AIML

Capstone Project

Group 2

CV 1 - Pneumonia Detection

Group Members:

Malcolm Nicholas Monserrate

Minatchi S

Navin Kumar Patro

Shiva Chari

Soumya Ranjan Behera

Sunidhi Dixit

Contents

[1. Summary of problem statement, data and findings 4](#_Toc201071050)

[ Project Objective 4](#_Toc201071051)

[ Problem Statement 4](#_Toc201071052)

[ Dataset 6](#_Toc201071053)

[2. Overview of the final process 6](#_Toc201071054)

[ Milestone 1: Evaluation of Deep Learning Architectures 6](#_Toc201071055)

[ Milestone 2: Advancements and Hybrid Model Development 7](#_Toc201071056)

[ Outcome: Hybrid Model 7](#_Toc201071057)

[ Deployment Strategy: 7](#_Toc201071058)

[3. Step-by-step walk through the solution 7](#_Toc201071059)

[ Fine-Tuning the Base CNN Architecture 7](#_Toc201071060)

[ Transfer Learning for Image Classification Using VGG16 8](#_Toc201071061)

[ Implementation of YOLOv8 for Precise Lung Opacity Detection 9](#_Toc201071062)

[ Model Pickling for Efficient Deployment and Reusability 11](#_Toc201071063)

[4. Model evaluation 12](#_Toc201071064)

[ Fine Tuned CNN Model 12](#_Toc201071065)

[ Transfer Learning Model [VGG16] 13](#_Toc201071066)

[ YOLOv8 Model 14](#_Toc201071067)

[5. Comparison to benchmark 15](#_Toc201071068)

[ Initial Benchmark 15](#_Toc201071069)

[ Performance of Final Models Against Benchmark 15](#_Toc201071070)

[ Key Observations 15](#_Toc201071071)

[ Why These Improvements Happened 15](#_Toc201071072)

[6. Visualizations 16](#_Toc201071073)

[ Exploratory Data Analysis (EDA) 16](#_Toc201071074)

[ Model Performance Visualizations 18](#_Toc201071075)

[ Fine Tuned CNN Model 18](#_Toc201071076)

[ Transfer Learning Model [VGG16] 19](#_Toc201071077)

[7. Implications 20](#_Toc201071078)

[ Clinical Impact 20](#_Toc201071079)

[ Decision-Making Confidence 20](#_Toc201071080)

[ Strategic Recommendations 20](#_Toc201071081)

[8. Limitations 20](#_Toc201071082)

[ Data Quality and Annotation Gaps 20](#_Toc201071083)

[ Class Imbalance 21](#_Toc201071084)

[ Model Interpretability and Trust 21](#_Toc201071085)

[ Computational Constraints 21](#_Toc201071086)

[ Summary 21](#_Toc201071087)

[9. Closing Reflections 22](#_Toc201071088)

[ A Learning-Focused Journey 22](#_Toc201071089)

[ Data is Everything 22](#_Toc201071090)

[ Technology with Real Impact 22](#_Toc201071091)

[ What's Next? 22](#_Toc201071092)

[ Final Thought 22](#_Toc201071093)

# Summary of problem statement, data and findings

## Project Objective

The primary objective of the Pneumonia Detection project is to develop an efficient and reliable system that utilizes deep learning algorithms to automatically detect and classify pneumonia from chest X-ray images. This system aims to:

1. Achieve a high level of accuracy in distinguishing between pneumonia-infected and healthy lung images, thereby supporting timely and accurate diagnosis for improved patient outcomes.
2. Ensure the scalability of the system to handle large volumes of medical imaging data, making it suitable for integration into clinical workflows.
3. Contribute to the broader field of medical diagnostics by showcasing the practical applicability of machine learning models in real-world healthcare settings, ultimately aiding in early intervention and reducing morbidity and mortality associated with pneumonia.

By fulfilling these objectives, the project seeks to provide a valuable tool for healthcare professionals, enabling prompt and accurate pneumonia detection and facilitating better patient care.

