TRANSFER LEARNING FOR RESPIRATORY DISEASE CLASSIFICATION IN CHEST X-RAYS

A Comparative Analysis of Transfer Learning vs. a From-Scratch Approach

DEPARTMENT OF COMPUTER SCIENCE,
SCHOOL OF ENGINEERING & TECHNOLOGY,
ISLAMIC UNIVERSITY OF SCIENCE AND TECHNOLOGY, AWANTIPORA, KASHMIR

PRESENTED BY:

- **SHAHID UL ISLAM (*MCA-23-07*)**

SUPERVISOR:

- **DR. SYED TANZEEL RABANI**

Submitted in partial fulfillment of the requirements for the award of the degree of "Master of Computer Applications"

PRESENTATION OUTLINE

1. Introduction & Motivation
2. Problem Statement & Objectives
3. Methodology & Experimental Setup
4. Results: Training and Test Performance
5. Discussion of Key Findings
6. Limitations & Future Work
7. Conclusion

INTRODUCTION : THE CLINICAL CHALLENGE

- Respiratory diseases like pneumonia and COVID-19 are major global health concerns, causing millions of hospitalizations annually.
- Chest X-rays (CXRs) are the primary diagnostic tool due to their accessibility and cost-effectiveness.
- However, manual interpretation is slow, subjective, and suffers from a global shortage of expert radiologists.
- This creates a critical need for automated, reliable diagnostic support tools.

Key Impact: Early detection can reduce mortality by 40-60% (WHO, 2023; CDC, 2023; Houck et al., The Lancet, 2004).

Problem Statement

MANUAL INTERPRETATION

The manual interpretation of CXRs is limited by human error, subjectivity, and diagnostic delays.

DEEP LEARNING MODELS

Deep learning models offer a solution but typically require massive datasets, which are scarce in the medical field

CORE PROBEM

How can we build an accurate, automated system for multi-class classification of respiratory diseases (Normal, Bacterial, Viral, COVID-19) despite data limitations?

KEY QUESTION

How effective is transfer learning compared to a model trained from scratch?

PROJECT OBJECTIVES

01

To prepare a comprehensive dataset of chest X-ray images for classification.

02

To establish a baseline by training a Custom CNN from scratch.

03

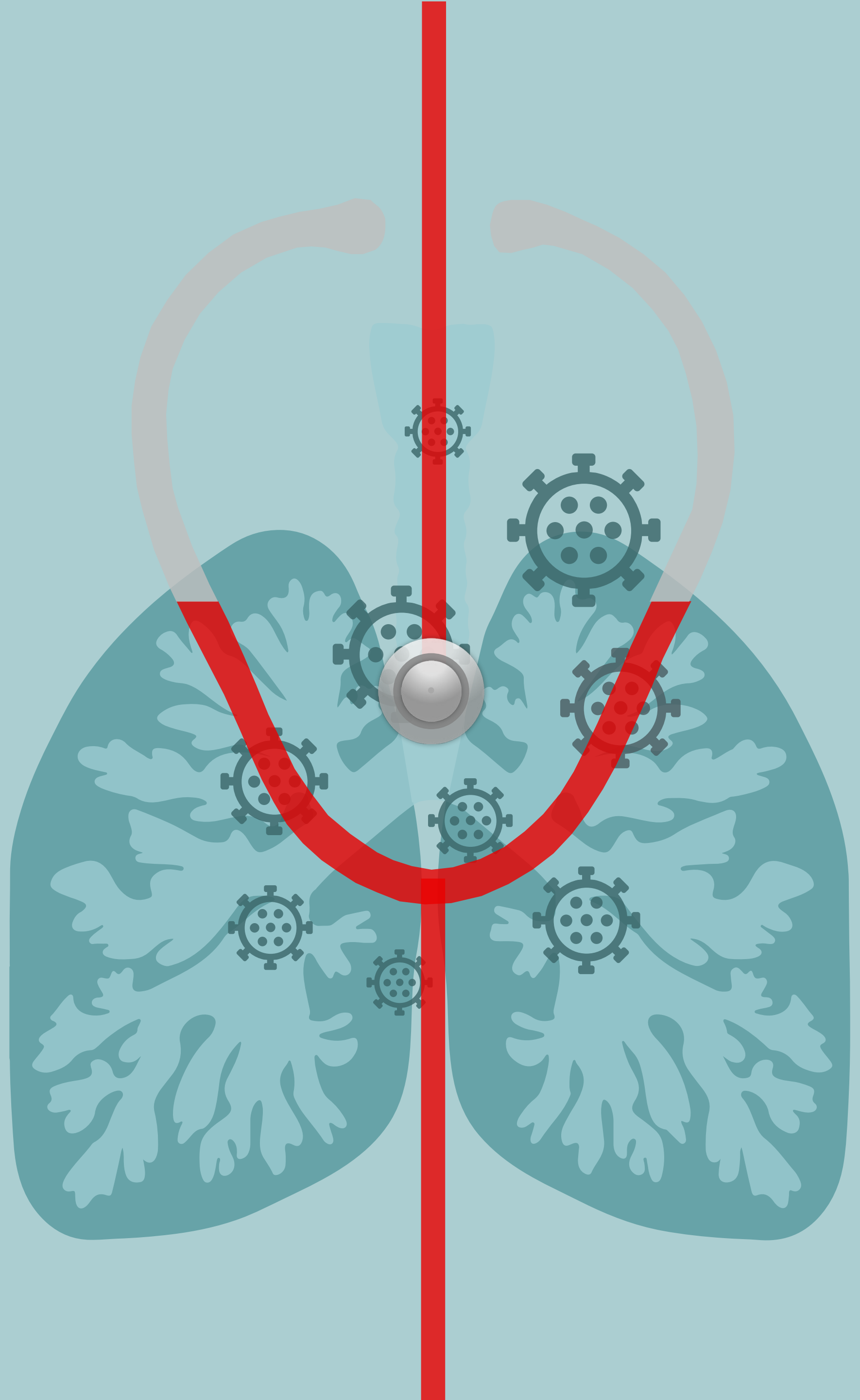
To implement and fine-tune a VGG16 model using Transfer Learning.

04

To implement and fine-tune a Vision Transformer (ViT) model using Transfer Learning.

05

To conduct a rigorous comparative analysis of all three models on performance and efficiency.



LITERATURE REVIEW & RESEARCH GAP

LITERATURE REVIEW

The literature confirms that fine-tuning pre-trained CNNs is the standard, effective approach for medical imaging (Tajbakhsh et al., 2016).

Vision Transformers (ViTs) are emerging as a powerful alternative but are less studied in this specific context (Dosovitskiy et al., 2021).

