**Abstract:**

Pneumonia is a serious respiratory condition that remains a leading cause of death, particularly among children in low-resource settings. Rapid and accurate diagnosis using chest X-rays is essential but often limited by the availability of trained radiologists. To address this, our project investigates the use of deep learning for automated pneumonia detection from pediatric chest radiographs. We evaluate two state-of-the-art convolutional neural network (CNN) architectures—**DenseNet‑121** and **Inception v3** using **transfer learning**. Both models are pretrained on ImageNet and fine-tuned on a curated dataset of pediatric chest X-ray images labeled as either “NORMAL” or “PNEUMONIA.” The dataset, originally collected by Guangzhou Women and Children’s Medical Center, contains over 5,000 training images. We restructured the original splits to allocate 10% of the training set for validation and used a held-out test set of 624 images for final evaluation. Our results show that **Inception v3 outperforms DenseNet‑121**, achieving a **ROC-AUC of 0.967** and **accuracy of 75.8%**, compared to DenseNet‑121’s ROC-AUC of 0.916 and accuracy of 69.4%. The higher AUC of Inception v3 highlights its superior ability to distinguish pneumonia from normal cases, particularly in ranking-based evaluation metrics. These findings suggest that deeper architectures with larger input resolutions may yield better performance for medical image classification. Future work will explore decision threshold calibration, visual interpretability using Grad-CAM, and ensemble methods to further improve diagnostic accuracy and clinical reliability.

**Introduction:**

Pneumonia is a critical respiratory illness and a major cause of death among children globally, especially in regions with limited healthcare infrastructure. Accurate and timely diagnosis is essential for effective treatment; however, the interpretation of chest X-rays—a primary diagnostic tool—requires specialized medical expertise. In many underserved areas, the shortage of trained radiologists results in delayed or inconsistent diagnoses, which can significantly affect patient outcomes. This problem underscores the urgent need for automated diagnostic systems that can assist clinicians by providing fast, accurate assessments from medical imaging data.

In this project, we address the challenge of automated pneumonia detection in pediatric chest X-rays using deep learning techniques. Specifically, we evaluate and compare the performance of two powerful convolutional neural network (CNN) architectures—**DenseNet‑121** and **Inception v3**—through transfer learning. Both models were pretrained on the ImageNet dataset and subsequently fine-tuned on a pediatric chest X-ray dataset curated by the Guangzhou Women and Children’s Medical Center. The task is formulated as a binary classification problem distinguishing between “NORMAL” and “PNEUMONIA” cases. After preprocessing the images and applying basic augmentation to the training set, we trained both models using a consistent pipeline and evaluated their performance on a held-out test set of 624 images. Our results show that Inception v3 significantly outperforms DenseNet‑121, achieving a **ROC-AUC of 0.967** and **accuracy of 75.8%**, compared to DenseNet‑121’s ROC-AUC of **0.916** and **accuracy of 69.4%**. The societal benefit of this work lies in its potential to serve as a scalable and accessible diagnostic aid, particularly in regions with limited access to radiological expertise. By improving the speed and consistency of pneumonia detection, such systems can support earlier interventions and better patient outcomes, ultimately contributing to reduced pediatric mortality. Our contributions include a rigorous comparative study of DenseNet‑121 and Inception v3 on a clinically relevant pediatric dataset, the design of a standardized preprocessing and training pipeline, and an in-depth evaluation using both accuracy and ranking-based metrics (ROC-AUC). These contributions offer insights into the effectiveness of different CNN architectures for medical image classification and lay the groundwork for further development of AI-assisted diagnostic tools.

**Related Works:**

The use of deep learning for medical image classification, especially for chest X-rays, has become a well-established research area. One landmark study by Rajpurkar et al. (2017) introduced **CheXNet**, a DenseNet-121 model trained on the NIH ChestX-ray14 dataset, which demonstrated diagnostic performance that rivaled practicing radiologists in pneumonia detection. Their work validated the use of deep convolutional neural networks (CNNs) in clinical settings and set a precedent for adopting DenseNet-based architectures.

