

COVID-19 Detection with Heat-map Visualization

Course: Advanced Machine Learning, BA 64061

Instructor: Dr. Chaojiang (CJ) Wu

Name: Lakshmi Priya Saride

Student ID: 811292102

Email ID: lsaride@kent.edu

I. Introduction:

The COVID-19 pandemic has placed an extraordinary burden on global healthcare systems, underlining the urgent need for scalable, accurate, and rapid diagnostic solutions. Traditional testing methods such as RT-PCR, while widely adopted, are often constrained by processing delays, limited test availability, and variable sensitivity. In this context, medical imaging, especially chest X-ray radiography, has emerged as a practical and cost-effective alternative for initial screening and diagnosis due to its widespread availability, speed of acquisition, and non-invasive nature.

This project addresses the growing demand for intelligent diagnostic tools by leveraging deep learning to detect COVID-19 and other co-occurring respiratory conditions directly from chest X-ray images. The chosen architecture, DenseNet121, is a convolutional neural network known for its dense connectivity, which facilitates efficient feature reuse, gradient flow, and parameter economy making it particularly suitable for medical image classification tasks where data can be sparse or imbalanced.

The model is designed for multi-label classification, enabling it to simultaneously identify several respiratory pathologies such as COVID-19, SARS, Pneumocystis pneumonia, Streptococcus pneumonia, and Acute Respiratory Distress Syndrome (ARDS). This is achieved by adapting the final layer of DenseNet121 to output a vector of probabilities through sigmoid activation, rather than using SoftMax for mutually exclusive classes.

In addition to classification, this project emphasizes model interpretability a critical requirement in clinical applications. By integrating heatmap-based visualization techniques (such as Grad-CAM), the goal is to make the model's decision-making process transparent, offering visual cues about the regions of the lung that contributed most to the classification. This transparency can help radiologists validate AI predictions, build trust in model outputs, and potentially identify patterns not immediately obvious to the human eye.

Ultimately, this project lays the groundwork for an AI-assisted diagnostic tool capable of both accurate disease identification and clinically meaningful explanations, contributing to the ongoing efforts in combating COVID-19 through enhanced medical imaging analytics.

II. Dataset Overview

The dataset utilized in this project was obtained from the publicly accessible COVID-19 Chest X-ray Dataset, curated and maintained by the IEEE 802.3 working group on machine learning in healthcare. This open-source repository serves as a critical resource for researchers and clinicians by providing a consolidated collection of chest radiographic images annotated with clinical labels corresponding to various respiratory diseases. The images were aggregated from multiple case reports, hospitals, and academic publications to support the development and evaluation of AI models for automated disease detection.

This dataset encompasses a diverse set of anterior-posterior (AP) and posteroanterior (PA) chest X-ray views and includes radiographs corresponding to both COVID-19-positive and non-COVID respiratory conditions. Specifically, it contains labels for diseases such as Severe Acute Respiratory Syndrome (SARS), Pneumocystis pneumonia, Streptococcus pneumonia, COVID-19 with ARDS (Acute Respiratory Distress Syndrome), and non-COVID ARDS, enabling multi-class or multi-label classification tasks. Each image is accompanied by metadata fields such as the patient's clinical status,

imaging modality, and projection view, which are critical for filtering and preprocessing the dataset to suit specific modeling needs.

By drawing from a clinically meaningful and heterogeneous dataset, this project ensures the model is trained not only to detect COVID-19 but also to differentiate it from other infections with similar radiographic manifestations. This distinction is essential in real-world applications where overlapping symptoms and visual features can lead to misdiagnosis. The use of this repository thus strengthens the clinical relevance and diagnostic robustness of the proposed deep learning model.