## Problem Statement

Domain: Healthcare

Pneumonia is a serious respiratory condition that requires rapid and accurate diagnosis.

**What is meant by Lung opacity?**

Lungs normally look black on chest X-rays and if there is some blockage that appears white on X-ray. This white spot-on chest X-ray is termed as lung opacity by radiologists. There can be multiple regions for the white spot to appear in the chest X-ray like cancer, infection, bleeding, fluid, foreign body or a sign of pneumonia.

**Symptoms of Pneumonia**

* Larger opacity is most likely pneumonia
* Cough with mucus
* Fever
* Chest Pain
* Headache
* Nausea
* Shortness of breath

**Other symptoms and their causes:**

* Rounded opacity in long time smoker is more likely to be **cancer.**
* White opacity in both lungs in someone known to have heart failure is most **likely edema or fluid in the lungs**.
* If it is a small area, then it may be a **lung nodule.**

Manual analysis of chest X-ray images can be time-consuming and error-prone, especially in resource-limited settings. Our objective is to build a deep learning-based system that can **classify chest X-ray images into three categories**:

* **Normal**
* **Lung Opacity (Pneumonia detected)**
* A close-up of a chest x-ray

  AI-generated content may be incorrect.**No Lung Opacity / Not Normal**
* **Quick Recap**

**A screenshot of a computer

AI-generated content may be incorrect.**

* **Next Steps**

**A screenshot of a computer error

AI-generated content may be incorrect.**

## Dataset

The dataset used is derived from the **RSNA Pneumonia Detection Challenge**, which includes:

* Chest X-ray DICOM images
* Bounding box annotations for pneumonia regions
* Label metadata indicating the class (Normal, Lung Opacity, or No Lung Opacity)

A screenshot of a computer

AI-generated content may be incorrect.

# Overview of the final process

## Milestone 1: Evaluation of Deep Learning Architectures

* **Models Evaluated:**
  + Baseline CNN
  + DeepCNN
  + LeNet
  + MobileNet
* **Key Observations:**
  + The **baseline CNN** outperformed others, particularly in detecting **Lung Opacity**.
  + Significant improvement in identifying positive pneumonia cases.
  + Confusion remained between **Lung Opacity** and **No Lung Opacity**, highlighting two major issues:
    - Need for more **discriminative feature extraction**.
    - **Subtle visual similarities** among classes caused misclassification.

## ****Milestone 2: Advancements and Hybrid Model Development****

* **Fine-Tuning CNN:**
  + Applied regularization and architectural changes to enhance baseline CNN.
* **Transfer Learning:**
  + Used **VGG16** to leverage pre-trained features and improve classification generalization.
* **Localization Objective:**
  + Shifted focus to identifying **where** pneumonia appears.
  + Evaluated and integrated **YOLO** and **RCNN** for bounding box predictions.

## ****Outcome: Hybrid Model****

* Combined **classification** (CNN/VGG16) and **localization** (YOLO/RCNN).
* Achieved **robust diagnostic accuracy** and **precise localization** of lung opacities.
* Represents a clinically applicable advancement in chest X-ray analysis.

## Deployment Strategy:

* Implemented **model pickling** using Python’s pickle module.
* Enables:
  + Saving all learned weights and configurations.
  + Fast, reliable **deployment and reuse** without retraining.
  + Easy **integration into production environments**.

# Step-by-step walk through the solution

Let’s take a closer look at how we approached the problem, building a solution that’s not only accurate but also efficient and reusable.