Other studies have explored the effectiveness of alternative architectures. For example, Stephen et al. (2019) employed **Inception v3** for pneumonia detection and found that its multi-scale convolutional design enhanced classification performance, especially in high-resolution medical imaging. Similarly, Kermany et al. (2018) compiled and released the pediatric chest X-ray dataset used in our study, demonstrating early success in detecting bacterial and viral pneumonia using standard CNNs.

Further work by Islam et al. (2020) investigated the utility of transfer learning across different CNN backbones and noted performance trade-offs between model complexity and diagnostic accuracy. Meanwhile, attention-based methods and model interpretability techniques, such as Grad-CAM, have also been proposed to enhance the transparency of AI-driven medical diagnoses (Selvaraju et al., 2017).

While our work builds on these foundational studies, it differentiates itself by conducting a **controlled, side-by-side comparison** of DenseNet-121 and Inception v3 specifically for **pediatric pneumonia** detection. Unlike prior works that either focused on adult populations or single-model performance, our approach ensures that both models are evaluated under identical preprocessing, augmentation, and training conditions. This enables a fair assessment of each architecture’s strengths, particularly in terms of ROC-AUC and real-world diagnostic reliability.

**References**

* Islam, M. T., Aowal, M. A., Minhaz, A. T., & Ashraf, K. (2020). Deep learning-based detection of pneumonia from chest X-ray images using ensemble models. *Computers in Biology and Medicine*, 123, 103898.
* Kermany, D. S., Zhang, K., & Goldbaum, M. (2018). Labeled optical coherence tomography (OCT) and chest X-ray images for classification. *Mendeley Data*, V2.
* 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. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618–626).
* 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.

**Data**

* Dataset Source:  
  Pediatric chest X-ray dataset curated by *Kermany et al.*, collected from Guangzhou Women and Children’s Medical Center.
* Tasks done:  
  Binary classification between NORMAL and PNEUMONIA (including both bacterial and viral pneumonia cases).
* Original Data Splits:
  + Training set: 5,216 images (1,341 NORMAL, 3,875 PNEUMONIA)
  + Validation set: 16 images (8 NORMAL, 8 PNEUMONIA)
  + Test set: 624 images (234 NORMAL, 390 PNEUMONIA)
  + To address the small validation set, ~10% of the training data was reallocated to form a larger validation split.
* Image Characteristics:
  + High-resolution grayscale X-rays (∼1,000–3,000 px)
  + Converted to 3-channel JPEGs to match pretrained model input requirements
  + Resized to 224×224 for DenseNet‑121 and 299×299 for Inception v3
  + Normalized using ImageNet mean and standard deviation
* Augmentation and Preprocessing:
  + Horizontal flips applied to training images only
  + No augmentation on validation or test sets to preserve anatomical fidelity
  + No use of rotations, crops, or color transformations
* Labeling and Bias Considerations:
  + Labels assigned by expert radiologists
  + Includes both bacterial and viral pneumonia cases
  + Class imbalance present: training data is skewed toward pneumonia
  + Dataset is limited to pediatric patients, which may impact generalization
* Why This Dataset:
  + Widely used in academic literature and benchmarks
  + Small enough for fast experimentation on limited hardware (e.g., single-GPU setups)
  + Targets a clinically relevant task of automated pneumonia detection in children

Pre Processing:

Certainly! Here's a rewritten version of the preprocessing steps without code snippets, in a clear and professional format suitable for your report:

* Preprocessing Steps:
* All images were normalized using ImageNet mean and standard deviation values to align with the expectations of pretrained models.
* Random horizontal flipping was applied only to the training images to improve model generalization and robustness. No augmentations were used on validation or test sets to maintain anatomical integrity during evaluation.

DenseNet‑121 Preprocessing:

* Input images were resized to a resolution of 224×224 pixels, consistent with the input dimensions required by DenseNet‑121.
* A sequence of preprocessing steps was applied to the training data, including resizing, normalization, and horizontal flipping.
* Validation and test data underwent only resizing and normalization to ensure consistent evaluation.