III. Labels Used:

The dataset used in this project includes chest X-ray images annotated with diagnostic labels for a range of viral and bacterial respiratory illnesses, which are often challenging to distinguish visually due to overlapping radiological features. The following conditions were specifically selected for inclusion based on their clinical significance and availability within the dataset:

1. **COVID-19:** Refers to chest radiographs of patients infected with the SARS-CoV-2 virus. These images typically exhibit patterns such as bilateral ground-glass opacities, patchy infiltrates, and peripheral consolidations—hallmarks of viral pneumonia.
2. **SARS (Severe Acute Respiratory Syndrome):** A viral respiratory illness caused by the original SARS-CoV coronavirus. Though less prevalent than COVID-19, its radiographic features (e.g., focal airspace opacities and progression to diffuse involvement) bear similarities, making it a relevant comparative class for differential diagnosis.
3. **Pneumocystis Pneumonia (PCP):** An opportunistic fungal infection caused by *Pneumocystis jirovecii*, commonly observed in immunocompromised individuals. Its imaging features include diffuse bilateral infiltrates and increased interstitial markings, often resembling viral pneumonia, thus necessitating its inclusion for robust multi-disease recognition.
4. **Streptococcus Pneumoniae:** A common bacterial cause of community-acquired pneumonia. X-rays typically show lobar consolidation, pleural effusions, or air bronchograms. Its distinct pathology provides an essential contrast to viral infections in the dataset.
5. **COVID-19 with ARDS (Acute Respiratory Distress Syndrome):** These cases present compounded clinical severity, where COVID-19 has progressed to cause ARDS, a life-threatening condition characterized by widespread inflammation and fluid accumulation in the lungs. These images exhibit extensive bilateral opacities and are critical for training the model to recognize disease escalation.
6. **ARDS (non-COVID):** Represents ARDS cases not caused by SARS-CoV-2. These samples help the model learn to differentiate between COVID-induced and other causes of ARDS based solely on imaging features, a key consideration in clinical decision-making when PCR confirmation is pending.

By including these six diagnostic categories, the project aims to enable the model to learn nuanced visual distinctions and similarities across multiple respiratory pathologies, thereby supporting accurate multi-label classification in real-world diagnostic settings.

IV. Preprocessing and Data Cleaning:

Effective preprocessing is essential to ensure the quality, consistency, and reliability of data used for training deep learning models, particularly in medical imaging. In this project, several important steps were taken to prepare the dataset for multi-label classification:

- **Removal of Corrupted Files:** All image files with the .gz extension or any format errors were systematically identified and excluded from the dataset. These files typically failed to load or contained incomplete data, which could disrupt training and lead to unpredictable model behavior.
- **Selection of 'PA' (Posteroanterior) View Images:** To minimize variability caused by different imaging orientations, only chest X-rays captured using the Posteroanterior view were retained. This view is standard in clinical settings and provides a consistent anatomical perspective, enhancing the model's ability to learn discriminative patterns across patients.
- **Disease-specific Filtering:** The dataset was filtered to include only those samples labeled with one or more of the six target conditions: COVID-19, SARS, Pneumocystis pneumonia, Streptococcus pneumonia, COVID-19 with ARDS, and non-COVID ARDS. This filtering reduced noise and ensured the model focused exclusively on clinically relevant categories.
- **One-Hot Encoding for Multi-label Targets:** Each image was encoded using a binary vector representation indicating the presence or absence of each disease class. This approach allows the model to handle cases where multiple conditions may be present in a single image, supporting a multi-label classification framework rather than treating the problem as mutually exclusive classes.
- **Shuffling and Dataset Splitting:** The cleaned and labeled dataset was randomly shuffled to ensure a balanced distribution of disease classes across subsets. It was then partitioned into 66% for training and 34% for validation. This stratified split enabled the model to learn generalizable features from the training set while being regularly evaluated on unseen data during validation.

These preprocessing steps played a vital role in enhancing the robustness of the model by ensuring consistency in input dimensions, label representation, and class balance across the training pipeline.