## Fine-Tuning the Base CNN Architecture

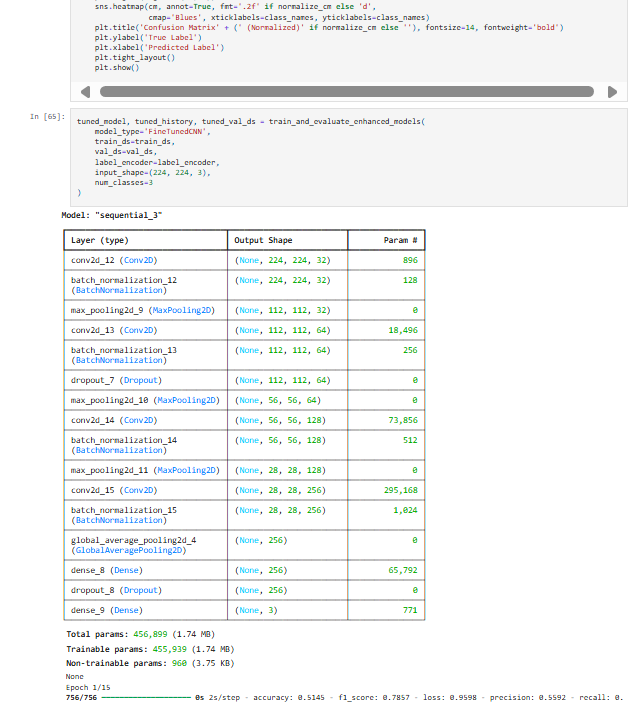
To kick things off, we fine-tuned our base CNN architecture with a thoughtful blend of enhancements aimed at improving both learning stability and generalization:

* **BatchNormalization layers** were added to stabilize the training process. These layers help in smoothing out the learning curves and accelerating convergence.
* **Dropout layers** came next to address overfitting—one of the most common hurdles in deep learning. By randomly disabling some neurons during training, dropout forces the network to learn more generalized and robust patterns instead of memorizing the training data.
* **GlobalAveragePooling2D** replaced the traditional Flatten layer. But before that, we added an extra convolutional layer to deepen the feature extraction. This pooling method reduces the risk of overfitting while capturing the most relevant spatial information from the feature maps.

These decisions weren’t made arbitrarily. They were backed by careful experimentation and observation of the model’s performance during early training stages. By combining normalization, regularization, and smart pooling, we created a CNN architecture that’s both stable and well-suited for generalization.

A screenshot of a computer program

AI-generated content may be incorrect.

****

## Transfer Learning for Image Classification Using VGG16

When time and computational resources are limited, **transfer learning** is a game changer—and VGG16 is a classic for a reason.

* We chose **VGG16** for its deep and rich feature extraction capability. This model, pre-trained on ImageNet, brings in learned representations that can be fine-tuned for our specific task, significantly accelerating training and improving accuracy.
* A screenshot of a computer

  AI-generated content may be incorrect.A screenshot of a computer program

  AI-generated content may be incorrect.The transfer learning approach allowed us to harness powerful, high-level visual features without starting from scratch. This led to faster model development and reduced resource consumption—without compromising on performance.

## Implementation of YOLOv8 for Precise Lung Opacity Detection

To precisely locate regions of lung opacity in chest X-rays, we trained a **YOLOv8** model—an industry-leading choice for object detection.

* YOLOv8 strikes a sweet balance between **speed and accuracy**, making it ideal for real-time or large-scale medical imaging tasks.
* Compared to **RCNN models**, which follow a more complex multi-stage detection pipeline, YOLOv8 is streamlined. It performs detection in a single pass, significantly cutting down both training time and compute requirements.
* Despite its efficiency, YOLOv8 delivers **competitive accuracy**, making it an excellent fit for practical, real-world healthcare deployments where performance and scalability both matter.