RESEARCH GAP

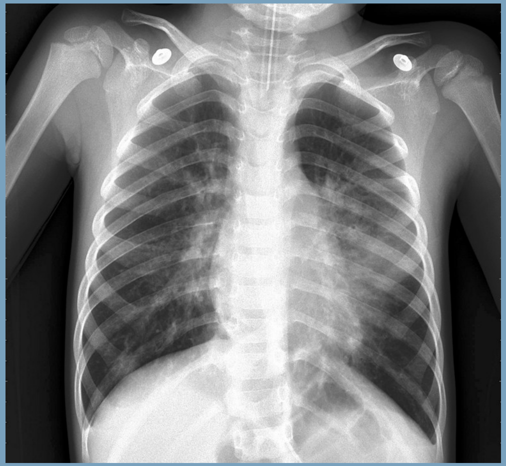
The Research Gap: A lack of direct, controlled studies comparing a classic CNN (VGG16), a ViT, and a from-scratch model on the same multi-class chest X-ray dataset.

Our Contribution: This project fills that gap by providing a direct empirical benchmark.

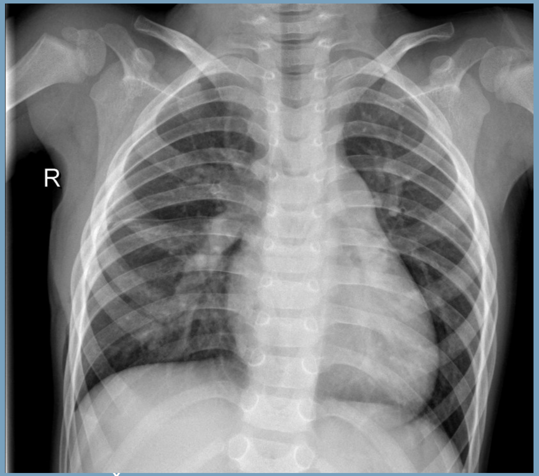
METHODOLOGY: OVERALL WORKFLOW

- **Data Collection (Kaggle Dataset)**
- **Preprocessing & Augmentation**
- **Data Splitting (80/10/10)**
- **Model Implementation (3 Arms)**
- **Training & Validation**
- **Testing & Performance Evaluation**

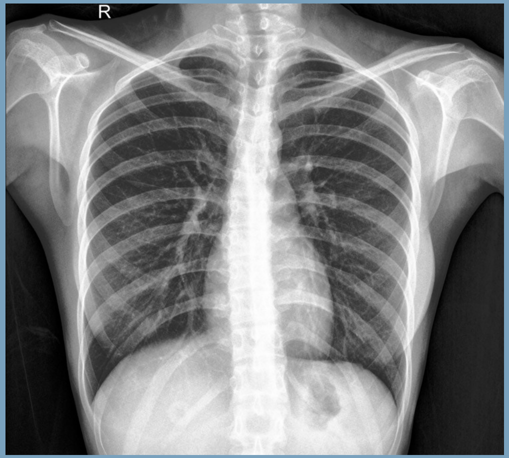
THE DATASET:
SOURCE AND PREPARATION



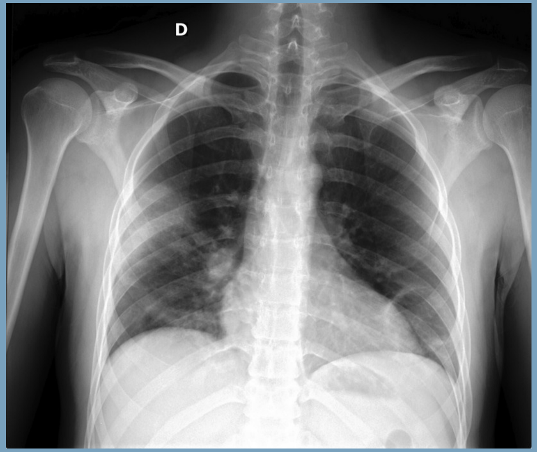
VIRAL PNEUMONIA



BACTERIAL PNEUMONIA



NORMAL



COVID-19

DATASET OVERVIEW

- **Source: Kaggle Pneumonia & COVID-19 Chest X-ray Dataset**
- **Total Images: 6,600 chest X-ray images**
- **Image Format: JPEG, PNG (256×256 to 1024×1024) Color: Grayscale medical imaging**

DATA QUALITY CHALLENGES

- **Varied image quality from different hospitals**
- **Inconsistent positioning and exposure levels**
- **Label noise in some COVID-19 samples**
- **Equipment differences across institutions**

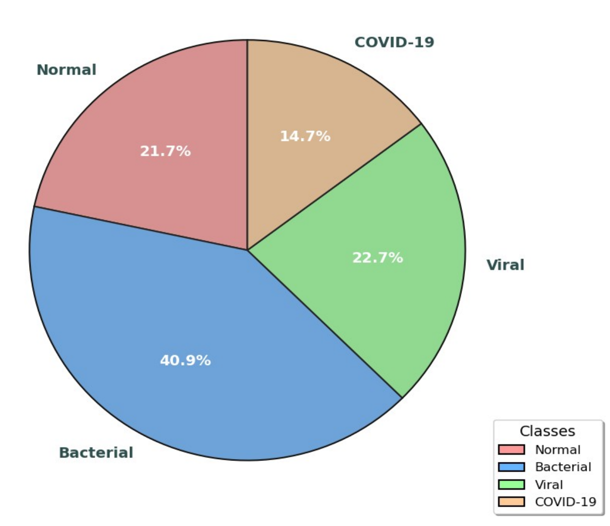
CLASS DISTRIBUTION

- **COVID19 : 980**
- **BACTERIAL : 2727**
- **VIRAL : 1512**
- **NORMAL : 1443**

Normal: 21.66% | Bacterial: 40.93%
Viral: 22.69% | COVID-19: 14.71%

CLASS IMBALANCE HANDLING

IMBALANCE STATISTICS




```
class_counts = {
    'Normal': 1443,
    'Bacterial': 2727,
    'Viral': 1512,
    'COVID-19': 980
    }
imbalance_severity =
max(class_counts.values()) /
min(class_counts.values())
# Result: 8.28 (High imbalance)
```

WHY WEIGHTED CROSS-ENTROPY LOSS

- Penalizes misclassification of minority classes more heavily
- Mathematically sound approach for imbalanced datasets
- Preserves original data distribution
- Computationally efficient compared to sampling methods

WEIGHT CALCULATION METHOD

Inverse Frequency Weighting

- Calculate Total Samples and Classes
- Compute Class Weights
- Assign Higher Weights to Minority Classes

LOSS FUNCTION IMPLEMENTATION

```
criterion =
nn.CrossEntropyLoss(

weight=torch.tensor([1.68,
0.61, 0.98, 5.02])
)
```

ALTERNATE METHODS CONSIDERED

- SMOTE: Risk of overfitting in medical images
- Oversampling: Increases training time significantly
- Focal Loss: Complex hyperparameter tuning
- ***Weighted CE: Simple, effective, interpretable***

DATA PREPROCESSING & AUGMENTATION

PREPROCESSING PIPELINE

Smart Preprocessing for Deep Vision: Optimized for CNNs and Transformers.