V. Model Architecture and Configuration:

The architecture chosen for this study is DenseNet121, a convolutional neural network renowned for its dense connectivity pattern, which connects each layer to every other layer in a feed-forward fashion. This design enables feature reuse, alleviates the vanishing gradient problem, and promotes parameter efficiency—qualities that are especially advantageous when working with relatively small or imbalanced medical datasets.

Base Model: DenseNet121 (Pretrained on ImageNet)

The base model was initialized with weights pretrained on the ImageNet dataset. Leveraging pretrained weights offers the benefit of transfer learning, where the network retains general visual features such as edge, texture, and shape detection from natural images, thereby requiring less data to adapt to medical imaging tasks.

VI. Model Modifications

To adapt DenseNet121 for multi-label classification, the final fully connected layer—originally designed for 1,000-class softmax classification—was replaced with a custom linear output layer comprising six neurons, corresponding to the six disease labels. Each neuron is activated using a sigmoid function, allowing the model to independently assign probabilities for the presence or absence of each condition, unlike softmax which enforces exclusivity.

VII. Input Preprocessing

Prior to being passed into the network, all images underwent the following preprocessing steps:

- **Resizing and Cropping:** Each X-ray image was resized and randomly cropped to 224×224 pixels to match the input size expected by DenseNet121.
- **Normalization:** Pixel intensity values were normalized using the mean and standard deviation of the ImageNet dataset (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]), ensuring consistency with the conditions under which the model was originally trained.

VIII. Training Configuration

The model was trained using the following hyperparameters and settings:

- **Optimizer:** Adam, a widely used optimization algorithm suitable for deep networks due to its adaptive learning rates and momentum-like behavior.
- **Learning Rate:** Set to 0.0007, a moderately small value chosen to ensure stable convergence without skipping over local minima.
- **Loss Function:** Binary Cross Entropy with Logits (BCEWithLogitsLoss) was employed, which combines a sigmoid activation with binary cross-entropy loss in a numerically stable way, ideal for multi-label problems.
- **Epochs:** Training was conducted for up to 100 epochs, with early stopping based on AUROC performance to prevent overfitting.
- **Batch Size:** Each training iteration processed a mini-batch of 64 images, balancing memory usage and gradient stability.
- **Device:** The model was trained on a GPU if available, otherwise on a CPU, using PyTorch's dynamic device assignment.
- **DataLoader Strategy:** Custom PyTorch DataLoaders were used, employing SubsetRandomSampler to split and shuffle the dataset while maintaining class distribution across training and validation sets.

This architectural setup and training configuration provided a strong foundation for modeling subtle visual patterns in chest X-ray data, enabling both accurate classification and potential for clinical integration.