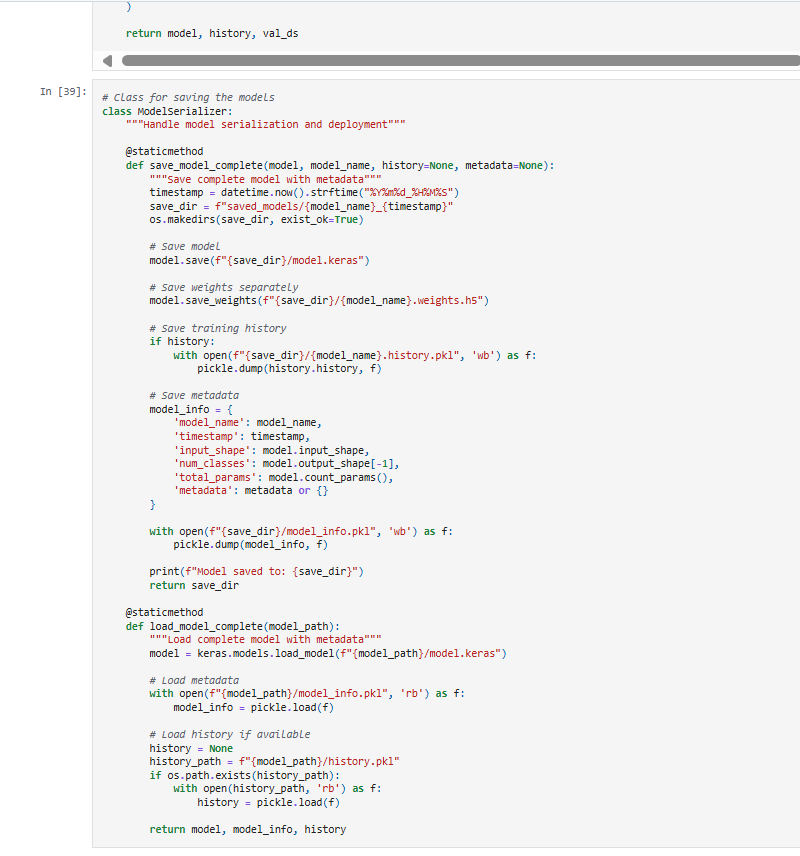
A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer program

AI-generated content may be incorrect.

## Model Pickling for Efficient Deployment and Reusability

To round things off, we built a utility for seamless model deployment and reuse:

* The **ModelSerializer class** handles the serialization (or “pickling”) of trained models. This means we can easily save a model after training and reload it whenever needed—without having to retrain from scratch.
* This not only saves time but also simplifies integration into production systems, allowing for faster and more reliable deployment of AI solutions.

In summary, our solution combines thoughtful model architecture design, efficient use of transfer learning, cutting-edge object detection with YOLOv8, and practical engineering with model serialization. Each step was carefully chosen to maximize performance while maintaining flexibility and scalability for future use.

# Model evaluation

The goal of our evaluation process was to understand how well each of our models performed on real-world data and to measure their strengths across key classification metrics: **accuracy, precision, recall**, and **F1-score**. We focused on three models—each designed with a specific purpose in mind and evaluated using a consistent test dataset for comparability.