01

Common Steps

- Load image in grayscale
- Apply CLAHE (Contrast Limited Adaptive Histogram Equalization) to enhance local contrast
- Normalize pixel values to range [0, 1] by dividing by 255
- Convert grayscale to RGB
- Convert NumPy array to PIL Image for compatibility with PyTorch transform

02

VGG16-Specific Transforms

- Resize: 224×224 (ImageNet standard)
- Normalization: ImageNet mean/std
- Data type: Float32

03

ViT-Specific Transforms

- Resize: 384×384 (ViT-Base optimal)
- Center crop: 224×224 patches
- Patch tokenization: 16×16 patches

DATA PREPROCESSING & AUGMENTATION

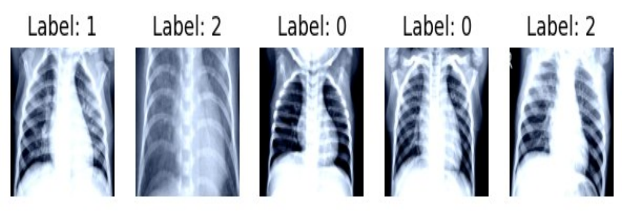
AUGMENTATION STRATEGIES

Optimizing Training Inputs: Augmentations Aligned with Architecture.

01

VGG16 AUGMENTATION

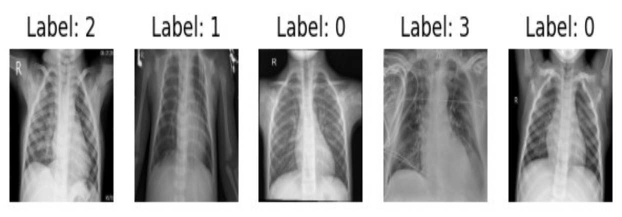
- Random Rotation
- Random Horizontal Flip
- Random Resized Crop
- Color Jitter (Brightness & Contrast)
- Random Affine (Rotation + Translation)



02

ViT B/16 AUGMENTATION

- Resize
- Center Crop
- Color Jitter (Hue & Saturation)
- Random Erasing



REASON FOR DIFFERENT PIPELINES

VGG16:

1. Uses aggressive augmentations to enhance invariance to position, scale, and lighting.
2. Helps CNNs learn robust features and generalize better on limited or varied datasets.

ViT B/16:

1. Requires minimal augmentations to preserve spatial and patch-level structure.
2. Heavy transforms can disrupt the token sequence and harm performance in ViTs.

Medical Imaging (Chest X-rays): Employs conservative preprocessing and augmentation to maintain diagnostic features without distorting subtle medical details essential for accurate classification.

EXPERIMENTAL SETUP: THE THREE MODELS

Arm 1

Custom CNN (From Scratch)

A 7-layer CNN with ~6.04M parameters.

Weights were initialized randomly.

Arm 2

VGG16 (Transfer Learning)

Classic CNN architecture, pre-trained on ImageNet.

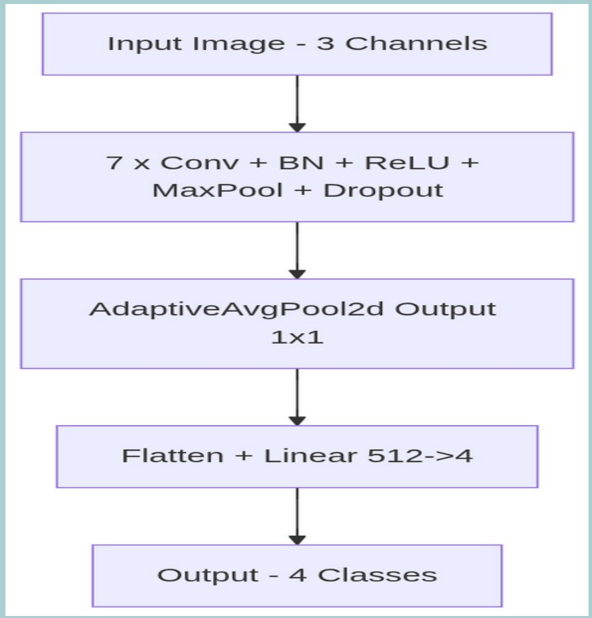
Fine-tuned on our dataset.