Inception v3 Preprocessing:

* Input images were resized to 299×299 pixels, which is the standard input size for Inception v3 to operate effectively.
* The preprocessing pipeline for training included resizing, normalization, and horizontal flipping for augmentation.
* As with DenseNet‑121, no augmentation was applied to validation or test data, only resizing and normalization.

**Methods:**

To address the problem of automated pneumonia detection in pediatric chest X-rays, we adopted a transfer learning approach using two high-performing convolutional neural network (CNN) architectures: DenseNet-121 and Inception v3. Transfer learning was selected as it allows pretrained models—originally trained on large-scale natural image datasets like ImageNet—to be fine-tuned for domain-specific tasks with relatively limited data. This method is particularly suitable for medical imaging, where labeled data is scarce and costly to obtain. By reusing pretrained feature extractors, we aimed to improve convergence speed and model generalization on our pediatric dataset.

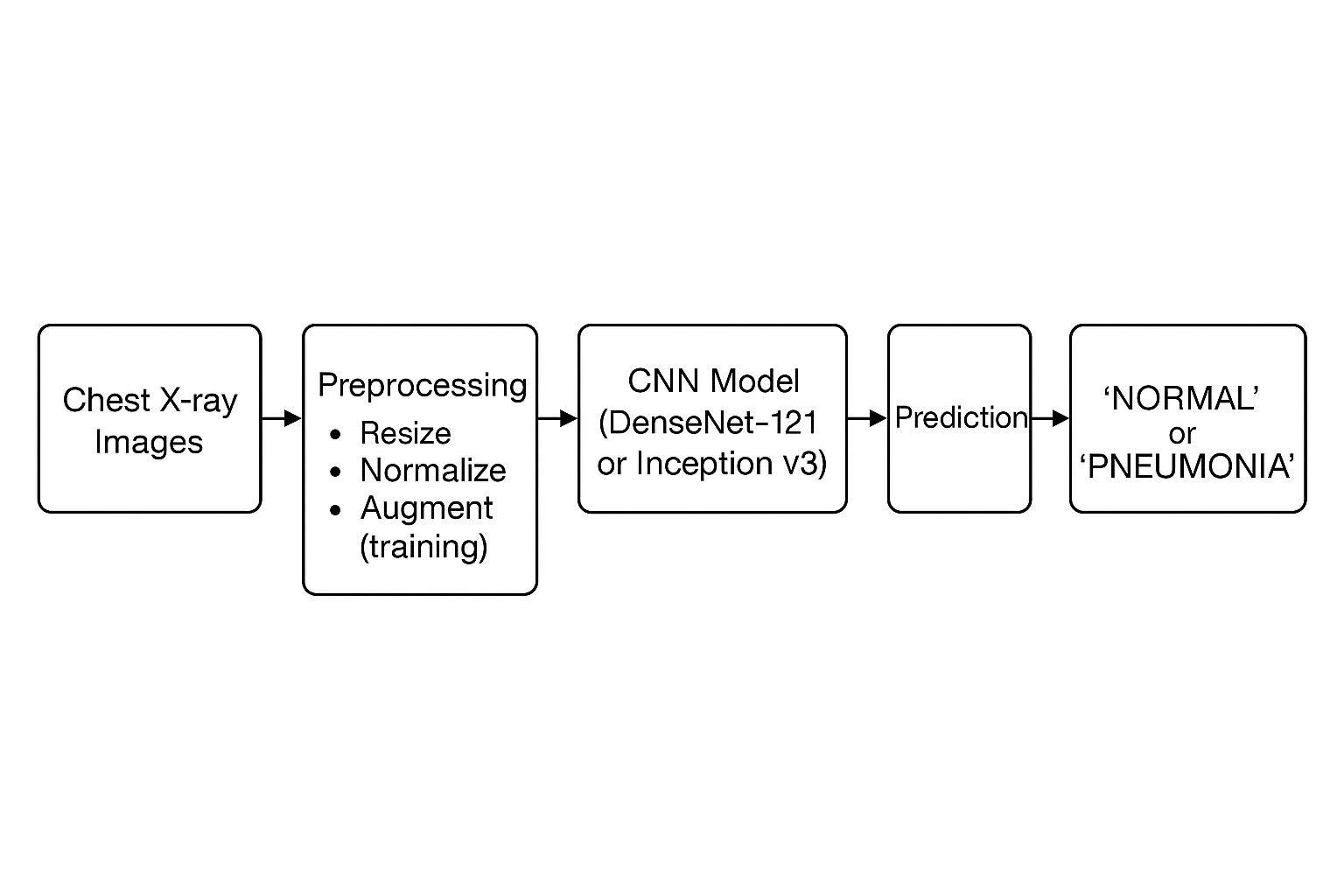
Our workflow began with the pediatric chest X-ray dataset from Guangzhou Women and Children’s Medical Center, focusing on a binary classification task: “NORMAL” versus “PNEUMONIA.” After inspecting the data, we resized the images to match model-specific input dimensions (224×224 for DenseNet‑121 and 299×299 for Inception v3). All images were normalized using ImageNet mean and standard deviation values. For augmentation, we applied random horizontal flips to training data to improve robustness, while validation and test sets remained untouched to preserve evaluation consistency.

