# Title: Pneumonia Detection Using Chest X-rays

Team: 04

- Srirama Murthy Chellu (SC23BK)
- Sameera Rompicherla (SR23BA)
- Yashwanth Gowram (YG23G)

Course : Data Science Meets Health Science

Date : 5/1/2024

Instructor : Olmo Zavala Romero

## Abstract

Pneumonia is a prevalent respiratory infection caused by bacteria or viruses, impacting a significant portion of the global population, especially in regions with limited healthcare resources and environmental challenges such as pollution and overcrowding. This condition often leads to pleural effusion, a complication where fluids accumulate in the lungs, causing breathing difficulties. Early and accurate diagnosis of pneumonia is critical for timely treatment and improving patient outcomes.

In this study, we address the research question of automating pneumonia detection using chest X-ray images through a computer-aided diagnosis system. Our methodology involves developing and training a Convolutional Neural Network (CNN) architecture, a deep learning model specialized in image analysis tasks. The CNN model is trained on a dataset of 5,863 chest X-ray images, learning to identify patterns and features indicative of pneumonia.

The key findings of our research include the successful development of a CNN-based computer-aided diagnosis system capable of accurately detecting pneumonia from chest X-ray images. Through rigorous evaluation and validation, we demonstrate the system's high diagnostic accuracy and reliability compared to manual interpretation.

Our study concludes that leveraging machine learning and deep learning techniques, such as CNNs, can significantly enhance pneumonia diagnosis by reducing the reliance on subjective human interpretation and improving diagnostic consistency. The automation of pneumonia detection using advanced technology holds promise for improving healthcare outcomes, particularly in resource-constrained settings.

# Table of Contents



# List of Figures and Tables

# 1.Introduction

Pneumonia, a prevalent respiratory infection affecting the bronchi, alveoli, and interstitial lungs, poses significant health risks worldwide. Its etiology primarily includes viral and bacterial pathogens, leading to inflammation and fluid accumulation in the lungs. This diversity in presentation and potential complications such as pleural effusion, consolidation, and atelectasis makes pneumonia a challenging condition for medical imaging analysis.

## Importance

In the context of medical science, the accurate and timely diagnosis of pneumonia is crucial for effective treatment and improved patient outcomes. Misdiagnosis or delayed diagnosis can lead to serious complications, including respiratory failure and sepsis, highlighting the critical need for reliable diagnostic tools.

## Objectives:

The primary objective of our study is to develop a computer-aided diagnosis system for automatic pneumonia detection using chest X-ray images. This system aims to enhance diagnostic accuracy, reduce human error, and facilitate faster diagnoses, especially in settings with limited access to expert radiologists.

## Deep learning Approach:

Our approach leverages deep learning, specifically a Convolutional Neural Network (CNN) architecture, known for its ability to extract intricate patterns and features from images. The CNN is trained on a large dataset of 5600 chest X-ray images, learning to distinguish between normal lung structures and pneumonia-affected areas.

## Effects on the Lung and Damages:

Pneumonia's impact on the lungs includes inflammation, fluid accumulation, and potential damage to bronchi, alveoli, and interstitial tissues. These effects can lead to respiratory difficulties and impaired gas exchange

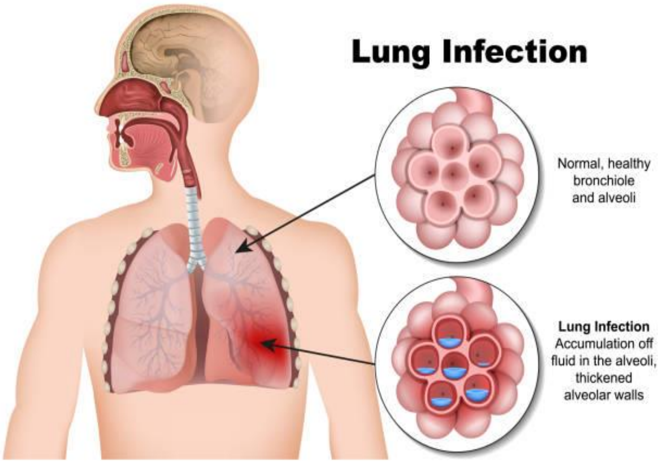



Figure 1. LUNG INFORMATION

# 2.Literature Review

## Summary of Previous Work in Medical Imaging:

Prior research in medical imaging has focused on various aspects of pneumonia diagnosis using chest X-ray images. Studies have explored traditional machine learning algorithms, such as support vector machines (SVMs) and random forests, for pneumonia detection. Additionally, efforts have been made to improve the accuracy of pneumonia diagnosis through feature engineering and pattern recognition techniques.

## Review of Deep Learning Methods in Similar Contexts:

Recent advancements in deep learning have revolutionized medical imaging analysis, particularly in pneumonia detection. Convolutional Neural Networks (CNNs) have emerged as a powerful tool for automatic feature extraction and classification in chest X-ray images. Several studies have demonstrated the effectiveness of CNNs in accurately identifying pneumonia-affected areas and distinguishing them from normal lung structures.