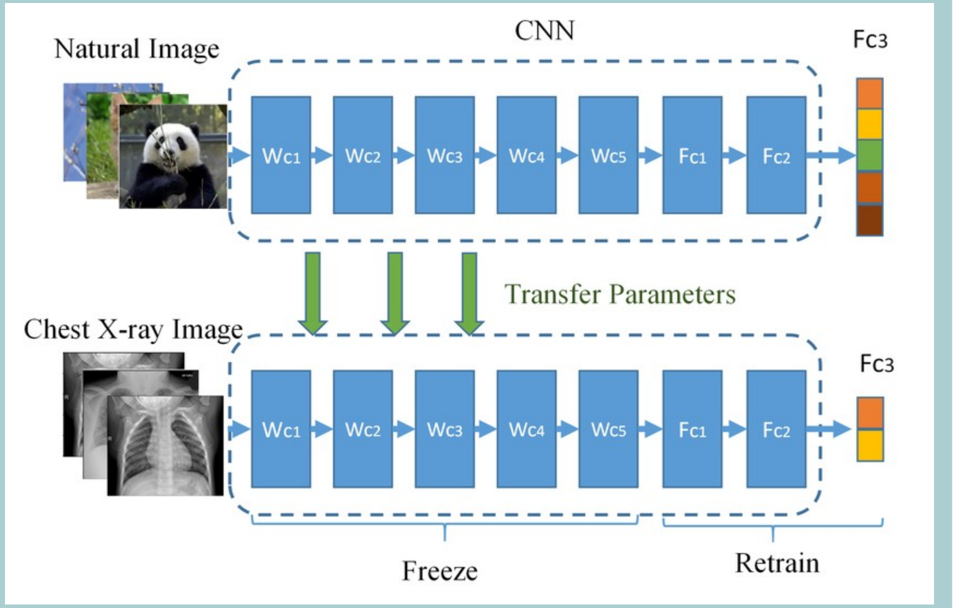
Arm 3

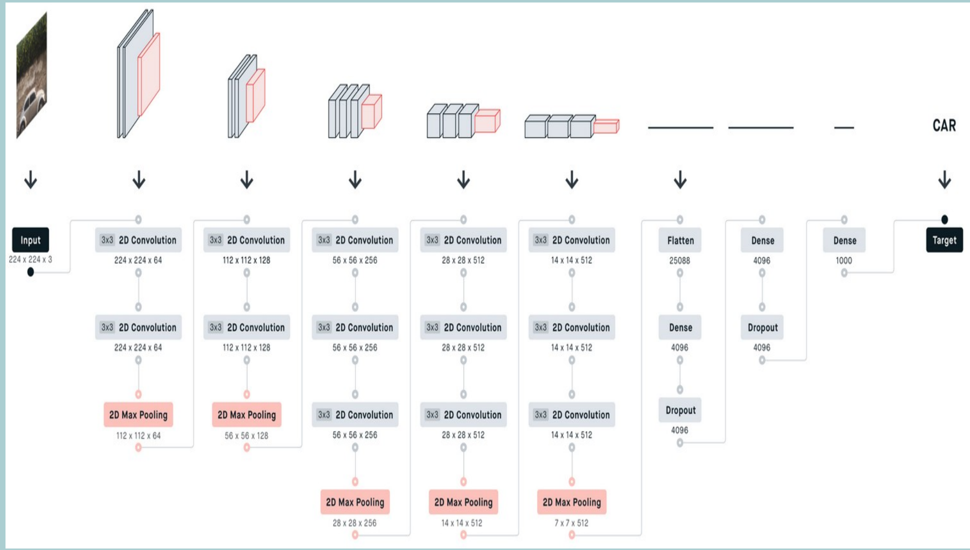
ViT (Transfer Learning)

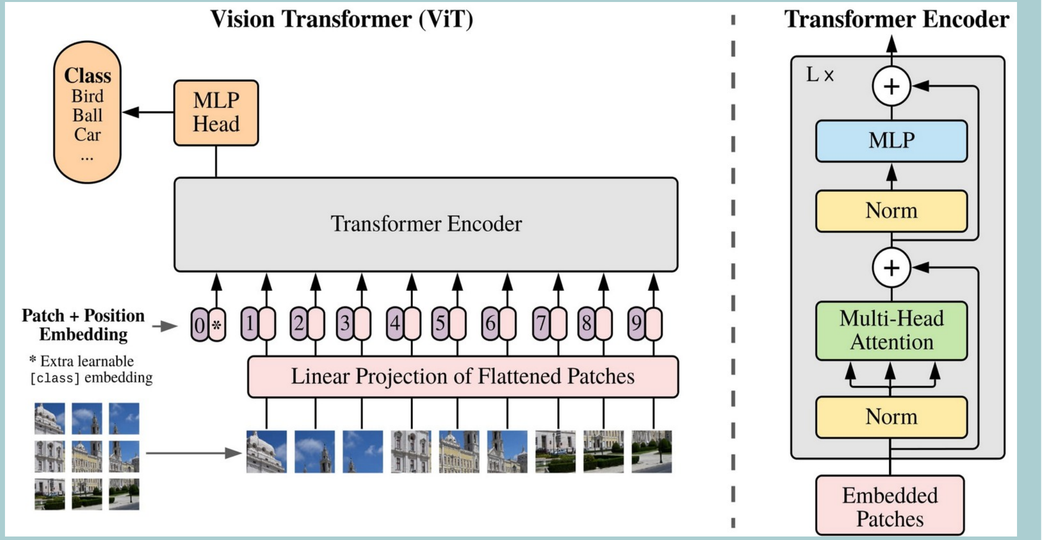
Transformer-based architecture, pre-trained on ImageNet.

Fine-tuned on our dataset.





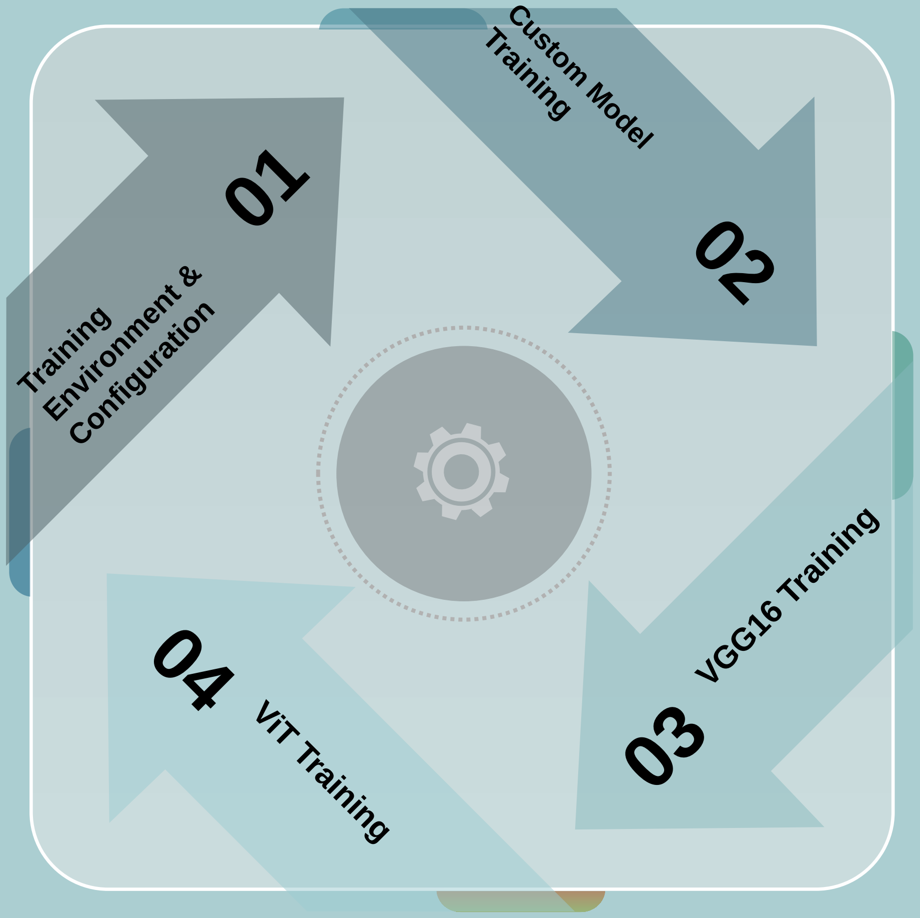




TRAINING PIPELINE

- Platform: Kaggle
- GPU: NVIDIA Tesla P100 (16GB VRAM)
- RAM: 25GB system memory
- Storage: 100GB persistent disk

Parameter	VGG16	ViT-Base	Custom
Batch Size	32	32	32
Initial LR	1e-3	5e-4	1e-4
Epochs	15	15	15
Dropout	0.5	0.1	0.3
Weight Decay	1e-4	1e-2	Not Used



- **7 Conv Layers**
- **1 BatchNorm layer**
- **6.05 M Trainable Parameters**

Optimizer:
- Used Adam optimizer
- Learning rate: 1e-3

Learning Rate Scheduler:
- Used ReduceLROnPlateau to reduce LR on validation loss plateau
- Reduces LR by a factor of 0.2
- Waits for 3 epochs (patience)