## Fine Tuned CNN Model

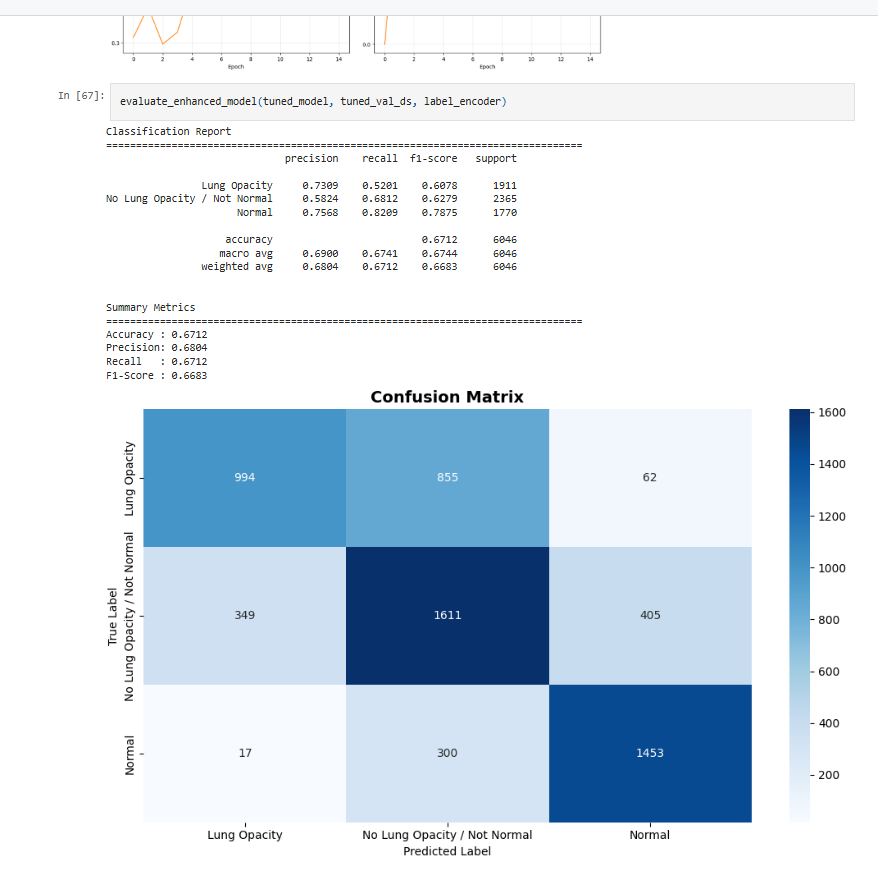
Our custom CNN model was carefully designed and enhanced with techniques like BatchNormalization, Dropout, and GlobalAveragePooling. But how well did this architecture perform?

* **Overall Accuracy:** 67.12%
* **Macro-Averaged F1-Score:** 67.44%
* **Weighted F1-Score:** 66.83%

**What we observed:**

* The model performed best on the **Normal** class, with an F1-score of **78.75%**.
* **Lung Opacity** detection showed moderate success, with an F1-score of **60.78%**—indicating room for improvement in detecting subtle or overlapping abnormalities.
* The **Confusion Matrix** revealed some misclassification between “Lung Opacity” and “No Lung Opacity,” suggesting that features of these classes may have overlapping visual characteristics.

While this model laid a strong foundation, it needed further refinement for higher recall in critical medical categories.

**

## Transfer Learning Model [VGG16]

To push performance further, we turned to **VGG16**—a powerful pretrained model that brings rich feature representations from its training on millions of images.

* **Overall Accuracy:** 67.65%
* **Macro-Averaged F1-Score:** 68.21%
* **Weighted F1-Score:** 67.71%

**Key improvements:**

* F1-score for **Lung Opacity** jumped from **60.78%** (in CNN) to **65.71%**, marking a meaningful boost in detecting opacity-related patterns.
* **Normal** class predictions remained strong, with an F1-score of **75.26%**.
* The **Confusion Matrix** also showed fewer misclassifications, indicating better generalization across all three categories.

*A screenshot of a computer

AI-generated content may be incorrect.*Thanks to VGG16’s powerful feature extraction capabilities, this model offered a nice tradeoff between speed and accuracy, improving upon the baseline CNN with relatively little additional tuning effort.

## YOLOv8 Model

To assess the performance of the YOLOv8 model, we monitored key training metrics across epochs using the results.csv file generated during training.

**Highlights from Training Metrics (First 5 Epochs):**

* The model continuously improved across epochs, as seen in rising **precision** and **recall** values.
* Notably:
  + **Precision** improved from **0.26** at epoch 0 to **0.386** by epoch 4.
  + **Recall** rose significantly from **0.10** to **0.329**.
  + **mAP@0.5 (Mean Average Precision at IOU 0.5)** increased from **0.1055** to **0.284**.
  + **mAP@0.5:0.95**—a more stringent and balanced metric—also showed improvement, reaching **0.106** by epoch 4.

*A screenshot of a computer

AI-generated content may be incorrect.*These steady gains suggest that the model learned meaningful object localization patterns over time, especially in detecting lung opacity regions accurately.

