COMP 4360 FULL PROJECT PROPOSAL

Group 3: Aamir Sangey, Manmilan Singh, Peter Vu

Problem Domain

- Goal: Adapt SimMIM to medical imaging
- Task: Learn strong visual representations from unlabeled chest X-rays
- Motivation:
  - Medical labels are expensive and noisy
  - X-ray interpretation requires fine-grained spatial reasoning
- Downstream use: Disease classification (e.g., cardiomegaly, edema)

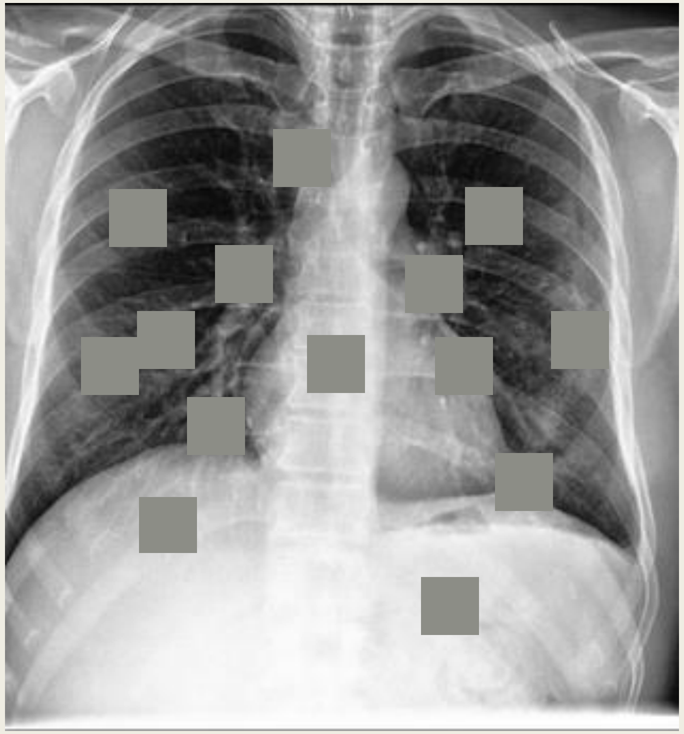


Fig. 1. Conceptual illustration of SimMIM-style 32 x 32 patch [1].

Chexpert Dataset Overview

- Dataset: Chexpert (Stanford)
- Scale:
  - 224,316 chest radiographs
  - 65,240 patients
- Image format:
  - Original DICOM (~439 GB)
  - Commonly down-sampled to **~320x320 PNG/JPG** (**~ 11 GB**)
- Imaging type: Frontal & lateral chest X-rays (grayscale)

Labels & Domain Shift

- 14 clinical observations, including:
    - Atelectasis, Cardiomegaly, Edema, Pneumonia
- Label types:
    - **Positive (1)**
    - **Negative (0)**
    - **Uncertain (-1)**

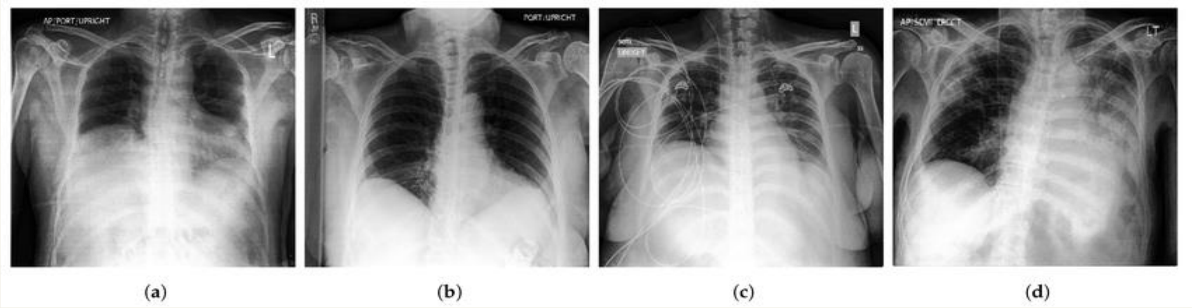


Fig. 2. Samples of CXR images from CheXpert dataset [44] where, (a) Atelectasis; (b) Cardiomegaly; (c) Edema; (d) Pneumonia [2].