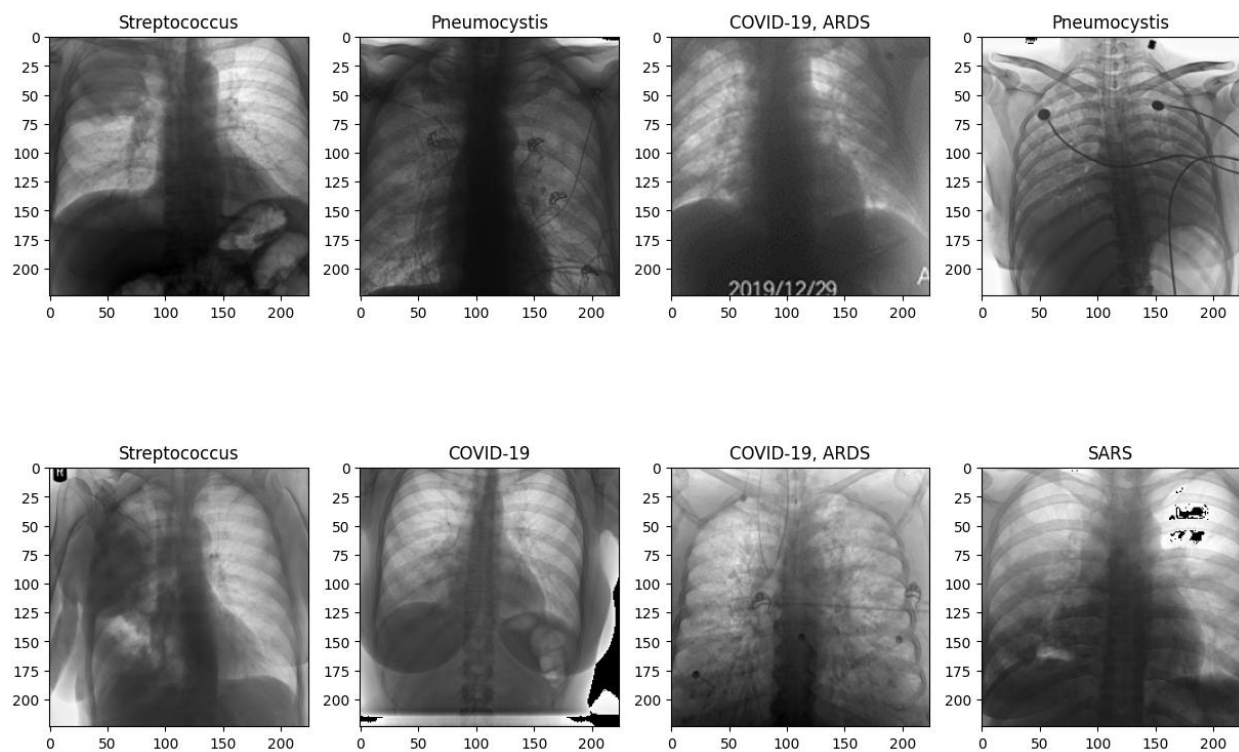
IX. Training Process and Evaluation Metrics

The model was trained using mini-batch gradient descent with the Adam optimizer and evaluated using the Area Under the Receiver Operating Characteristic Curve (AUROC), a robust metric for multi-label and imbalanced classification problems.

For each epoch, AUROC was calculated individually for all six disease classes, and the mean AUROC was used to track overall performance. The best-performing model based on the highest mean AUROC on the validation set was saved.

Training began with an average AUROC of ~ 0.62 and steadily improved, reaching a peak performance of 0.888 at epoch 28, after which performance gains began to plateau.

