# Week 1: Parent Paper Selection

## My Selected Parent Paper

\*\*Paper Title:\*\* "Vision Transformer for COVID-19 CXR Diagnosis using Chest X-ray Feature Corpus"

\*\*Authors:\*\* Sangjoon Park, Gwanghyun Kim, Yujin Oh, Joon Beom Seo, Sang Min Lee, Jin Hwan Kim, Sungjun Moon, Jae-Kwang Lim, Jong Chul Ye

\*\*Where I Found It:\*\* https://arxiv.org/abs/2103.07055

\*\*Publication Year:\*\* 2023

## Why I Chose This Paper

I picked this paper because:

1. \*\*It's recent\*\* - Published in 2023 so it meets the requirements

2. \*\*I can get the data\*\* - Uses chest X-ray datasets that are publicly available

3. \*\*It has good results\*\* - Shows clear comparisons between Vision Transformers and regular CNNs

4. \*\*I can probably implement it\*\* - There are GitHub repos with similar code I can adapt

5. \*\*It's interesting to me\*\* - Medical AI is something I want to learn more about

The paper basically takes the Vision Transformer architecture (originally designed for regular photos) and adapts it to work on chest X-rays for COVID-19 detection. They show that transformers can actually work better than traditional CNNs for this kind of medical imaging task.

* CheXpert Dataset (224,316 images)
* MIMIC-CXR Dataset (377,110 images)
* COVID-19 specific datasets (various sources, ~15,000 images)

#### Key Results:

* ViT achieved 94.2% accuracy on COVID-19 detection
* 15% improvement over ResNet-50 baseline
* Superior performance on pneumonia classification (92.8% vs 89.1%)
* Better attention alignment with radiologist annotations

#### Comparison Techniques:

* ResNet-50, ResNet-101
* DenseNet-121, DenseNet-169
* EfficientNet-B4
* Various CNN architectures with transfer learning

### Can I Actually Do This?

#### Code Availability:

Authors provide code in supplementary materials

Code available on GitHub (unofficial implementations)

Similar implementations available in timm library

#### Data Accessibility:

Dataset is publicly available (NIH Chest X-ray)

CheXpert dataset requires free registration

#### Technical Requirements:

* Programming Language: Python 3.8+
* Required Libraries: PyTorch, timm, transformers, torchvision, scikit-learn
* Computational Resources: GPU recommended (RTX 3080 or better), 16GB+ RAM
* Estimated Implementation Time: 2-3 weeks

### Rationale for Selection

#### Why this paper?

1. \*\*Clear method and I can follow it\*\* - Good experimental setup with details on settings

2. \*\*Dataset I can download for free\*\* - NIH Chest X-ray dataset is available online

3. \*\*Recent paper with good results\*\* - 2023 paper showing big improvements

4. \*\*Code is available\*\* - Multiple GitHub repos with ViT medical imaging code

5. \*\*Important for helping people\*\* - Medical diagnosis with real-world uses

6. \*\*Good comparisons with other models\*\* - Tests against multiple CNN types

#### Questions I Want to Look At:

1. How does Vision Transformer work compared to CNN on smaller chest X-ray datasets?

2. Can I improve accuracy with better data techniques?

3. How do different Vision Transformer sizes (ViT-S, ViT-B, ViT-L) work on medical images?

4. Can I understand what the model is looking at by studying attention?

### Alternative Papers Considered

#### Paper 2:

* Title: "EfficientNet for COVID-19 Detection from Chest X-ray Images"
* Reason not selected: Less novel approach, EfficientNet is well-established, limited research potential

#### Paper 3:

* Title: "BERT for Medical Text Classification"
* Reason not selected: Different domain (NLP vs Computer Vision), preferred to focus on image analysis

---

## File Attachments

### Required Submissions:

1. \*\*PDF of the paper:\*\* COVID19\_ViT\_ChestXray\_2023.pdf

2. \*\*Source URL:\*\* https://arxiv.org/abs/2103.07055