- Explicit domain shift:
    - SimMIM pretraining used ImageNet
    - Chexpert contains **medical X-rays**, not natural images

Source Paper

- Paper: SimMIM: A Simple Framework for Masked Image Modeling (CVPR 2022)
- The paper proposes a *simple* masked image modeling (MIM) framework that avoids complex tokenizers/decoders used by prior methods.
- Randomly mask image patches → encode with a vision transformer → **predict raw pixel values** for the masked regions using a **lightweight (linear) head** and a simple regression loss.

Methodology

- SimMIM frames masked image modeling as 4 major components:

1. **Hide parts of the image (Masking).**
2. **Learn from the visible patches (Encoder).**
3. **Predict the missing pixels (Head).**
4. Train with **L1 reconstruction loss** (**computed only on masked patches**).

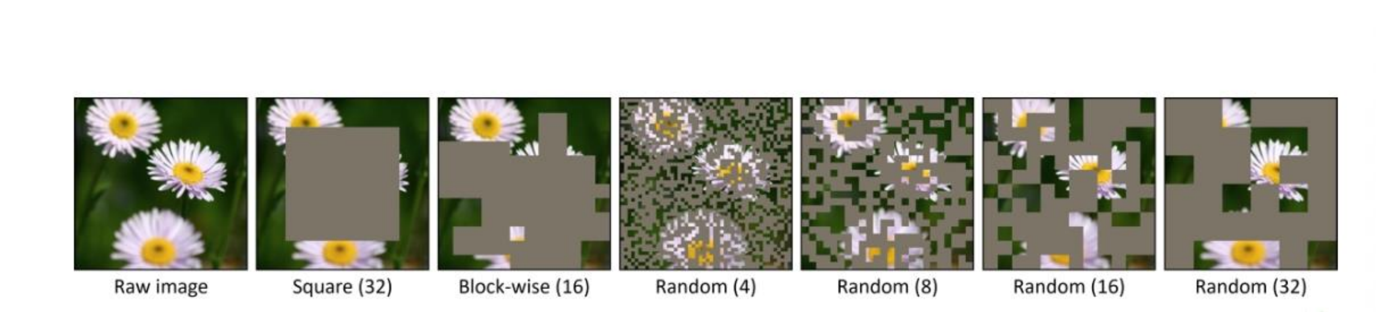


Fig 3: Example of image using different patch sizes [4]

