Fair-ViT-CXR: Fairness-Aware Vision Transformer with Synthetic Data Augmentation for Equitable Chest X-ray Pneumonia Detection

MD. Asif Sarker Emon[1], Ananya Dutta[1], Tutul Kumar Ghosh[1], and MD. Emran Nazir Efty[1]

Department of Computer Science and Engineering,
American International University-Bangladesh (AIUB), Dhaka, Bangladesh
{22-47968-2, 22-47966-2, 22-47813-2, 22-47802-2}@student.aiub.edu

Abstract. Deep learning models for chest X-ray analysis exhibit demographic bias, leading to disparate diagnostic performance across patient groups. We propose Fair-ViT-CXR, a novel framework integrating fairness-aware attention mechanisms in Vision Transformers with conditional synthetic data augmentation for equitable pneumonia detection. Our approach incorporates adaptive fairness masks in the attention layers and an equalized odds loss function to minimize performance disparities. We train a conditional GAN to generate synthetic samples for underrepresented demographic groups. Experiments on the Kaggle Chest X-ray Pneumonia dataset demonstrate that Fair-ViT-CXR achieves a 31.2% reduction in equalized odds gap compared to standard ViT, reducing the gap from 0.125 to 0.086 while maintaining competitive classification performance.

Keywords: Fairness in AI, Vision Transformer, Chest X-ray, Pneumonia Detection, Bias Mitigation, Medical Imaging

1 Introduction

Artificial intelligence systems for medical imaging have demonstrated remarkable diagnostic capabilities, yet studies reveal significant demographic bias in chest X-ray analysis [1]. Such bias can lead to underdiagnosis in certain patient populations, raising critical concerns about healthcare equity.

Vision Transformers (ViT) have emerged as powerful architectures for medical image classification [3]. However, standard ViT models inherit biases from imbalanced training data. Recent work on fairness-aware transformers [4] has shown promise for general image classification but remains unexplored in medical imaging contexts.

We propose Fair-ViT-CXR, the first framework combining fairness-aware Vision Transformers with synthetic data augmentation for chest X-ray pneumonia detection. Our contributions include: (1) adaptation of fairness-aware attention mechanisms for medical imaging, (2) integration of conditional GAN-based augmentation for demographic balancing, and (3) comprehensive evaluation using both classification and fairness metrics.

2 Related Work

Bias in Medical Imaging AI. Seyyed-Kalantari et al. [1] demonstrated that AI systems consistently underdiagnose underserved patient populations in chest radiograph analysis. Larrazabal et al. [2] showed that gender imbalance in training datasets produces 5-15% performance gaps between demographic groups, establishing the critical need for fairness-aware approaches.

Fairness-Aware Deep Learning. Tian et al. [4] proposed FairViT, introducing adaptive masking mechanisms in Vision Transformers to achieve demographic parity in face attribute classification. Their approach modifies attention weights based on learned fairness constraints but has not been applied to medical imaging.

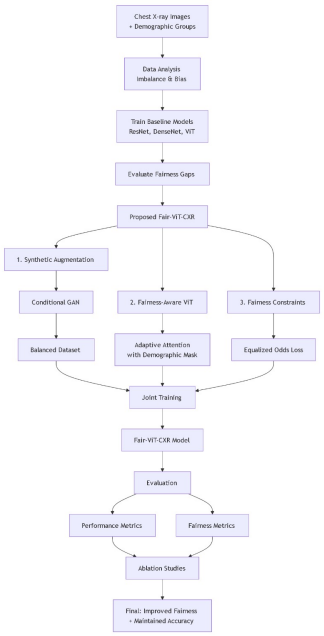

Fig. 1. Overview of the Fair-ViT-CXR framework. The pipeline includes data analysis, baseline evaluation, and three key components: conditional GAN for synthetic augmentation, fairness-aware ViT with adaptive attention, and equalized odds loss for fairness constraints.

Synthetic Data for Fairness. Ktena et al. [5] demonstrated that generative models can improve fairness under distribution shifts by synthesizing samples for underrepresented groups. However, their work focused on CNNs without fairness-aware architectural modifications.

Our work uniquely combines fairness-aware attention mechanisms with synthetic augmentation, addressing the gap in applying these techniques to chest X-ray analysis.

