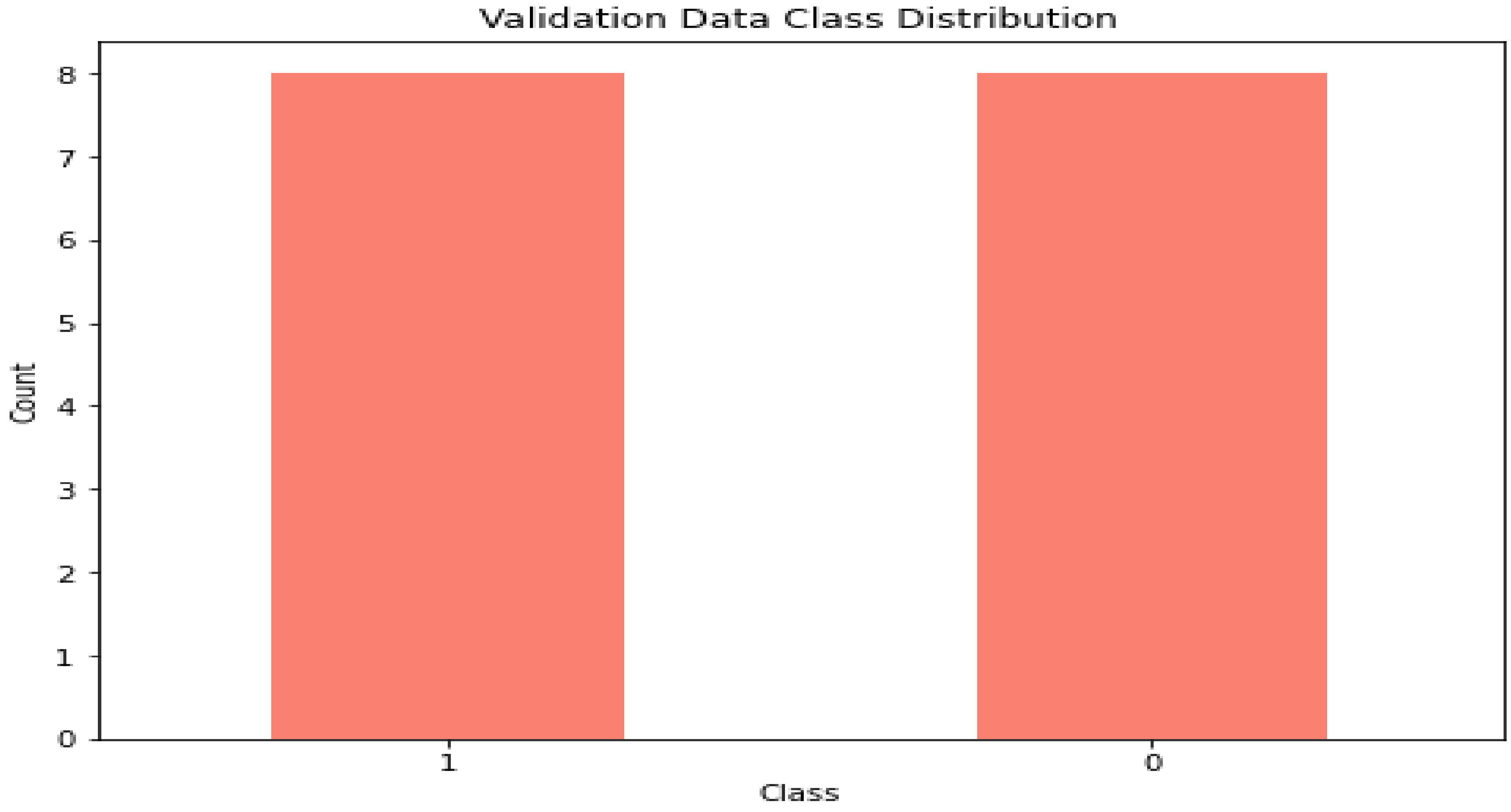
BAR GRAPH DISPLAYING TRAINING DATA CLASS DISTRIBUTION

Show the number of images in the dataset for normal(0) vs pneumonia(1), with pneumonia being the highest