## Identification of Gaps in Existing Research:

Despite the progress in medical imaging and deep learning methods for pneumonia detection, several gaps remain in the existing research. These include the need for:

- Improved interpretability of deep learning models for clinical decision-making.
- Robustness and generalization of CNNs across diverse patient populations and imaging conditions.

- Integration of advanced image augmentation and preprocessing techniques to enhance model performance.
  Validation and deployment of computer-aided diagnosis systems in real-world clinical settings to assess their impact on healthcare outcomes.

# 3. Methodology

## 3.1 Data Collection:

- Overview: The dataset includes 5,863 paediatric chest X-ray images in JPEG format, categorized into two classes: Normal and Pneumonia.
- Source: Sourced from Guangzhou Women and Children's Medical Center, Guangzhou, the images represent anterior-posterior X-rays from children aged one to five.
- Structure: Organized into three directories: train, test, and validation, each containing specific subfolders for the image categories.
- Distribution:
- Training: 1,313 Normal and 3,847 Pneumonia images.
- Testing: 234 Normal and 390 Pneumonia images.
- Validation: 36 Normal and 36 Pneumonia images.
- Clinical Integration: All imaging was conducted as part of routine clinical care, ensuring real-world applicability.
- Access: The dataset is available for research on Kaggle (https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia)

Train DataFrame shape: (5160, 2)

Test DataFrame shape: (624, 2)

Validation DataFrame shape: (72, 2)