Optimizer:
- Used AdamW optimizer
- Learning rate: 1e-3
- Weight decay: 1e-4 (for regularization)
- β values: (0.9, 0.999) for momentum

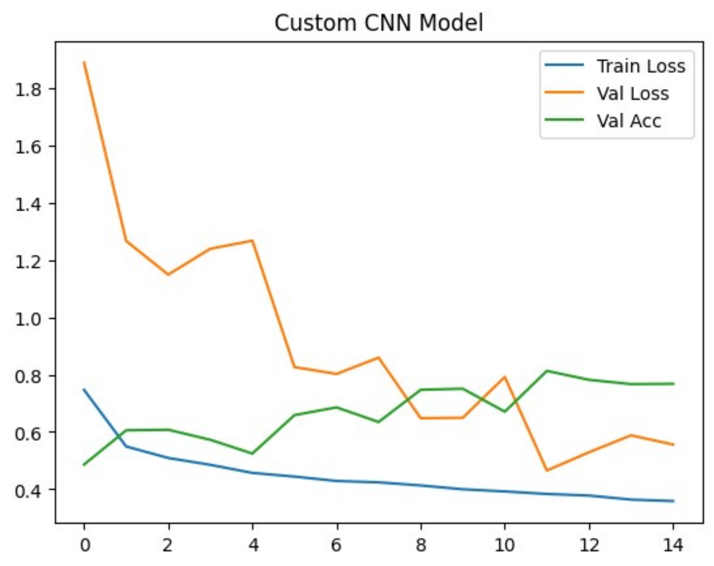
Learning Rate Scheduler:
- Used ReduceLROnPlateau to reduce LR on validation loss plateau
- Reduces LR by a factor of 0.5
- Waits for 5 epochs (patience)
- Minimum LR capped at 1e-6

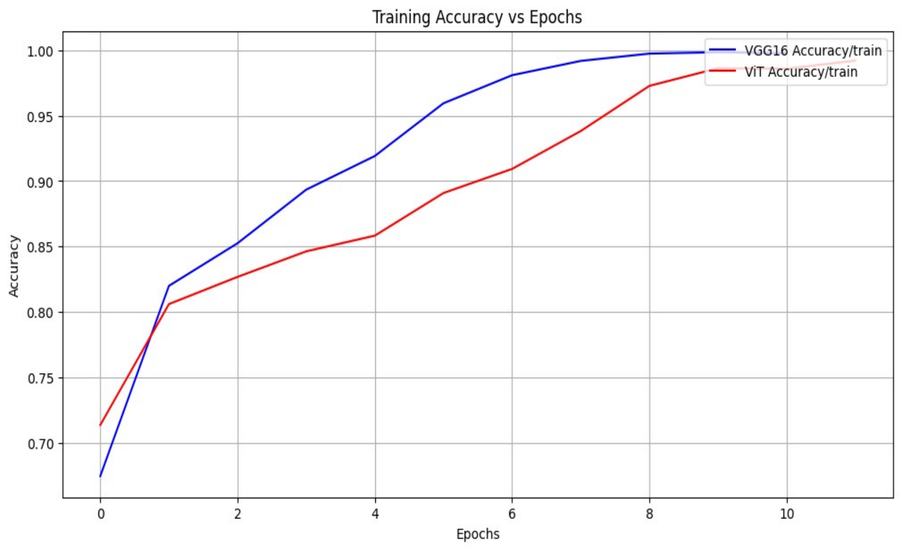
Optimizer:
- Used AdamW
- Learning rate: 5e-4
- Weight decay: 0.01 (to prevent overfitting)

Learning Rate Scheduler:
- Used ReduceLROnPlateau to reduce LR on validation loss plateau
- Reduces LR by a factor of 0.5
- Waits for 5 epochs (patience)
- Minimum LR capped at 1e-6

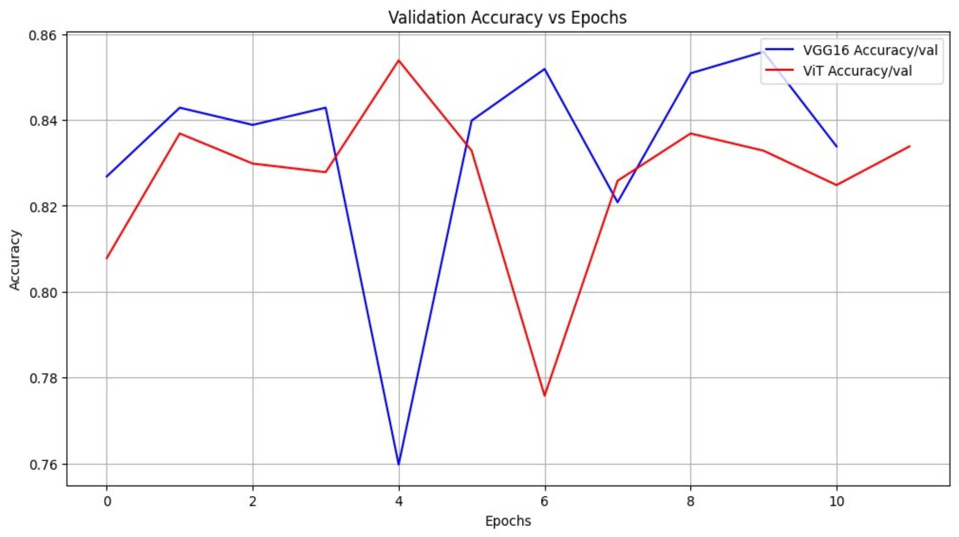
RESULTS: TRAINING PERFORMANCE ANALYSIS

Transfer Learning Models (VGG16 & ViT): Learned very efficiently, achieving high validation accuracy in just a few epochs before early stopping.





From-Scratch Model: Had a much slower and more volatile learning curve, requiring 12 epochs to reach its peak validation accuracy of 81.28%.