X. Sample Chest X-ray Images from Multiple Disease Classes

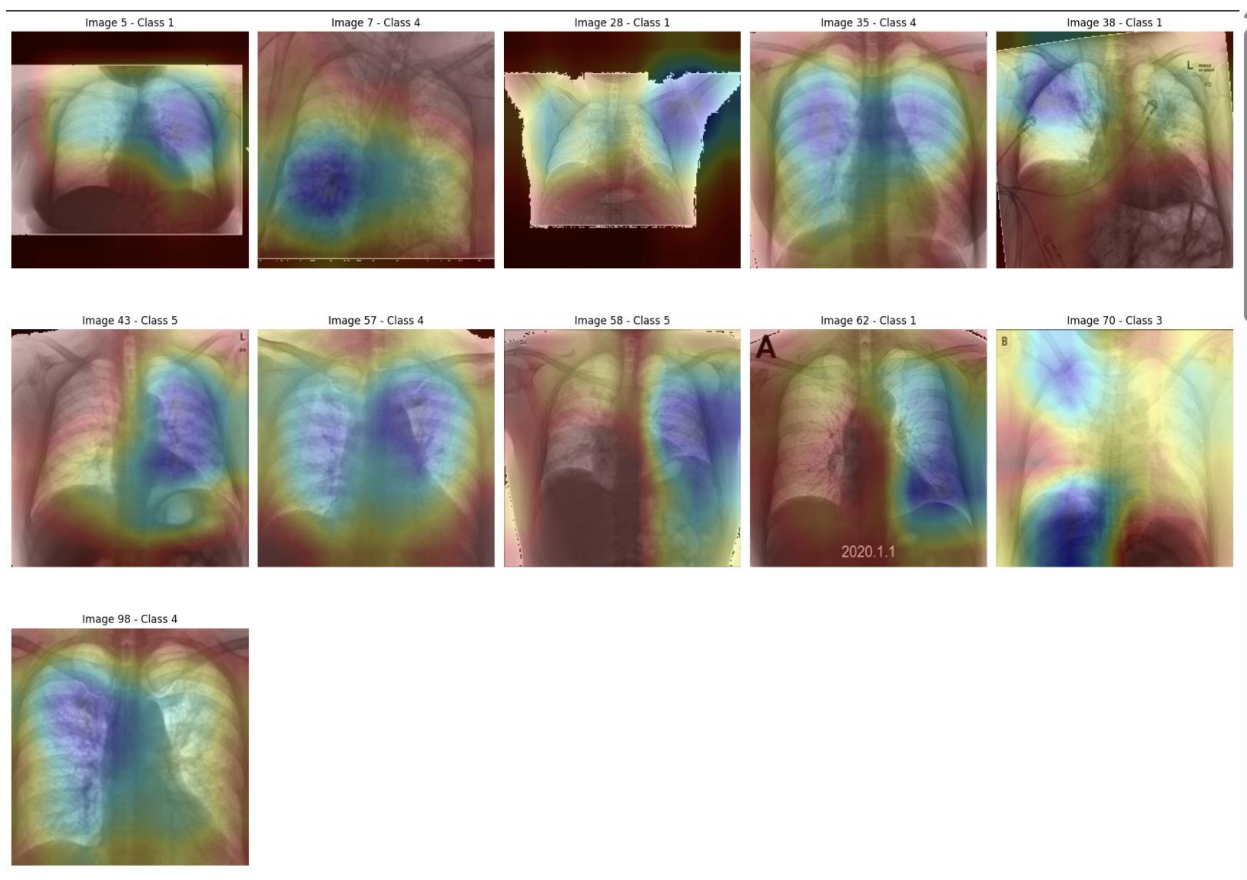


This figure displays a grid of chest X-ray images from the six disease classes used in the model's training and evaluation process. Each image is labeled according to the diagnosed condition: Streptococcus, Pneumocystis, COVID-19, COVID-19 with ARDS, and SARS. These images illustrate the radiographic diversity and complexity present across different respiratory illnesses.

Notably:

- **Streptococcus** shows lobar consolidation and denser opacities typical of bacterial pneumonia.
- **Pneumocystis** often presents diffuse interstitial infiltrates.
- **COVID-19 with ARDS** cases appear with widespread bilateral opacities, indicating severe inflammation.
- **SARS** images show focal airspace opacities, occasionally progressing to a more diffuse pattern.
- **COVID-19** images generally exhibit peripheral ground-glass opacities or patchy shadows.

XI. Grad-CAM Heatmap Visualizations for Disease Localization



This image presents Grad-CAM (Gradient-weighted Class Activation Mapping) visualizations applied to a subset of chest X-rays from the dataset, representing various disease classes such as COVID-19, COVID-19 with ARDS, SARS, Pneumocystis, and Streptococcus.