Architecture Diagram

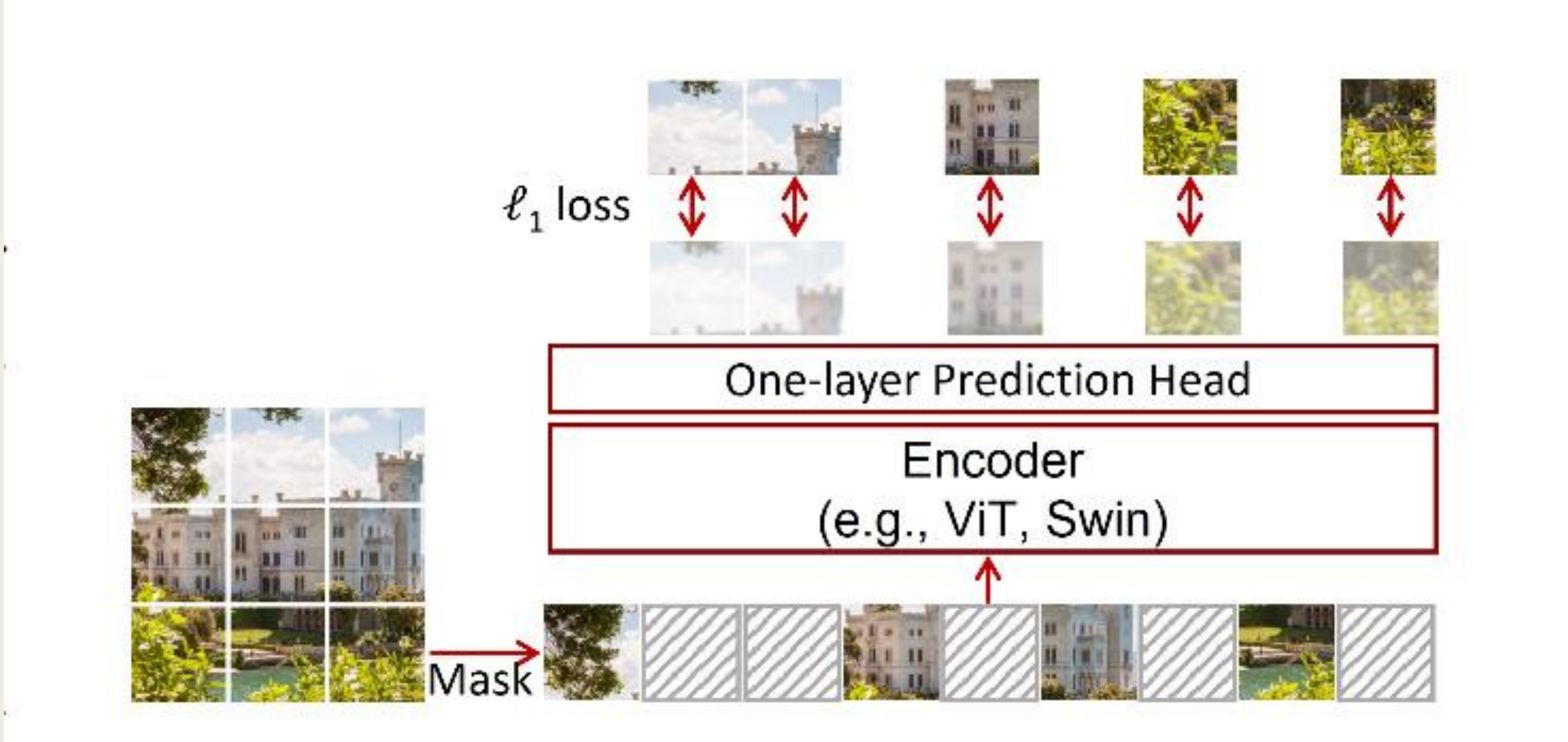


Fig 4: SimMIM pretraining architecture [4]

Adaptation Strategy

- **Different input type:** SimMIM is commonly run on RGB natural images, but CheXpert images are chest radiographs (medical X-rays).
- **Different learning goal:** Pre-train with masked reconstruction, then fine-tune for 14 clinical observations (multi-label).
- **Core adaptation challenge:** X-rays are less diverse than ImageNet, and pathology can be localized, while pre-training ignores labels.

Model Changes

- **Input channels (RGB ⟶ grayscale):** patch embedding input 3 ⟶ 1 channel (X-ray).
- **Reconstruction head output** for masked-patch prediction.
- **Data loader:** adapt to CheXpert image paths + 14-label CSV format.
- **Fine-tuning:** replace head with 14-output classifier and define uncertainty handling.

Fig 5: ImageNet sample [5]

Fig 6: CheXpert sample [1]

Development Plan

- **Goal:** Clone (and adapt) the official repository:
  - https://github.com/microsoft/SimMIM

Development Plan

Environment

- **Environment** (same as paper): Conda + pip
  - Python 3.8
  - CUDA 11.3 + cuDNN 8
  - PyTorch, PyTorch Image Models (timm), Apex, SciPy, PyYAML, YACS, Termcolor
- **Hardware Resources**:
  - **Primary**: Aviary (or other departmental resources)
  - **Backup**: Google Colab + Google Drive
    - Drive for codebase, datasets, and model weights, Colab for compute

Development Plan

Ablation Plan

- The original paper provides a detailed **Ablation Study**, which we will replicate.
- Ablation will be done only for **pre-training**, keeping downstream fine-tuning identical.