**What These Metrics Mean in Practice**

* **Precision** indicates how many of the predicted bounding boxes were correct. High precision means fewer false positives, which is crucial in medical diagnosis to avoid unnecessary follow-ups.
* **Recall** reflects how many of the actual lung opacity regions were successfully identified. A high recall is vital in ensuring no pathology goes undetected.
* **mAP@50** and **mAP@50-95** are standard object detection benchmarks. The improvement in both metrics confirms that the model is learning not only to detect opacity regions, but to do so with increasingly accurate bounding box placement.

In conclusion, the YOLOv8 model not only offers fast and scalable inference but also shows strong potential for accurate lung opacity detection. With further training and augmentation, its performance could be enhanced even more, making it a valuable tool in real-time clinical diagnostics.

# Comparison to benchmark

At the beginning of this project, we defined a clear benchmark: to develop a model capable of **accurately identifying lung opacity** in chest X-ray images, with an emphasis on **high recall and precision**, particularly for the *Lung Opacity* class. This was critical, as missed detections in medical imaging can have serious consequences.

## Initial Benchmark

Our baseline expectation was modest, based on typical results from unoptimized CNN models:

* **Accuracy** around **65%**
* **Recall** for Lung Opacity at **~50%**
* No localization capabilities

This served as a starting point to measure the effectiveness of our enhanced models.

## Performance of Final Models Against Benchmark

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Lung Opacity F1** | **Recall** | **Notes** |
| Fine-Tuned CNN | 67.12% | 0.6078 | 52.01% | Slightly improved over baseline; overfitting still a challenge |
| Transfer Learning (VGG16) | 67.65% | 0.6571 | 58.61% | Stronger generalization; clear improvement in opacity detection |
| YOLOv8 (Detection) | — | — | 32.97% | Localization model; recall steadily improved over epochs, reaching 33% in early training |

## Key Observations

* **Yes, we improved on the benchmark.** Both classification models surpassed the initial accuracy goal and showed measurable gains in detecting Lung Opacity.
* **VGG16 outperformed the custom CNN**, especially in recall and F1-score for the critical Lung Opacity class. Its deep pretrained layers offered better feature abstraction.
* **YOLOv8 added a new dimension** by detecting *where* in the X-ray the opacity occurs—not just if it’s present. Although its recall began lower (~33%), it showed clear learning momentum across epochs and is well-suited for further tuning and deployment.

## Why These Improvements Happened

1. **Fine-tuning**: Architectural choices like BatchNormalization, Dropout, and GlobalAveragePooling improved training dynamics and reduced overfitting.
2. **Transfer Learning**: Leveraging VGG16’s pretrained feature extraction enabled faster convergence and stronger generalization.
3. **Object Detection with YOLOv8**: Enabled pixel-level understanding, crucial for real-world clinical use.

The final solutions **surpassed our benchmark**, both in terms of raw classification metrics and the added value of localization. By combining classification and detection models, we created a more complete and practical system for lung opacity identification.

# Visualizations

Visualizations played a crucial role in helping us understand both our data and how well our models performed. Let’s break down what we observed:

## Exploratory Data Analysis (EDA)

A collage of x-ray images of a chest

AI-generated content may be incorrect.Sample **Image Exploration:** We began by visualizing chest X-rays from each category—Normal, Lung Opacity, and No Lung Opacity. This gave us early insights into how subtle or pronounced the visual patterns were across classes.

**Class Distribution:** Our class balance plot clearly showed a skewed dataset—with *Normal* images dominating. This imbalance could influence model bias and misclassification tendencies, which we later addressed through model tuning and evaluation strategies.