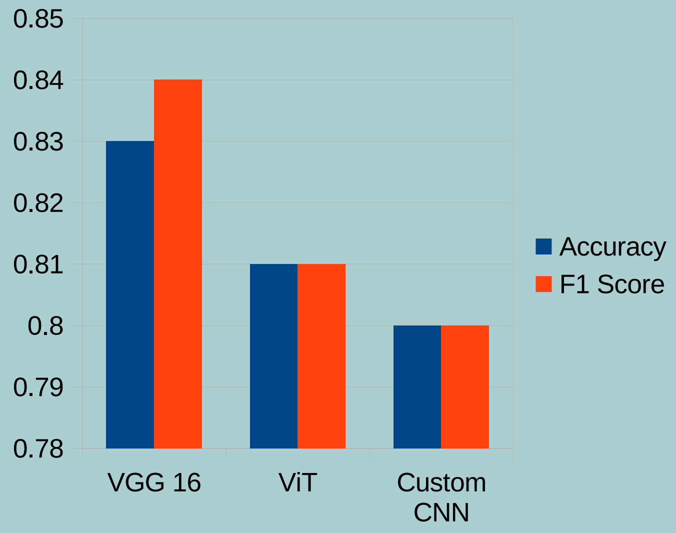


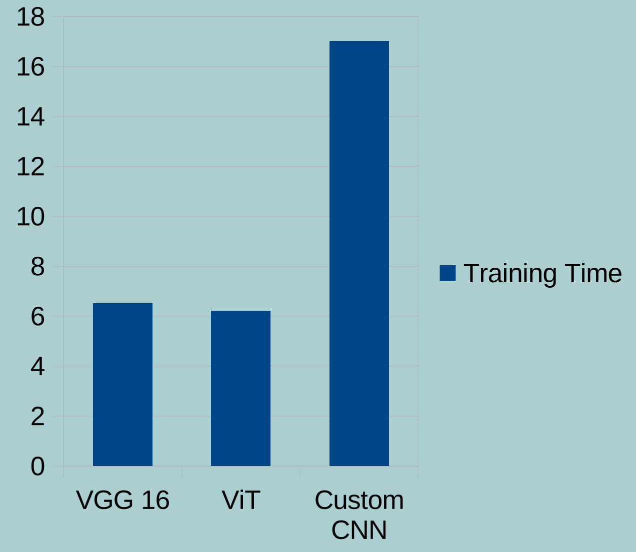
Conclusion: Transfer learning provides a massive advantage in training speed and stability.

RESULTS: OVERALL PERFORMANCE COMPARISON

Main results table:

Model	Training Strategy	Accuracy	F1-Score	Training Time
VGG16	Transfer Learning	83.0%	0.84	~6.5 min
ViT	Transfer Learning	81.0%	0.81	~6.2 min
Custom CNN	From Scratch	80.0%	0.80	17 min



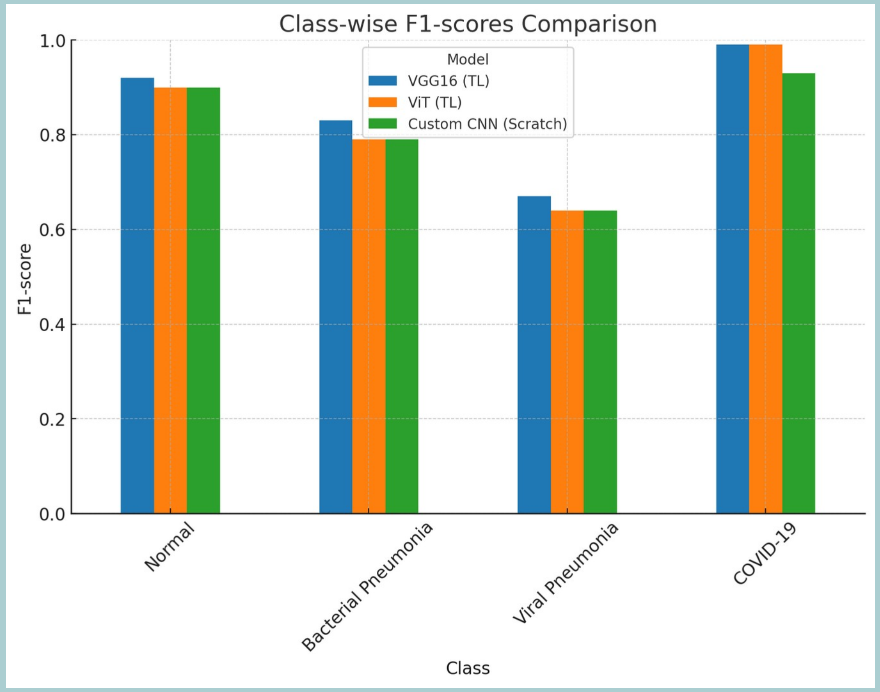


Key Takeaway:
VGG16 with transfer learning was the most accurate model.
The from-scratch model was the least accurate and took over twice as long to train.

RESULTS: CLASS-WISE PERFORMANCE (F1-SCORE)

Class-wise F1-score table:

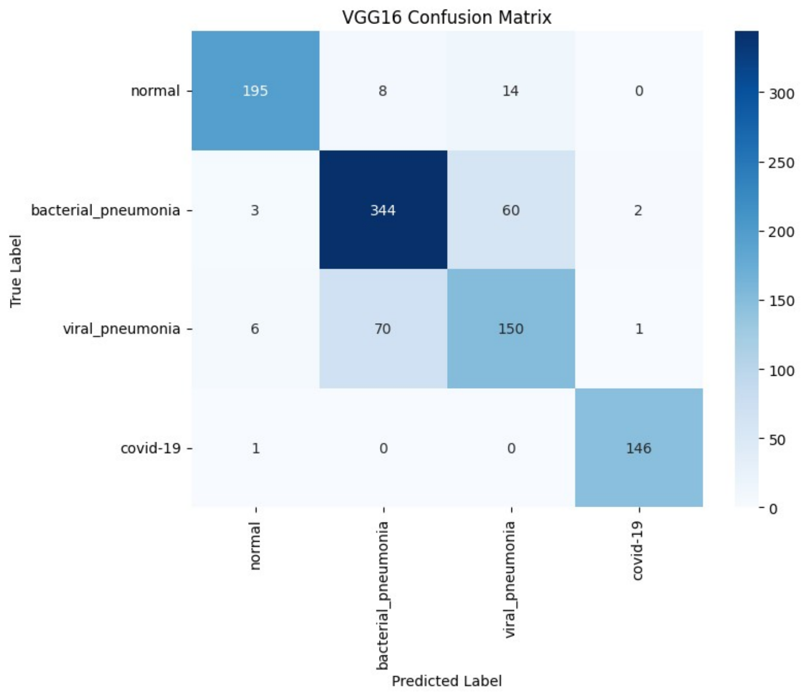
Class	VGG16 (TL)	ViT (TL)	Custom CNN (Scratch)
Normal	0.92	0.90	0.90
Bacterial Pneumonia	0.83	0.79	0.79
Viral Pneumonia	0.67	0.64	0.64
COVID-19	0.99	0.99	0.93