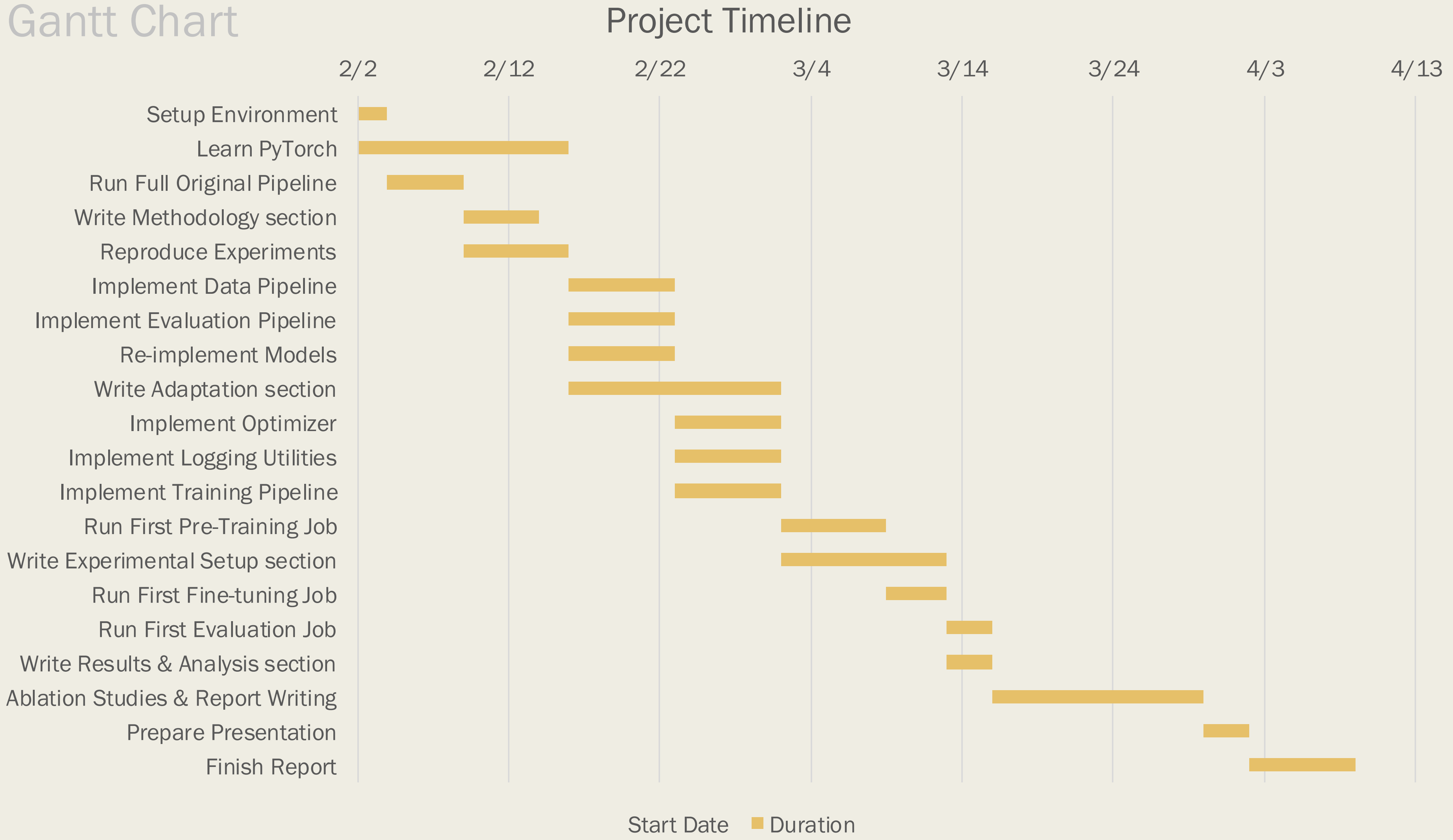
1. **Masking Strategy**: mask type, masked patch size, mask ratio,…
2. **Prediction Head**: Linear => 2-layer MLP
3. **Prediction Target** (Loss Function): L1 => L2

Project Management

- **Milestones:**
  - **Milestone 1 (2/2 - 2/15):** Understand + Run the Official Implementation
  - **Milestone 2 (2/16 - 3/1):** Implement Fully the Adapted Model
  - **Milestone 3 (3/2 - 3/15):** Run Experiments and Evaluate Implementation
  - **Milestone 4 (3/16 - 3/29):** Run Ablation Studies
- Weekly meetings to check in on progress
- Final Report will be populated in parallel to development and experimentation