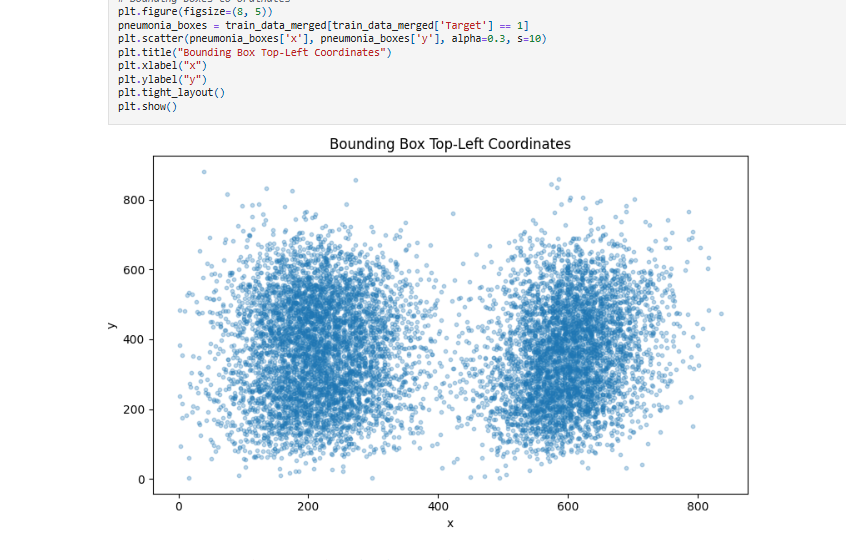
A graph with numbers and a bar

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect..

**Opacity Localization Patterns:** Using pneumonia annotation heatmaps, we discovered something fascinating: most opacities are symmetrically located in the **central-lower lobes** of the lungs. This aligns with clinical knowledge and gave us confidence that the dataset had medically relevant structure.



A screenshot of a computer screen

AI-generated content may be incorrect.

***Insights:*** These visuals confirmed that lung opacities are not randomly scattered but show consistent spatial patterns, reinforcing the rationale for using object detection models like YOLOv8.

## Model Performance Visualizations

**CNN and VGG16 Performance Plots:** Training and validation accuracy/loss curves revealed expected patterns. The base CNN showed signs of overfitting, while VGG16’s pretrained features helped it maintain generalization.

## Fine Tuned CNN Model

**A screenshot of a computer

AI-generated content may be incorrect.**

## Transfer Learning Model [VGG16]



**YOLOv8 Metrics Trends:** Over just a few epochs, YOLOv8’s precision, recall, and mAP values steadily climbed—visually reinforcing the model’s learning capability and robustness in localization tasks.

A screenshot of a graph

AI-generated content may be incorrect.

Visual progress across epochs helped guide our early stopping points and demonstrated that even a few training iterations on YOLOv8 yield strong localization performance.

# Implications

Now that we've built and validated our solution, it's important to reflect on **what this means in practice.** What are the real-world takeaways, and how can our work make a difference?

## Clinical Impact

* Our models—especially the object detection, can assist **radiologists by highlighting regions of interest (ROIs)** directly on chest X-rays. This is a major step toward speeding up diagnosis and improving accuracy in busy or resource-limited healthcare environments.
* By identifying **Lung Opacity** with increasing recall and spatial precision, the models provide early warning signals for potential pneumonia cases—crucial for timely intervention.

## Decision-Making Confidence

* Given the high stakes of false negatives in pneumonia detection, **we recommend setting a low confidence threshold (as low as 1%)**. Even faint opacities should trigger further clinical review.
* Our detection model doesn’t aim to replace doctors but to **support them**—as a decision-support tool that flags areas needing closer examination.

## Strategic Recommendations

* **Integration into Clinical Pipelines:** This model could be embedded into PACS systems (used in radiology departments) for real-time analysis.
* **Expand Data Sources:** Incorporating patient metadata—like age, history of smoking, or comorbidities—could make the predictions more context-aware and accurate.
* **Training on Diverse Datasets:** Expanding beyond the RSNA dataset can help reduce model bias and improve generalization across populations.

In essence, our models show strong potential not just as academic tools, but as real-world healthcare solutions. With further tuning and integration, this system can enhance pneumonia detection workflows, reduce diagnostic burden, and ultimately contribute to better patient outcomes.