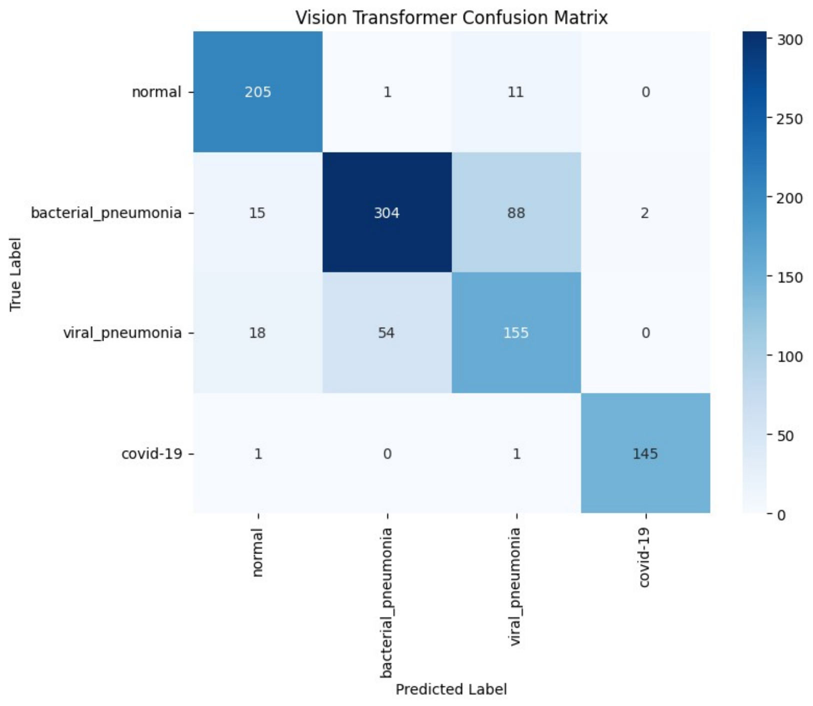
Talking Points:

- *Transfer learning models were nearly perfect at detecting COVID-19.*
- *VGG16 was the most consistent performer across all classes.*
- *All models struggled most with 'Viral Pneumonia'.*

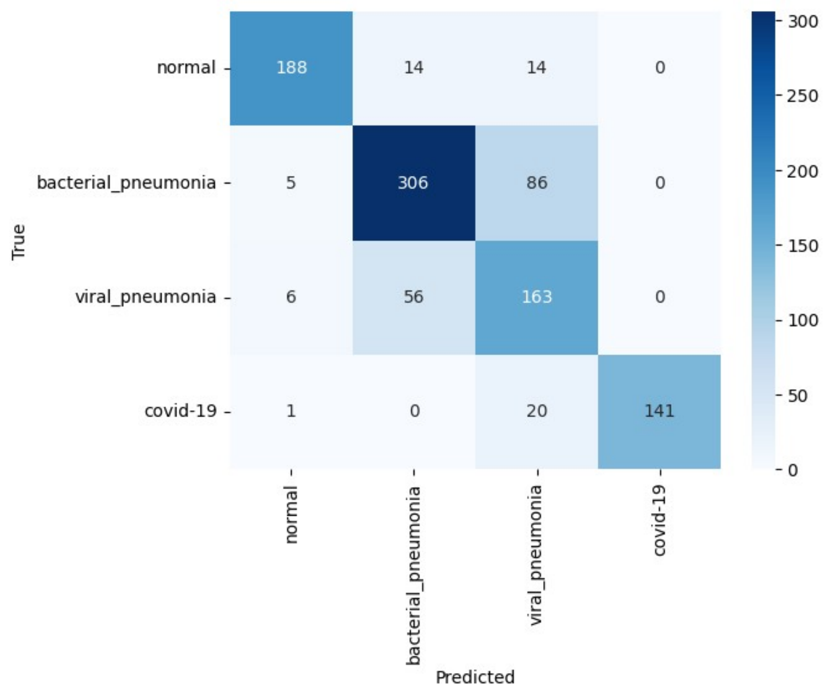
RESULTS: ANALYSIS OF CONFUSION MATRICES



VGG 16



ViT



Custom CNN

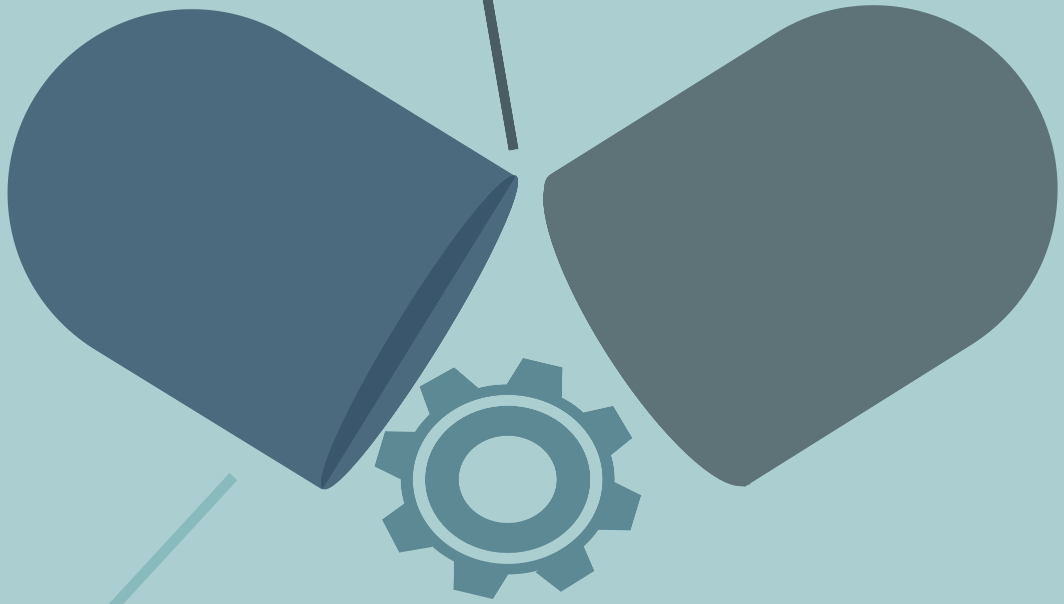
- ***VGG16 & ViT:*** *Showed very strong performance on COVID-19 and Normal cases, with most confusion occurring between the two pneumonia types.*

- ***Custom CNN (From Scratch):*** *Showed major confusion when classifying 'Viral Pneumonia' and missed a significant number of true COVID-19 cases (low recall).*

DISCUSSION 1: THE DECISIVE ADVANTAGE OF TRANSFER LEARNING

The Main Finding: The performance gap between the transfer learning models and the from-scratch model is the most significant result.

Transfer learning models start with a rich "visual vocabulary" from ImageNet and only need to adapt it, leading to higher accuracy and much faster training.