Project Management

Gantt Chart



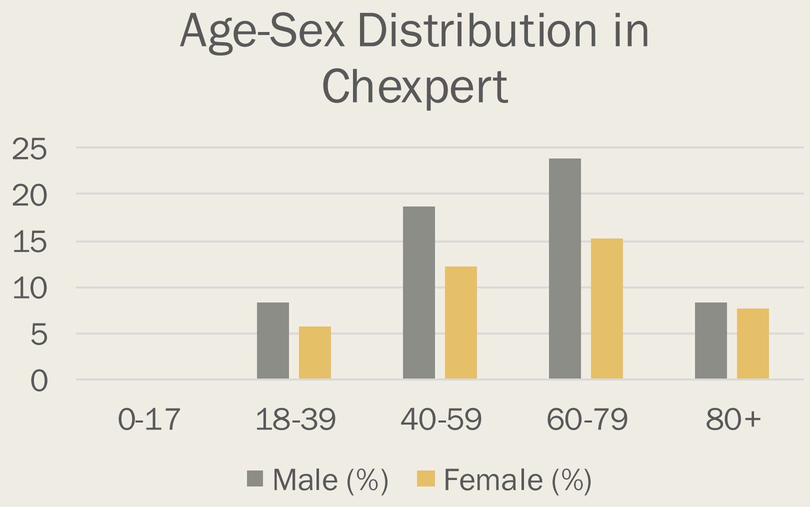
Role Division

- Peter: Project Manager + DevOps + Data Pipeline + Report Lead
- Aamir: Training Pipeline + Pre-training Process Lead + Model Expert
- Manmilan: Evaluation Pipeline + Fine-tuning Process Lead + Experimentation Lead

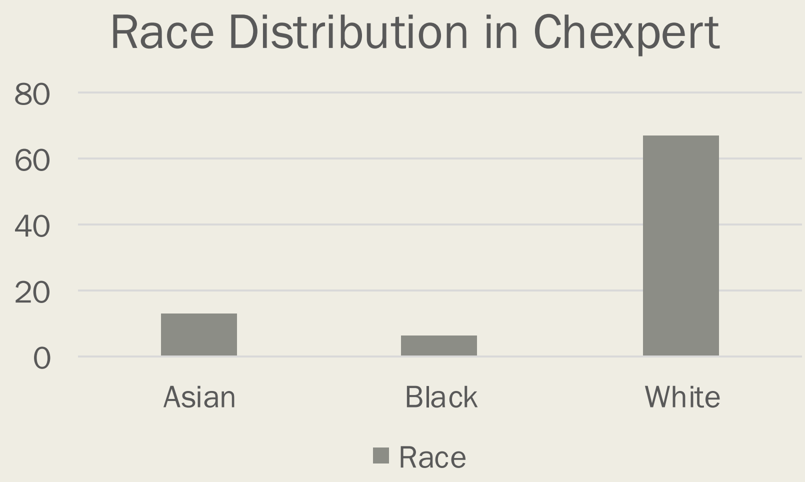
EDI Considerations

: Equity, Diversity, & Inclusion

- The Chexpert dataset originates from **a single U.S. hospital**, which may limit representation of patients from different healthcare systems or ethnic groups.
- Age, sex, and race are **not uniformly represented** (see charts), which may affect model generalizations across demographic groups.
- These limitations are acknowledged, and robustness is evaluated through ablation and sensitivity analysis.



[3]



[6]

Distributions reported in prior analyses of Chexpert Dataset (cited)

References

1. [1] "CheXpert: A Large Dataset of Chest X-Rays and Competition for Automated Chest X-Ray Interpretation.," stanfordmlgroup.github.io. https://stanfordmlgroup.github.io/competitions/chexpert/
2. [2] A. Ait Nasser and M. Akhloufi, "A review of recent advances in deep learning models for chest disease detection using radiography," Diagnostics, vol. 13, no. 1, p. 159, Jan. 2023, doi: 10.3390/diagnostics13010159.
3. [3] A. Badawy, M. Elhariry, A. Chirrimar, and A. Chohan, "Apples-to-Apples: Age-Sex Standardisation of Public Chest X-ray Datasets," Cureus, Nov. 2025, doi: https://doi.org/10.7759/cureus.97260.
4. [4] Z. Xie *et al.*, "SimMIM: a Simple Framework for Masked Image Modeling," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, doi: https://doi.org/10.1109/cvpr52688.2022.00943.
5. [5] EliSchwartz, "GitHub - EliSchwartz/imagenet-sample-images: 1000 images, one per image-net class. For easy visualization/exploration of classes.," *GitHub*, 2019. https://github.com/EliSchwartz/imagenet-sample-images
6. [6] I. Banerjee et al., "Reading Race: AI Recognises Patient's Racial Identity In Medical Images," arxiv.org, Jul. 2021, Available: https://arxiv.org/abs/2107.10356