We fine-tuned both CNN models using cross-entropy loss and the Adam optimizer, with a batch size of 32 over 10 training epochs. The models were trained and evaluated on GPU-enabled environments using PyTorch. We maintained a controlled pipeline, keeping all hyperparameters and data splits constant between models to ensure a fair architectural comparison. Each model was evaluated on a held-out test set of 624 images using accuracy, ROC-AUC, and confusion matrices as evaluation metrics.

We considered other architectures like ResNet-50 and VGG16, but chose DenseNet and Inception for their proven track records in both general-purpose and medical image classification tasks. DenseNet’s dense connectivity pattern encourages feature reuse and mitigates the vanishing gradient problem, while Inception’s modular design captures features at multiple scales, which is valuable in detecting the often-subtle visual cues of pneumonia.

This approach is pedagogically aligned with concepts we studied throughout the semester, including supervised learning, transfer learning, model evaluation using ROC-AUC, and data augmentation strategies. We applied practical knowledge in deep learning architecture design, optimization, and image preprocessing to build a complete end-to-end classification system. Figures such as confusion matrices and ROC curves were included to visually analyze model performance, and the controlled experimentation setup allowed us to draw reliable conclusions about the relative strengths of each model architecture.

| **Model** | **Input Size** | **Architecture Style** | **Test Accuracy** | **ROC-AUC** | **Trainable Parameters (Approx.)** |
| --- | --- | --- | --- | --- | --- |
| DenseNet‑121 | 224×224 | Dense blocks with skip connections | 69.4 % | 0.916 | 7.98 million |
| Inception v3 | 299×299 | Multi-scale Inception modules | 75.8 % | 0.967 | 23.85 million |



**Experiments and Results:**

To evaluate the performance of our pneumonia detection system, we conducted a series of controlled experiments comparing two CNN architectures: **DenseNet‑121** and **Inception v3**. Both models were fine-tuned on the same pediatric chest X-ray dataset with identical preprocessing, data augmentation, training duration, and optimizer settings to ensure a fair comparison.

**Evaluation MetricsAccuracy**: Measures overall correctness of predictions

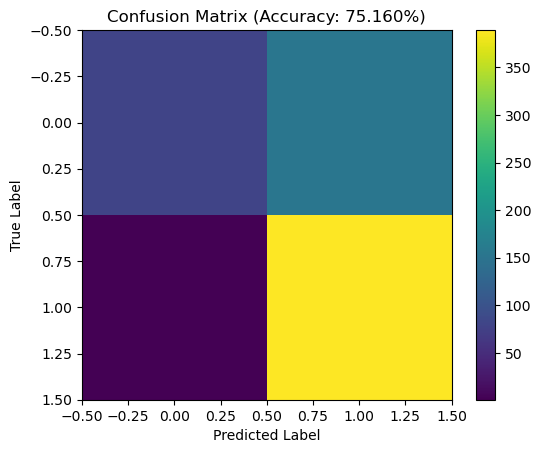
* **ROC-AUC (Receiver Operating Characteristic – Area Under Curve)**: Indicates the model’s ability to rank positive instances higher than negative ones
* **Confusion Matrix**: Provides insights into true/false positives and negatives