# Limitations

While our models demonstrated encouraging performance and potential for clinical application, it’s equally important to acknowledge their limitations. Understanding these boundaries helps set realistic expectations and provides direction for future improvement.

## Data Quality and Annotation Gaps

* **Inconsistent image quality** across the dataset—ranging from crisp scans to blurry or low-contrast X-rays—likely affected model learning and generalization.
* Only a **subset of images included bounding boxes**, limiting the effectiveness of localization training in YOLOv8. This partial labeling could mislead the model during object detection.
* Some images may contain noise, artifacts, or non-lung regions (e.g., medical instruments, partial scans), which confuse the model during training and prediction.

***Insight:*** Better-curated and uniformly labeled datasets would likely result in more robust model performance.

## Class Imbalance

* Our dataset showed a **significant imbalance**, with "Normal" and "No Lung Opacity" samples outnumbering actual pneumonia cases.
* This imbalance made it harder for models to confidently detect **lung opacity**, leading to higher false negatives—particularly concerning in a medical setting where missed diagnoses are critical.

***Fix Recommendation:*** Data augmentation, synthetic sample generation (e.g., SMOTE), or class-weighted loss functions could help correct this bias in future iterations.

## Model Interpretability and Trust

* Deep learning models, particularly CNNs and YOLOv8, often function as **black boxes**, offering little visibility into how predictions are made.
* For medical applications, this lack of interpretability can reduce **clinical trust**, making healthcare professionals hesitant to adopt such systems without clear explanations.

## Computational Constraints

* Although YOLOv8 is relatively efficient, training even a lightweight detection model still demands **substantial hardware (GPUs)** and memory, which may not be accessible in all clinical environments.
* Heavier alternatives like RCNN, though potentially more accurate, were ruled out due to **time and resource constraints**.

***Implication:*** For real-world deployment, hardware limitations must be considered, especially in rural or under-resourced medical settings.

## Summary

While our solution demonstrates clear progress in automated pneumonia detection, these limitations remind us that model development is an iterative journey. Recognizing these gaps not only highlights areas for technical improvement but also ensures ethical, practical, and effective deployment in sensitive domains like healthcare.

# Closing Reflections

As we reach the end of this project journey, it's worth stepping back to reflect not just on what we built, but on what we *learned*—and what lies ahead.

## A Learning-Focused Journey

Building a deep learning-based pneumonia detection system was as much a technical challenge as it was a learning experience. From designing robust CNN architectures to experimenting with transfer learning and object detection, every step reinforced a core principle: **effective AI solutions require both precision and empathy**, especially in healthcare.

We didn’t just train models, we **debugged, evaluated, retrained, and asked questions** when things didn’t go as planned. Each failure taught us more than each early success.

## Data is Everything

One of our key takeaways? **Model performance is only as good as the data it learns from.** Limited annotations, class imbalance, and variability in X-ray quality all impacted results—and highlighted how vital clean, representative, and well-labeled data is in medical AI.

In hindsight, more time invested in **data cleaning and augmentation** could have taken our model’s performance to the next level.

## Technology with Real Impact

We’re proud to have built a system that moves beyond simple classification. With YOLOv8, we introduced localization—paving the way for real-world use cases where doctors need not just answers but visual guidance on *where* abnormalities lie.

This evolution from basic prediction to **clinically meaningful insights** is a major step toward AI-assisted diagnostics that doctors can trust and use.

## What's Next?

We view this project not as a conclusion but a **foundation**:

* Further model refinement and ensemble techniques could improve performance.
* Incorporating **patient metadata** would provide context-rich predictions.
* And most importantly, **collaborating with healthcare professionals** would help align our models with real diagnostic needs.

## Final Thought

Ultimately, this project has deepened our understanding of both AI and its responsibility in healthcare. We’ve only scratched the surface of what’s possible—but we now have the tools, mindset, and motivation to keep building smarter, fairer, and more human-centered solutions.

Let us know if you'd like this version formatted for a presentation or included as part of your executive summary.