BAR GRAPH DISPLAYING TEST DATA CLASS DISTRIBUTION

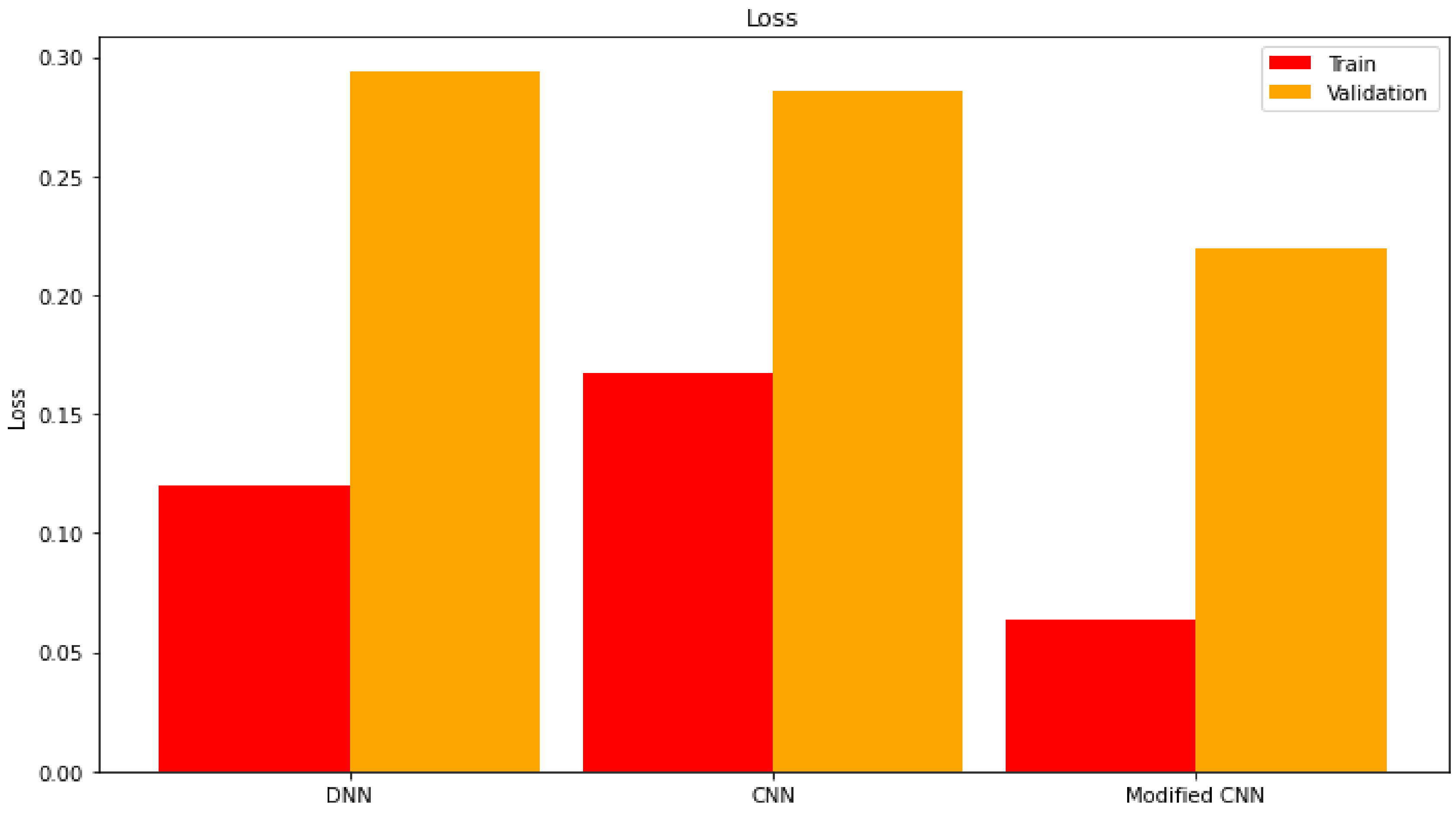
Show the number of images in the dataset for normal(0) vs pneumonia(1), with pneumonia being the highest