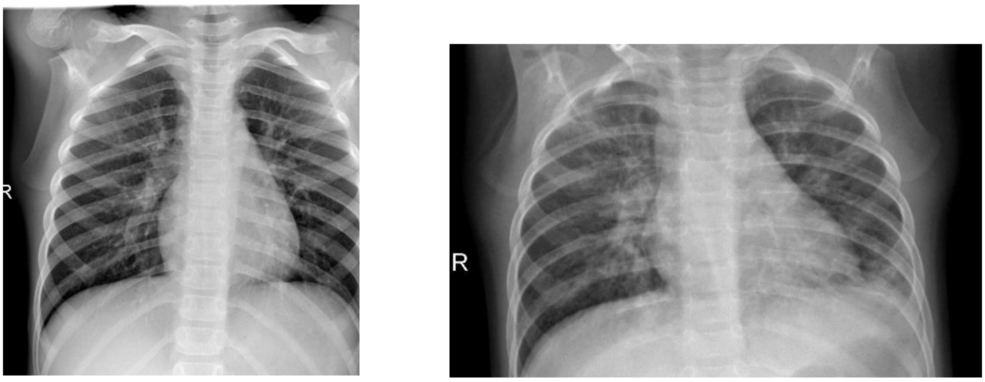



Figure 2. Normal Chest XRAY & Pneumonia Chest XRAY

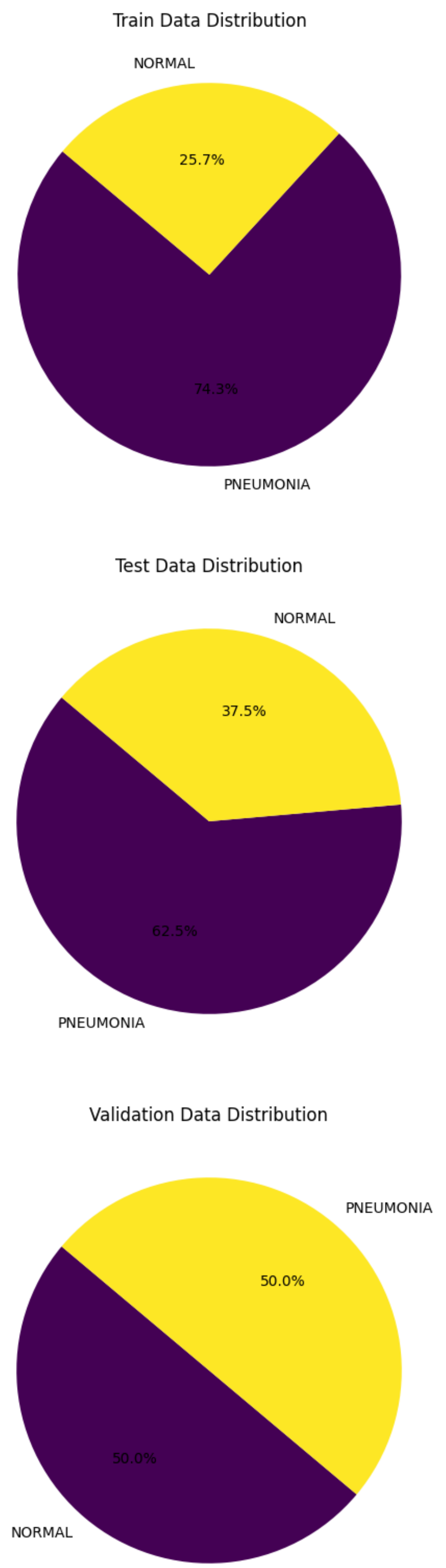



Figure 3.Data visualization on pie charts

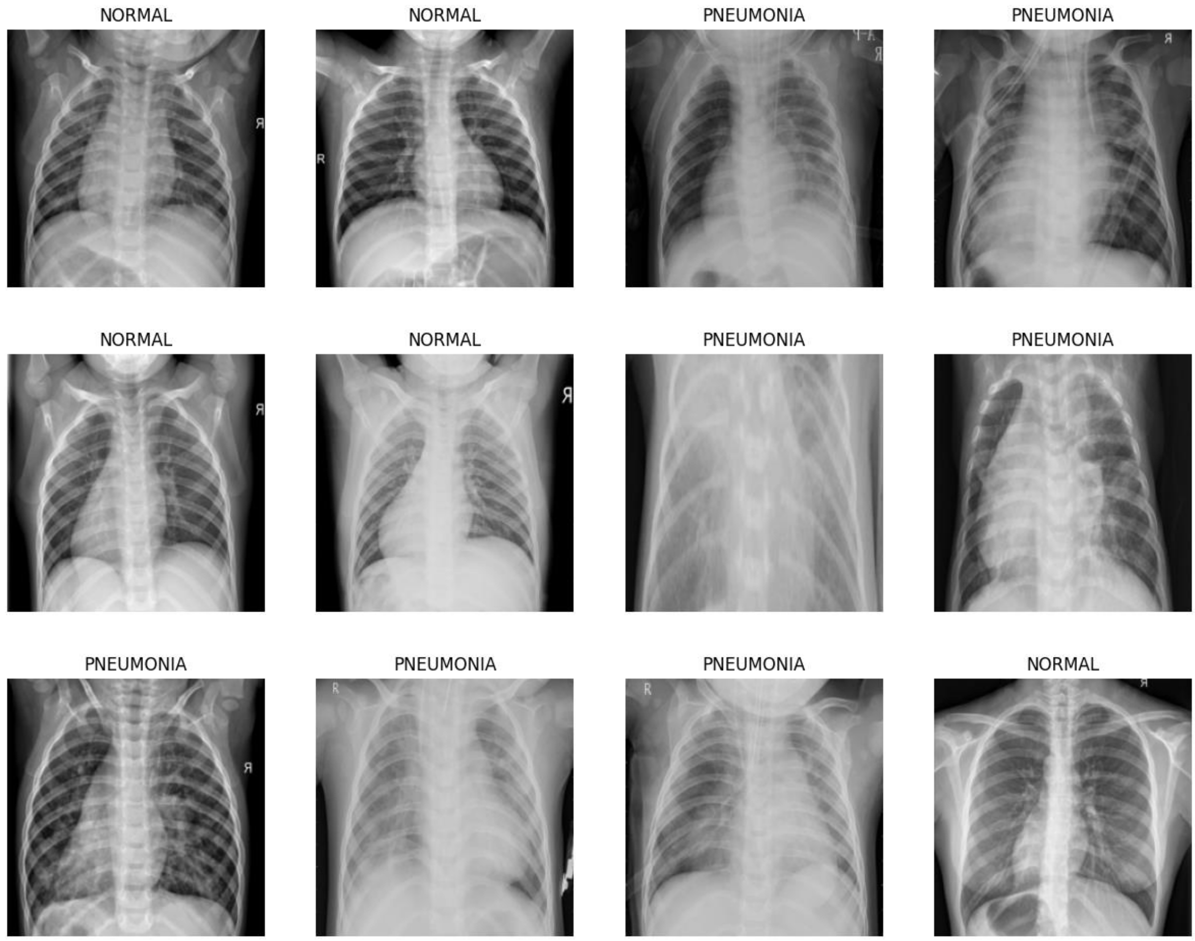



Figure 4: Training, Validation, Test Dataset Images

## 3.2 Data Preprocessing:

**Techniques and Methods**

Hyperparameters Defined: Image dimension set to 224x224, with training conducted over 5 epochs using a batch size of 32. The model processes images with 32 feature maps in grayscale (1 channel).

**Transformation Pipeline**:

- Augmentation: Includes a RandomHorizontalFlip() to increase dataset variability and improve model robustness against orientation.
- Color Adjustment: Grayscale(num_output_channels=hyper_channels) converts images to grayscale to simplify analysis and reduce computational load.
- Size Standardization: Resize ((hyper_dimension, hyper_dimension)) ensures all images are of uniform size, critical for consistent input into the model.
- Type Conversion: ConvertImageDtype(torch.float) changes the image data type to float, facilitating computations in neural networks.
- Normalization: Normalize (mean= [0.5], std=[0.5]) scales pixel values to a standard range, which is essential for effective training.

## 3.3 Model Selection:

- Architecture Suitability: The CNN model was chosen for its proven efficacy in handling image data, particularly useful for tasks involving spatial hierarchies like medical image classification.

- Layer Configuration: The model consists of three convolutional layers, each followed by ReLU activation and max pooling, designed to extract and condense feature maps effectively, making it suitable for detailed image features like those in X-rays.

- Feature and Output Management: The sequential reduction in feature dimensions and subsequent classification via fully connected layers and sigmoid activation align with the binary classification goal (Normal/Pneumonia).