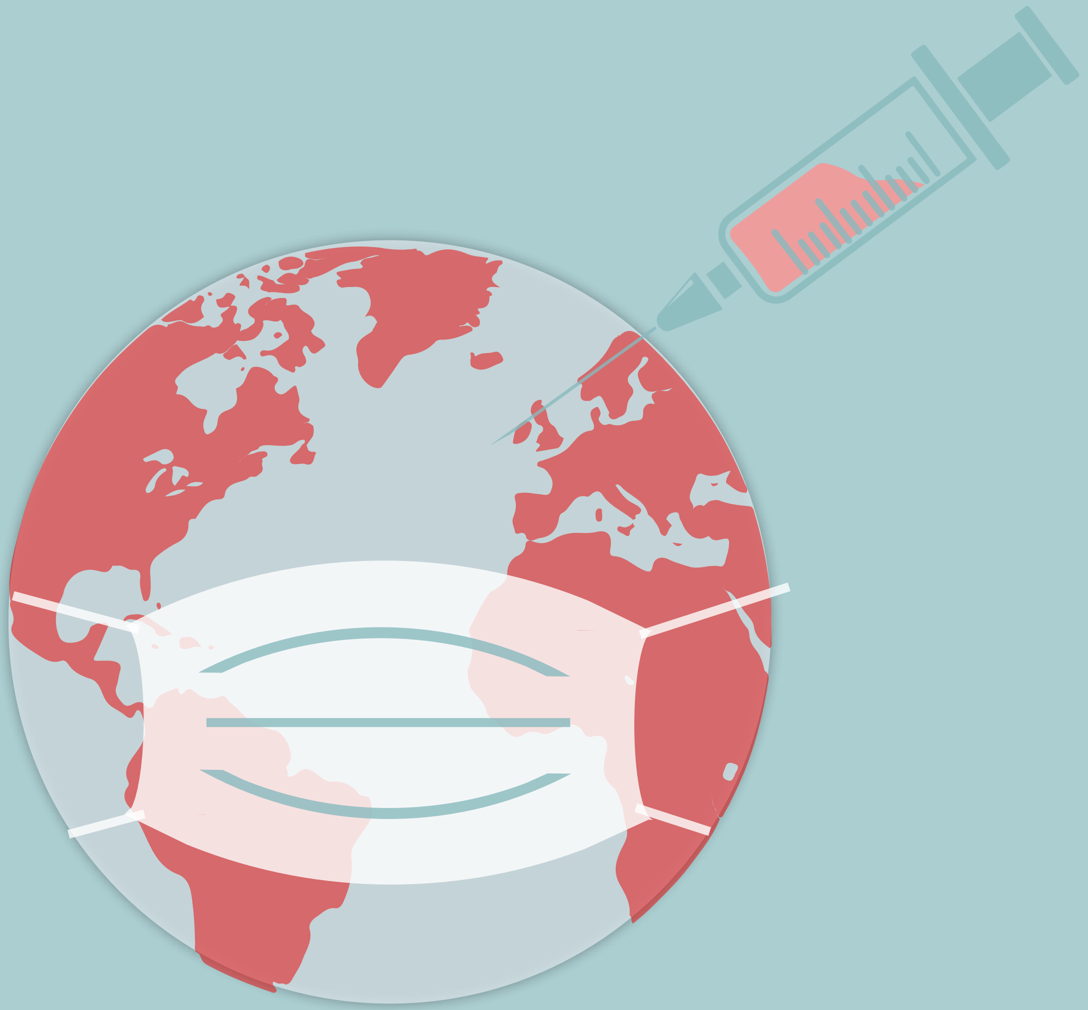
Why? The from-scratch model had to learn basic visual features (edges, textures) from a small dataset, which is inefficient and difficult.

DISCUSSION 2: VGG16 VS. VIT & INDUCTIVE BIAS

- **Why did VGG16 outperform ViT? The concept of inductive bias.**
- **CNNs (VGG16): Have a strong built-in bias for images (locality, translation equivariance). This makes them very data-efficient.**
- **Transformers (ViT): Have a weaker bias and need more data to learn spatial relationships.**
- **Conclusion: On this moderately-sized dataset, the VGG16's strong inductive bias gave it a performance edge.**

CNN kernels implicitly assume neighboring pixels are related and translate features across image positions—an efficient prior not available in Transformers.

LIMITATIONS OF THE STUDY



- **Dataset:** Performance is tied to a single public dataset; results may not generalize to different hospitals or populations.
- **Not a Clinical Tool:** This is a proof-of-concept, not a clinically validated diagnostic tool.
- **Architectural Scope:** The comparison was limited to three specific architectures.
- **Image Data Only:** The models do not use other clinical data (patient history, symptoms, etc.).

FUTURE WORK & RECOMMENDATIONS

- **Validate models on diverse, multi-institutional clinical datasets.**
- **Explore more advanced architectures (e.g., EfficientNet, hybrid models).**
- **Integrate multimodal data (EHR data, patient history) to improve accuracy.**
- **Implement Explainable AI (XAI) to build clinical trust.**
- **Conduct prospective clinical studies to measure real-world impact.**

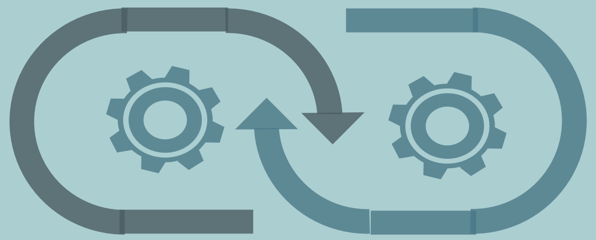
CONCLUSION

1

This research successfully developed and compared three deep learning models for respiratory disease classification.

2

The findings demonstrate that transfer learning is decisively superior to a from-scratch approach in terms of both accuracy and training efficiency.



3

Among the transfer learning models, the VGG16 CNN provided the most robust and balanced performance, likely due to its strong inductive bias.

4

The work validates the use of transfer learning as a powerful strategy for medical imaging and provides a clear benchmark for future research.

REFERENCES

[1] Kim H.E. et al. (2022). *Transfer Learning for Medical Image Classification: A Literature Review*. BMC Medical Imaging.

[2] Salehi M. et al. (2021). *Automated Detection of Pneumonia using Deep Transfer Learning with Pediatric Chest X-rays*. British J. Radiology.

[3] Irvin J. et al. (2019). *CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels*. arXiv:1901.07031.

[4] GiBi13. *Pneumonia & COVID-19 Image Dataset*. Kaggle.
🔗 https://www.kaggle.com/datasets/gibi13/pneumonia-covid19-image-dataset

[5] Raghu M. et al. (2019). *Transfusion: Understanding Transfer Learning for Medical Imaging*. NeurIPS.

[6] Tajbakhsh N. et al. (2016). *CNNs for Medical Image Analysis: Full Training or Fine Tuning?* IEEE Transactions on Medical Imaging.

[7] Dosovitskiy A. et al. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. ICLR.

[8] Khan S.H. et al. (2022). *Transformers in Vision: A Survey*. ACM Computing Surveys.

[9] Wang X. et al. (2017). *ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks*. CVPR.

THANK YOU!