BAR GRAPH DISPLAYING VALIDATION DATA CLASS DISTRIBUTION



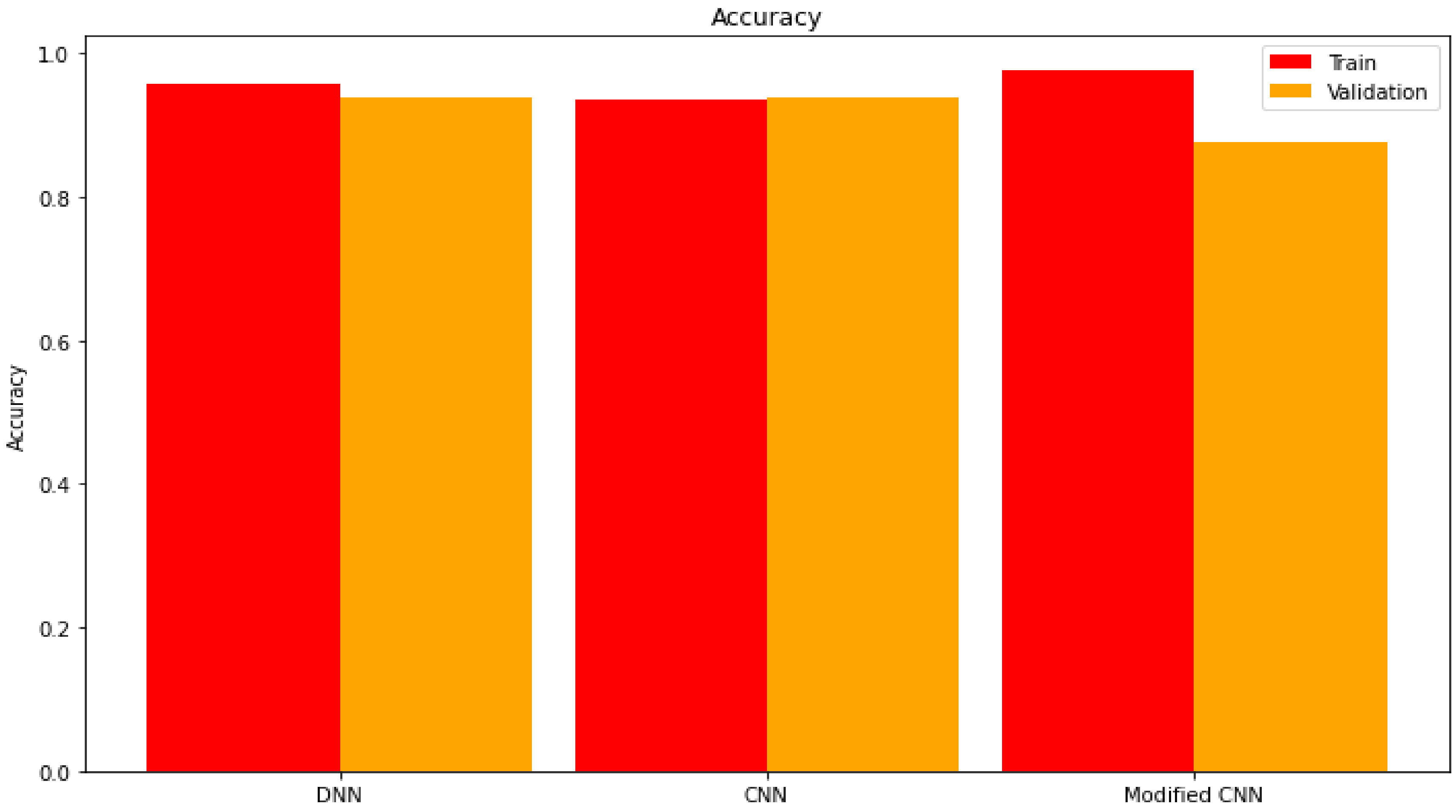
Show the number of images in the dataset for normal(0) vs pneumonia(1), with pneumonia and normal being equal

MODEL EVALUATION



The bar chart compares the performance of the models across training and validation metrics, showing that the Modified CNN model performed better since it had the least loss.

MODEL EVALUATION



The bar chart compares the performance of the models across training and validation metrics, showing that the Modified CNN model performed better since it had the best accuracy levels.

RECOMMENDATIONS

- Regularization: Prevent overfitting and enhance generalization.
- Hyperparameter Tuning: Optimize model configurations for peak performance.
- Ensemble Methods: Combine diverse models for improved accuracy and robustness.
- Cross-Validation: Rigorously assess model performance across multiple data splits.
- Data Augmentation: Expose models to diverse data to improve real-world capabilities.

NEXT STEPS

- Severity Assessment: Beyond identifying pneumonia's presence, the AI model can determine its severity (mild, moderate, severe), providing clinicians with crucial information for treatment decisions.
- Pneumonia Type Identification: Differentiating between bacterial and viral pneumonia can guide appropriate treatment plans, as antibiotics are only effective for bacterial infections.
- Precise Localization: The model can pinpoint the specific lung areas affected by pneumonia, aiding further investigation and targeted treatment.
- Cloud Accessibility: Deploying the model on a cloud platform ensures widespread access for hospitals and clinics, facilitating broader utilization and impact.