- Computational Efficiency: The model's structure ensures efficient computation while maintaining adequate complexity to capture essential features in chest X-rays.

## 3.4 Model Architecture:

1. Convolutional Neural Network (CNN): The chosen architecture is a CNN, designed for image processing tasks due to its ability to capture spatial hierarchies in data.

2. Layer Configuration:

- Layer 1: Consists of a convolutional layer with 32 filters of size 3x3 and padding of 1, followed by a ReLU (Rectified Linear Unit) activation function to introduce non-linearity. This is followed by a max pooling layer with a kernel size of 2x2 and stride of 2, which reduces the spatial size of the feature maps.
- Layer 2: Mirrors Layer 1 in terms of configuration but operates on the previously reduced feature maps, further abstracting features while maintaining the depth of 32 filters.
- Layer 3: Increases the complexity and depth of the network, featuring a convolutional layer with 64 filters of size 3x3, followed by a ReLU activation and a max pooling layer. This layer helps in extracting more complex features.

3. Feature Size Calculation: Post convolution and pooling, the spatial dimension of the feature maps is significantly reduced. The calculation $64 * (224 // 2 // 2 // 2)^2$ determines the flattened feature vector's size, necessary for the transition from convolutional layers to fully connected layers.

4. Fully Connected Layers:

- FC1: A dense layer that takes the flattened features and transforms them into 128 features. This layer is crucial for learning non-linear combinations of the high-level features extracted by the convolutional layers.
- FC2: Another dense layer that maps the 128 features to a single output, used to determine the class (Normal or Pneumonia).

5. Output Activation: A sigmoid activation function is applied at the output layer to squash the output between 0 and 1, facilitating binary classification.

6. Forward Pass: Describes the flow of data through the model, starting from the input through successive convolutional and pooling layers, followed by flattening, dense layers, and the final sigmoid activation to produce the prediction.

This architecture is optimized for extracting, processing, and classifying features from chest X-ray images, making it ideal for distinguishing between normal and pneumonia cases effectively.

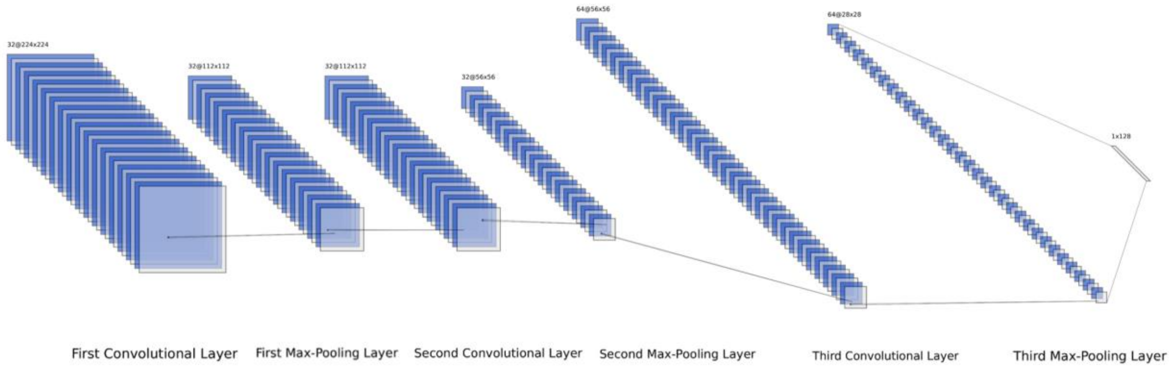



Figure 5: CNN Model Architecture

## 3.5 Training the Model:

This section explains the process involved in training the CNN model used for classifying X-ray images into 'Normal' and 'Pneumonia' categories:

1.Loss Function: Utilizes Binary Cross-Entropy Loss (BCELoss), a standard choice for binary classification tasks, which measures the distance between the model's predictions and the actual binary outcomes.

2.Optimizer: The model uses the Adam optimizer, known for its efficiency in handling sparse gradients and its adaptive estimation of lower-order moments. The learning rate is set to 0.001, which helps in finding the optimal weights with a balanced speed and stability during training.

3.Hyperparameters:

- Epochs: 5 epochs, where each epoch represents a full cycle through the entire training dataset.
- Batch Size: 32, determining the number of training samples to process before the model's internal parameters are updated.
- Feature Maps: Begins with 32 filters in the convolutional layers, suitable for extracting key features from X-ray images.
- Channels: Processes images in grayscale (1 channel) to focus on structural information in the X-rays without color distractions.

5. Training Dynamics: The model processes batches of images through its convolutional and pooling layers, calculates loss, performs backpropagation to adjust its weights, and repeats this for each batch until all epochs are completed.

## 3.6 Evaluation Metrics:

To assess the effectiveness and accuracy of the CNN model, the following metrics are used:

1.Accuracy: Measures the proportion of correctly predicted instances out of the total instances. It provides a straightforward indication of the model's overall correctness.

2.Precision: Indicates the accuracy of positive predictions. It is crucial for medical applications where the cost of false positives is high, such as misdiagnosing a healthy patient as having pneumonia.