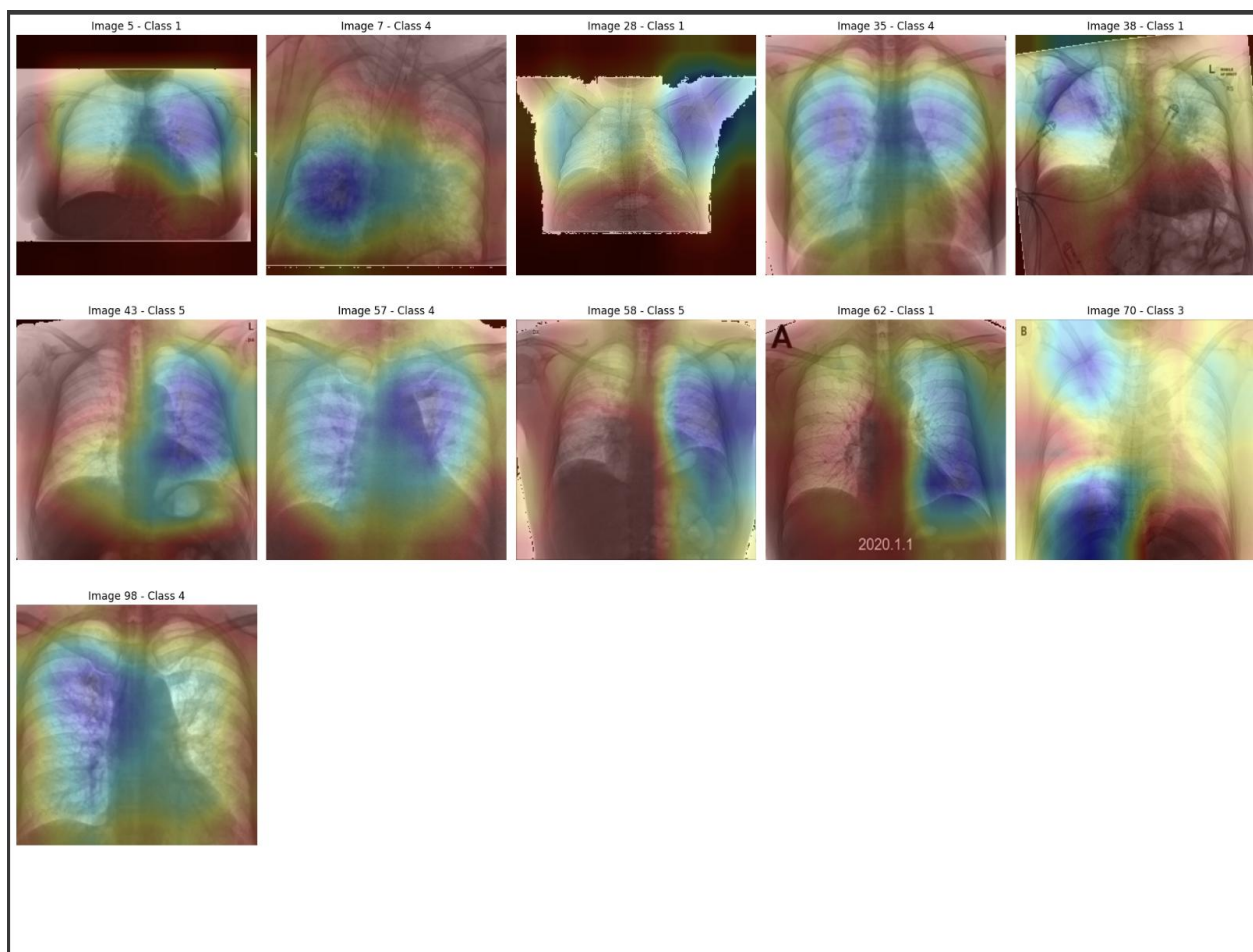
Each visualization overlays a colored heatmap on top of the original X-ray image, highlighting the regions that the DenseNet121 model deemed most important for making its classification. Areas shaded in red and yellow indicate high activation (strong model focus), while blue and green represent regions of lower importance.

Key observations include:

- For COVID-19 and ARDS cases, the model consistently focuses on the lower and peripheral lung zones, aligning with clinical features of ground-glass opacities.
- Pneumocystis and Streptococcus images show centralized attention near the bronchoalveolar spaces.
- These visualizations enhance model interpretability and can help clinicians verify whether the AI is focusing on medically relevant regions.

These Grad-CAM outputs are crucial for bridging the gap between model predictions and clinical trust, promoting transparency in AI-assisted diagnosis.