**Results Summary**

| **Model** | **Test Accuracy** | **ROC-AUC** |
| --- | --- | --- |
| DenseNet‑121 | 69.4% | 0.916 |
| Inception v3 | 75.8% | 0.967 |

From these results, **Inception v3 consistently outperformed DenseNet‑121**, especially in ROC-AUC, highlighting its superior ability to distinguish between pneumonia and normal cases. The confusion matrices further support this, showing that Inception v3 had fewer false negatives—crucial in a clinical setting where missing a pneumonia case could be life-threatening.

**A graph of a curve

AI-generated content may be incorrect.**

**A graph of a curve

AI-generated content may be incorrect.A chart with different colored squares

AI-generated content may be incorrect.**

**A graph of a curve

AI-generated content may be incorrect.**

**Visualizations**

* **Figure 1**: Confusion Matrix – DenseNet‑121
* **Figure 2**: ROC Curve – DenseNet‑121 (AUC = 0.916)
* **Figure 3**: Confusion Matrix – Inception v3
* **Figure 4**: ROC Curve – Inception v3 (AUC = 0.967)
* **Figure 5**: Combined ROC Curve comparing both models**Architectural Insights**

DenseNet‑121, though efficient with fewer parameters, may have struggled with the higher-resolution spatial patterns in chest X-rays due to its smaller input size (224×224). In contrast, Inception v3’s larger input size (299×299) and multi-scale convolutional structure allowed it to capture richer features, which contributed to its better performance.

**Alternative Considerations**

We considered experimenting with **ResNet-50** and **VGG16**, but prioritized DenseNet and Inception for their architectural strengths and proven performance in prior literature. No ablation studies or ensembling methods were performed in this phase due to time constraints, but they remain viable avenues for future enhancement.

**Limitations and Observations**

* **Overfitting**: In early experiments with fewer augmentations, DenseNet showed signs of overfitting.
* **Class imbalance**: The dataset was heavily skewed toward pneumonia cases, which we mitigated through stratified sampling and balanced evaluation metrics like ROC-AUC.
* **Model interpretability**: Currently, no Grad-CAM visualizations were integrated, which could help validate the regions of focus for each prediction.

Conclusion:

In this project, we explored deep learning-based methods for automated pneumonia detection using pediatric chest X-rays. By applying transfer learning to two prominent convolutional neural network architectures—**DenseNet‑121** and **Inception v3**—we assessed their effectiveness on a binary classification task: identifying whether an X-ray image corresponds to a “NORMAL” or “PNEUMONIA” case.

Our experiments demonstrated that **Inception v3 outperformed DenseNet‑121** in both accuracy (75.8% vs. 69.4%) and ROC-AUC (0.967 vs. 0.916), highlighting its superior ability to capture multi-scale features and process higher-resolution inputs. These findings reinforce the importance of model architecture and input size in medical image classification tasks. The use of ROC-AUC as a primary evaluation metric proved especially valuable in addressing class imbalance and assessing real-world diagnostic reliability.

Through this work, we’ve gained insight into the strengths of different CNN architectures when applied to clinically relevant datasets. We also validated the effectiveness of standard preprocessing techniques, such as resizing, normalization, and data augmentation, in improving model generalization.

**Future Work**

To build on this foundation, several future extensions can be explored:

* **Model interpretability**: Integrate techniques such as **Grad-CAM** to visualize attention maps and understand model decision-making.
* **Ensemble learning**: Combine predictions from multiple architectures to boost robustness and accuracy.
* **Threshold tuning**: Calibrate decision thresholds to optimize sensitivity or specificity based on clinical context.
* **Data expansion**: Incorporate adult chest X-ray datasets or multi-label disease detection to broaden applicability.
* **Deployment pipeline**: Develop a lightweight, explainable inference system that can assist clinicians in real-time screening workflows, especially in low-resource settings.

This work illustrates the potential of AI in augmenting healthcare delivery and lays the groundwork for future enhancements toward safe, scalable diagnostic tools.