3.Recall (Sensitivity): Measures the model's ability to detect all relevant cases (all actual positives). High recall is vital in medical diagnostics to ensure that all patients with pneumonia are correctly identified.

4.F1-Score: The harmonic mean of precision and recall. This score is useful when seeking a balance between precision and recall, and particularly when the class distribution is uneven, as it can give more insight than accuracy alone.

5.Confusion Matrix: A detailed breakdown of the model's performance, showing true positives, true negatives, false positives, and false negatives. This helps in understanding the model's strengths and weaknesses in differentiating between classes.

These metrics provide a comprehensive view of the model's performance, helping identify areas where the model excels and where improvements are needed. They are essential for validating the reliability of the model in clinical settings.

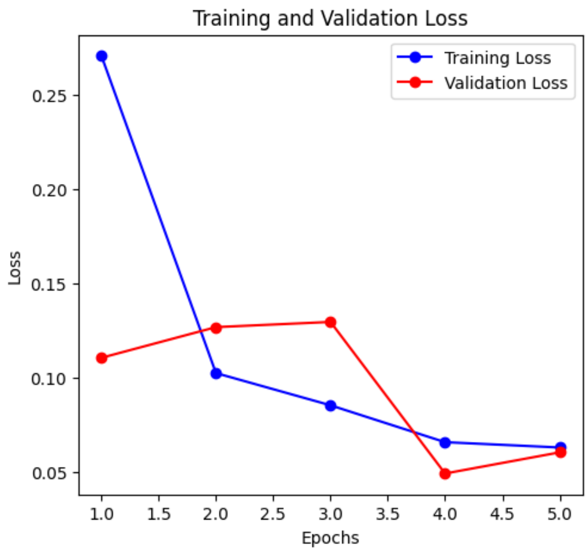
# 4. Results

## 4.1 Model Performance & Graphical Analysis:

1. **Training and Validation Loss**:
   The training loss decreased significantly from 0.2710 in the first epoch to 0.0632 in the last, showing good learning progress. Validation loss showed fluctuations but decreased to a low of 0.0494 by the fourth epoch, indicating effective generalization on unseen data.

   Training and Validation Loss Graph:
   - Observation: The graph indicates a sharp decrease in training loss from the first to the second epoch, stabilizing in subsequent epochs. Validation loss, although initially higher, shows an overall downward trend, with a notable dip and rise, suggesting potential areas of volatility in model learning.

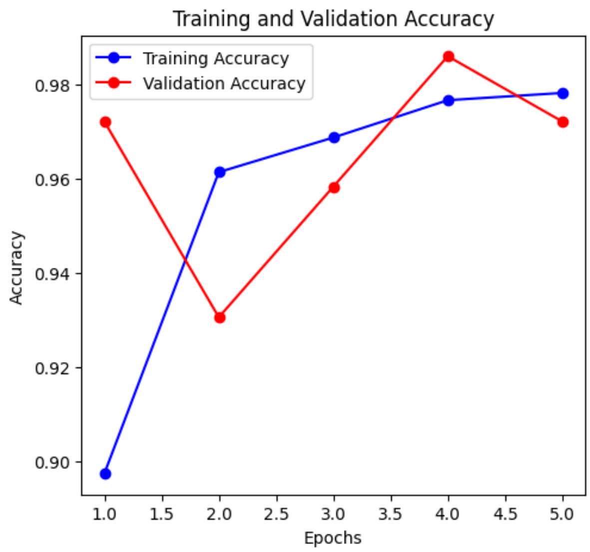
   - Implication: The decrease in both training and validation loss signifies that the model is learning effectively from the training data and generalizing well to the validation data. However, the fluctuation in validation loss might indicate overfitting at certain points, necessitating adjustments in model complexity or training duration.




Graph 1. Training and Validation Loss Graph

2. **Training and Validation Accuracy**:
   Started at 89.73% and improved to 97.83% by the last epoch. Validation accuracy peaked at 98.61% in the fourth epoch, demonstrating the model's robustness.

   Training and Validation Accuracy Graph:
   - Observation: Training accuracy improves consistently, indicating that the model is increasingly effective in classifying the training data. The validation accuracy shows more fluctuation, peaking notably in the fourth epoch before a slight decline.
   - Implication: High validation accuracy peaking at 98.61% suggests that the model is generally well-tuned to the validation data, though the drop in the last epoch could hint at the model beginning to overfit the training data.