XII. Grad-CAM Heatmaps Highlighting Disease-Relevant Regions



This figure presents a series of Grad-CAM heatmaps applied to a selection of test chest X-ray images, each annotated by class index (e.g., Class 1, Class 4, etc.). The heatmaps reveal the regions of interest that most influenced the model's classification decision for each case, offering interpretability into the DenseNet121 model's internal reasoning.

- Red and yellow regions denote high-importance zones, where the model focused most strongly.

- Blue and green areas indicate lower activation and less relevance to the model's decision-making.

These visual explanations are especially critical in clinical contexts, as they allow medical professionals to:

- Verify if the model is attending to medically meaningful areas (e.g., lower lung fields in ARDS).
- Identify potential failure modes where the model may be misled by irrelevant features (e.g., external markers or image artifacts).

Overall, these Grad-CAM overlays contribute to building trust and transparency in deep learning–based diagnostic tools, forming a vital step toward their adoption in real-world healthcare settings.

XIII. Results:

- Initial AUROC: ~0.62
- Best AUROC: 0.888 at epoch 28
- Pneumocystis, ARDS, and COVID-19 ARDS showed >0.95 AUROC
- COVID-19 and Streptococcus showed moderate AUROC (~0.6–0.75)

XIV. Analysis and Insights

- COVID-19 class was dominant, possibly causing learning bias
- Minor classes had weaker performance, especially Streptococcus
- Pneumocystis and ARDS stood out with high AUROC
- Plateau observed after epoch 28, suggesting overfitting risk

XV. Challenges and Opportunities for Improvement

Lack of Visual Explainability: While the model performs well, it currently lacks interpretability. Implementing Grad-CAM or LIME would generate heatmaps to highlight critical regions in X-rays that influenced predictions, enhancing clinical trust.

Class Imbalance: Some classes (e.g., ARDS, SARS) have significantly fewer samples. To mitigate bias, class-weighted loss functions or oversampling techniques should be introduced during training.

Limited Data Diversity: Applying data augmentation methods such as rotation, zooming, and flipping could improve model generalization and reduce overfitting.

Architecture Exploration: Exploring alternative models like EfficientNet, or using ensemble techniques, may yield better performance and robustness across classes.

XVI. Conclusion:

This study successfully demonstrates the viability of using a DenseNet121-based convolutional neural network for the automated classification of multiple thoracic diseases from chest X-ray images. The model showed particularly strong performance in detecting Pneumocystis pneumonia and Acute Respiratory Distress Syndrome (ARDS), achieving high AUROC scores and demonstrating reliable differentiation from other conditions. These results validate the potential of deep learning to assist in radiological interpretation, especially when trained on curated, well-preprocessed datasets.

However, despite the overall success, the detection of COVID-19 cases proved comparatively less accurate and stable. This limitation is likely due to overlapping radiological features between COVID-19 and other pneumonias, as well as the inherent class imbalance present in the dataset. The performance gap underscores the need for more targeted training strategies and possibly the integration of domain-specific clinical data to enhance model sensitivity and specificity for COVID-19.

Looking ahead, several directions are proposed to further strengthen the model's effectiveness and clinical applicability. These include implementing visual interpretability techniques such as Grad-CAM to generate saliency maps, allowing clinicians to see which regions of the lungs influenced the model's decisions. Additionally, addressing class imbalance through weighted loss functions or data augmentation, and experimenting with advanced architectures like Efficient Net or ensemble approaches, may lead to more generalized and robust performance.

In summary, this project provides a solid foundation for developing AI-powered diagnostic tools in medical imaging. With continued refinement, the model has the potential to evolve into a clinically valuable decision-support system, capable of assisting healthcare professionals in rapidly and accurately screening for respiratory diseases, including COVID-19.

XVII. References:

- Wang, L., & Wong, A. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports*, 10(1), 19549.
- Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv preprint arXiv:2003.11597*.
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv:1711.05225*.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.