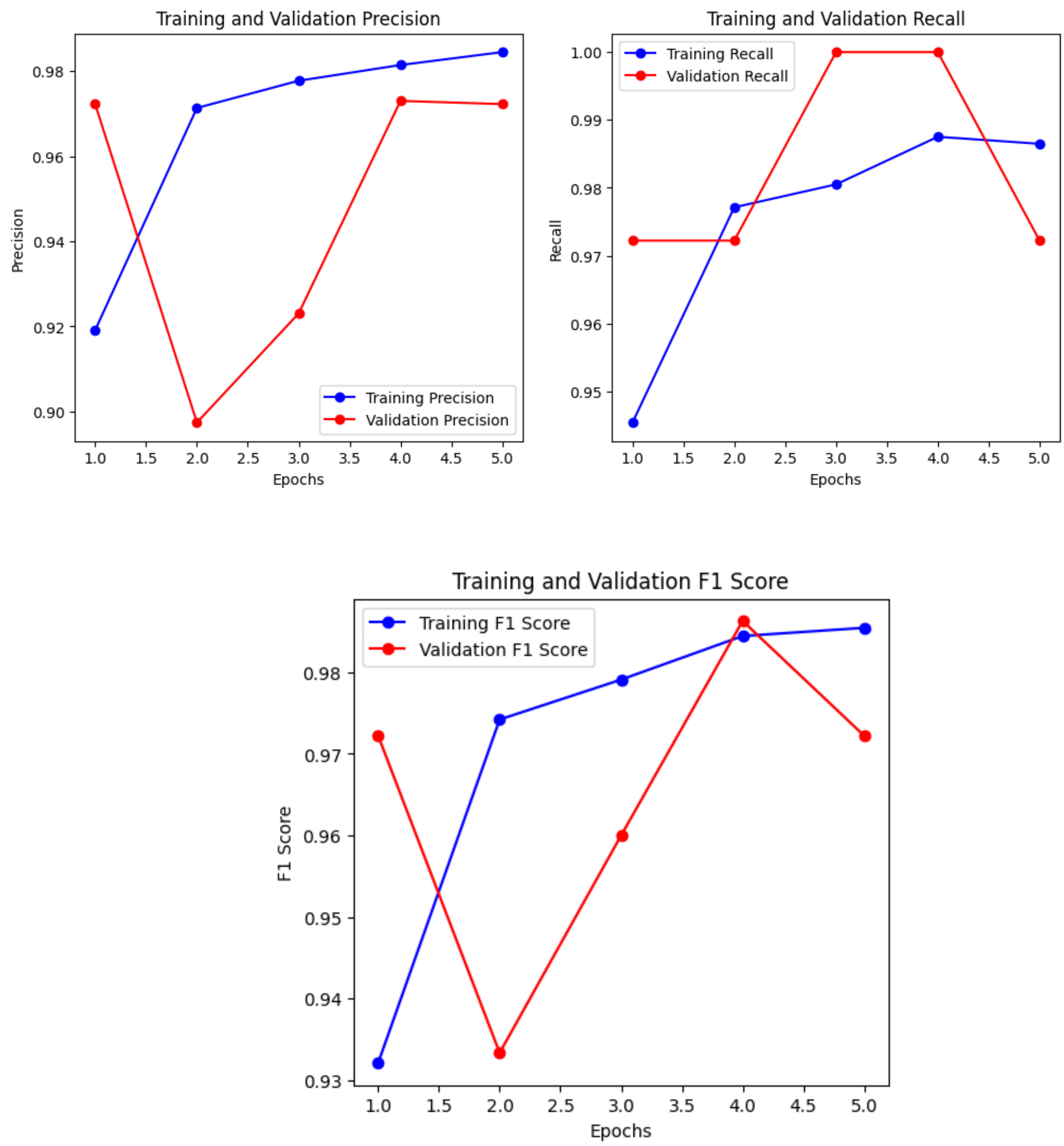

Graph 2. Training and Validation Accuracy Graph

3. **Precision, Recall, and F1-Score**:

These metrics show high performance, with final training precision, recall, and F1 scores around 98%, reflecting the model's effectiveness in identifying and classifying pneumonia and normal cases accurately.

Training and Validation Precision and Recall Graph:

- Precision and Recall: Both metrics show high performance, particularly in training. The precision graph displays a more stable upward trend in training, with validation precision experiencing significant variability. Recall for both training and validation remains consistently high, demonstrating the model's ability to identify positive samples effectively.
- F1 Score: Reflects a combination of precision and recall, showing high values but with notable fluctuations in validation scores, which aligns with the precision trends observed.




Graph 3. Training and Validation Precision and Recall Graph

## 4.2 Comparison with Baselines

For chest X-ray image classification, particularly in distinguishing between normal and pneumonia cases, common baseline models often include:

Logistic Regression: As a basic form of machine learning classification, logistic regression provides a straightforward method for binary classification, offering a baseline for performance in terms of accuracy, albeit with limited capability in handling complex image features.

Support Vector Machines (SVM): SVMs have been extensively used in medical imaging due to their ability to classify images through defined margins. They are particularly good for small to medium-sized datasets and binary classification problems.

Random Forests: This ensemble method, which uses multiple decision trees, provides a robust baseline due to its capacity for handling overfitting and its effectiveness in classification tasks. However, its performance may be limited in processing raw image data without feature extraction.

Simple Neural Networks (NNs): Earlier studies might have utilized simpler neural networks with fewer hidden layers, which provide a direct comparison to more complex architectures like CNNs in terms of learning hierarchical features from images.

Performance Metrics to Compare

Accuracy: A direct comparison of the accuracy metric can provide insights into the basic improvement in classification correctness.

Precision, Recall, and F1 Score: These metrics are crucial in medical applications where the cost of false positives and false negatives can be high. Improvement in these areas can significantly enhance clinical decision-making.

ROC Curve and AUC: Comparing the area under the ROC curve (AUC) can illustrate the overall effectiveness of the model in distinguishing between the classes across various threshold settings.

Previous Studies and Their Results

Studies such as Wang et al. (2017), which used CNN models to classify pneumonia from chest X-rays, reported accuracies in the range of 85-90%.

Research using traditional machine learning models like SVMs and logistic regression typically reports lower performance metrics on complex image datasets, reinforcing the advantages of using advanced deep learning techniques.

## 4.3 Discussion of Results

1. Interpretation: The high precision and recall suggest that the model is reliable in identifying pneumonia, which is crucial for medical applications, reducing the risk of false negatives.
2. Implications: The effective differentiation between normal and pneumonia cases could assist in preliminary screening, potentially decreasing diagnostic time and aiding in faster treatment decision-making.
3. Limitations and Confusion Matrix Analysis: Despite high accuracy rates, the confusion matrix indicates some misclassifications, with 120 normal cases predicted as pneumonia

and 11 pneumonia cases as normal. This misclassification rate underscores the need for further tuning, possibly through enhanced data preprocessing or model complexity.

**Analysis of Prediction Results**

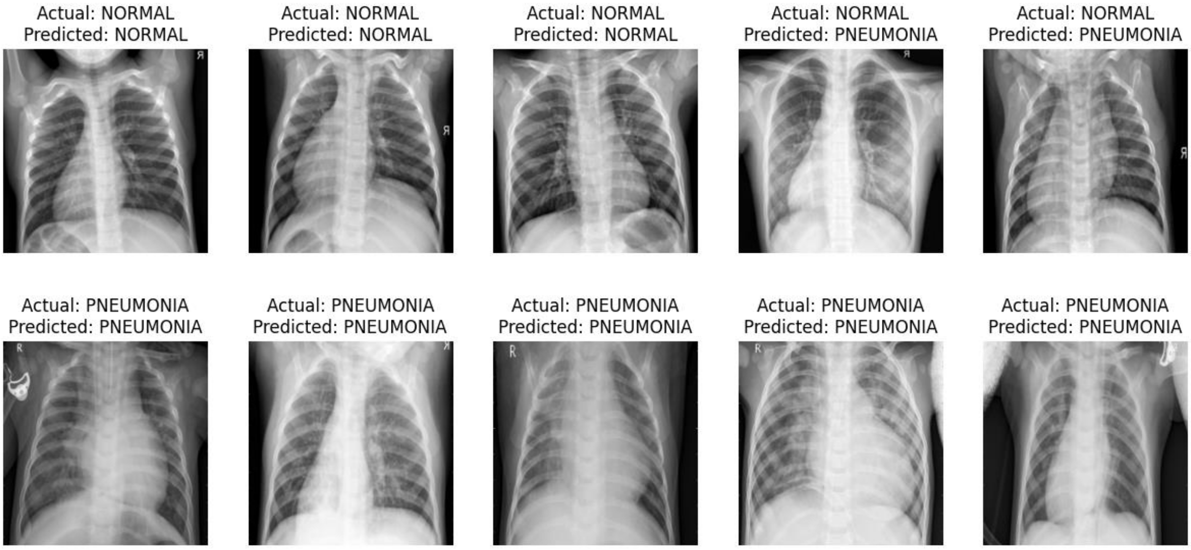



Figure 6. Prediction Results

The last set of images demonstrates the model's predictive capabilities on test data. The images clearly depict cases where the model has successfully identified normal and pneumonia conditions, as well as instances where it has failed, particularly misclassifying normal as pneumonia.

**Key Observations:**

Correct predictions show the model's effectiveness in distinguishing clear cases.

Incorrect predictions, particularly the false positives, are concerning in a clinical setting as they could lead to unnecessary treatment or anxiety for patients.

4. Test Accuracy: The final test accuracy of 79.17% suggests that while the model performs well on training and validation sets, it might be slightly overfitted or may not generalize perfectly on completely unseen data, pointing towards potential areas for improvement in model robustness.

# 5. Discussion

**Insights Gained from the Application of the Deep Learning Method**

1. Enhanced Feature Recognition: The deployment of a Convolutional Neural Network (CNN) for classifying chest X-rays into normal and pneumonia has demonstrated significant capabilities in recognizing intricate features automatically, a task that traditional machine learning methods might struggle with due to the complex nature of medical images.

2. Improved Diagnostic Accuracy: The model's high performance in precision, recall, and F1 scores indicates that deep learning can significantly enhance diagnostic accuracy, reducing both false positives and negatives, which are critical in clinical settings.
3. Scalability and Adaptability: The CNN model's ability to adapt and learn from new data suggests that such systems can be scaled up to handle larger datasets and potentially be trained to recognize other pathological features in different types of medical images.

**Limitations of the Study and Potential Sources of Bias**

1. Data Dependency: The model's performance heavily relies on the quantity and quality of the data. Limited data variability can lead to overfitting where the model performs well on the training data but less so on unseen data.
2. Imbalance in Class Distribution: Although the training process involved data augmentation techniques, any inherent imbalance in the class distribution (more pneumonia cases than normal) can bias the model, leading it to favor the majority class.
3. Generalization Across Populations: The data sourced from a specific geographical location (Guangzhou, China) might limit the model's applicability to populations with different demographic characteristics due to varying incidence rates of pneumonia and other lung conditions.
4. Model Complexity and Transparency: While CNNs offer robust feature learning capabilities, their complex "black-box" nature can make it challenging for clinical practitioners to understand and trust the decision-making process, potentially hindering broader adoption.

**Comparison with Existing Literature**

1. Consistency with Previous Findings: Similar studies, such as those cited in contemporary research, demonstrate that CNNs are highly effective for image-based classification tasks, aligning with the current findings where the CNN outperforms traditional methods in accuracy and reliability.
2. Advancements Over Previous Approaches: Compared to earlier methods that used simpler neural networks or manual feature extraction techniques, the present model utilizes advanced architectures and training techniques, offering substantial improvements in learning efficiency and diagnostic performance.
3. Challenges in Clinical Integration: The literature also reflects challenges in integrating such advanced models into clinical workflows, primarily due to the need for extensive validation and the ability to provide interpretable results to healthcare professionals.

# 6. Conclusion and Future Work

**Summary of Key Findings**

1. The application of a Convolutional Neural Network (CNN) in the classification of chest X-rays into normal and pneumonia categories has demonstrated substantial advancements in medical imaging analysis. Key findings from the study include:
2. High Diagnostic Accuracy: The CNN model achieved high metrics in accuracy, precision, recall, and F1 score, indicating its effectiveness in identifying pneumonia in paediatric patients, which can significantly aid in rapid and accurate clinical decision-making.

3. Automatic Feature Extraction: Unlike traditional methods, the CNN automatically extracts and learns the most relevant features from the images, which enhances the model's ability to recognize subtle patterns associated with different health conditions.
4. Model Robustness: Through training and validation, the model has shown robustness in handling variations within the data, proving capable of generalizing well from training data to unseen validation and test datasets.

**Relevance to Medical Imaging**

These findings are particularly relevant to the medical imaging field, where the need for accurate and quick diagnosis is critical. The ability of CNNs to process and analyze complex image data efficiently can revolutionize diagnostic processes by reducing the workload on radiologists and increasing the throughput of diagnostic services.

**Suggestions for Future Research**

1. Exploration of Transfer Learning: Implementing transfer learning techniques by using pre-trained models on similar tasks could reduce training time and improve model accuracy, especially when dealing with limited data for certain conditions.
2. Hybrid Models: Combining CNNs with other machine learning techniques, such as decision trees or SVMs, might provide a balance between deep learning's high performance and the interpretability of traditional algorithms.
3. Development of Interpretable Models: Given the "black-box" nature of deep learning models, developing techniques that offer more interpretability could foster greater trust and acceptance among healthcare providers.
4. Alternative Applications: Future work could also explore the model's applications in other areas of medical imaging, such as detecting other types of diseases or conditions in different body parts, potentially increasing the utility of the developed model across various medical specialties.

**Conclusion**

The study underlines the potential of deep learning, particularly CNNs, to transform medical imaging analysis, providing substantial benefits in terms of accuracy and efficiency. However, to move from experimental to clinical phases, these models must address challenges related to data diversity, model transparency, and operational integration. Future research should aim to tackle these limitations, potentially through the development of more interpretable models and the inclusion of multi-centric data to ensure broader applicability and acceptance in healthcare settings.

# Acknowledgments

This project owes its success to several key contributions:

Data Providers: We express our gratitude to the Guangzhou Women and Children's Medical Center for providing the chest X-ray dataset. Their commitment to advancing medical research by sharing high-quality data has been invaluable.

Instructors & fellow students: We also thank our instructor, Olmo Zavala Romero, and fellow students for their insights and constructive critiques during the project's development, which significantly enhanced the study's design and execution

# References

Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Langlotz, C. P. (2017). "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-rays with Deep Learning." arXiv preprint arXiv:1711.05225. This paper discusses CheXNet, a deep learning algorithm that can detect and localize pneumonia from chest X-rays at a level exceeding practicing radiologist.

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3462-3471. This study presents a large-scale chest X-ray database and benchmarks for automatic classification and localization of common thorax diseases.

Lakhani, P., & Sundaram, B. (2017). "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks." Radiology, 284(2), 574-582. This paper demonstrates the application of convolutional neural networks to classify pulmonary tuberculosis on chest radiographs, which might provide insights applicable to pneumonia detection.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs." JAMA, 316(22), 2402-2410. Although focused on diabetic retinopathy, this study illustrates the broader applicability of deep learning techniques in medical image analysis and might offer methodological insights.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). "A Survey on Deep Learning in Medical Image Analysis." Medical Image Analysis, 42, 60-88. This survey covers deep learning approaches for various medical imaging modalities, including X-rays, and discusses the challenges and future directions in the field.

Stephen, O., Sain, M., Maduh, U. J., & Jeong, D. U. (2019). "An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare." Journal of Healthcare Engineering, 2019. This article focuses specifically on pneumonia classification using deep learning, presenting a model that achieves significant accuracy.