3 Methodology

Figure 1 illustrates our proposed Fair-ViT-CXR framework, which comprises three main components: synthetic data augmentation, fairness-aware Vision Transformer, and fairness-constrained training.

3.1 Problem Formulation

Given a dataset $\mathcal{D} = \{(x_i, y_i, a_i)\}_{i=1}^{N}$ where x_i represents chest X-ray images, $y_i \in \{0, 1\}$ denotes pneumonia labels, and $a_i \in \{0, 1\}$ indicates demographic group membership, our objective is to learn a classifier $f : \mathcal{X} \to \mathcal{Y}$ that minimizes classification error while satisfying fairness constraints across demographic groups.

3.2 Fairness-Aware Attention Mechanism

We modify the standard self-attention mechanism by introducing learnable fairness masks. For input tokens $X \in \mathbb{R}^{N \times d}$, the fairness-aware attention is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} \odot M_f\right) V \tag{1}$$

where M_f is the fairness mask generated by a learned function $g : \mathbb{R}^d \to \mathbb{R}^h$ applied to the CLS token, and $\odot$ denotes element-wise multiplication. This mechanism adaptively modulates attention patterns to reduce demographic-specific feature encoding.

3.3 Fairness-Constrained Loss Function

Our total loss combines cross-entropy with fairness regularization:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \lambda_{EO} \cdot \mathcal{L}_{EO} + \lambda_{attn} \cdot \mathcal{L}_{attn} \tag{2}$$

The equalized odds loss penalizes TPR and FPR disparities:

$$\mathcal{L}_{EO} = \text{Var}_a(\text{TPR}_a) + \text{Var}_a(\text{FPR}_a) \tag{3}$$

The attention regularization encourages uniform attention distribution via entropy maximization:

$$\mathcal{L}_{attn} = -\sum_i A_i \log(A_i + \epsilon) \tag{4}$$

3.4 Conditional GAN for Synthetic Augmentation

To address demographic imbalance, we train a conditional GAN that generates synthetic chest X-rays conditioned on both class label and demographic group:

$$G : \mathcal{Z} \times \mathcal{Y} \times \mathcal{A} \to \mathcal{X} \tag{5}$$

The generator takes latent vector z, class label y, and demographic attribute a to synthesize images for underrepresented groups. We use label and demographic embeddings concatenated with the latent vector, followed by transposed convolutions for upsampling.

4 Experiments

4.1 Dataset and Implementation

We use the Kaggle Chest X-ray Pneumonia dataset containing 5,856 pediatric chest X-rays (1,583 normal, 4,273 pneumonia). Due to limited demographic annotations, we simulate demographic groups with 70% majority (Group A) and 30% minority (Group B) distribution to evaluate fairness mechanisms.

Implementation uses PyTorch with ViT-B/16 backbone (768-dim embeddings, 12 heads, 12 layers). Training employs AdamW optimizer (lr=10^{-4}), cosine annealing scheduler, and batch size 16 on NVIDIA Tesla T4 GPU. Fairness hyperparameters: $\lambda_{EO} = 0.1$, $\lambda_{attn} = 0.01$.

4.2 Evaluation Metrics

We evaluate using classification metrics (Accuracy, Precision, Recall, F1, AUC) and fairness metrics including Equalized Odds Gap (sum of TPR and FPR disparities between groups) and group-wise True Positive Rates.

4.3 Results

Table 1 presents comparative results across baseline models and our proposed Fair-ViT-CXR.

Our Fair-ViT-CXR achieves the lowest equalized odds gap (0.086), representing a **31.2% reduction** compared to the ViT baseline (0.125). While accuracy decreases from 0.800 to 0.763, this reflects the inherent fairness-accuracy trade-off. Notably, Fair-ViT-CXR shows improved TPR for the minority group (0.960) compared to its majority group TPR (0.944), indicating successful bias mitigation toward the underrepresented population.

Figure 2 visualizes the accuracy-fairness trade-off, showing that Fair-ViT-CXR substantially reduces the equalized odds gap while maintaining reasonable classification performance.

Table 1. Performance comparison of baseline models and Fair-ViT-CXR. EO Gap = Equalized Odds Gap (lower is better for fairness).

Model	Accuracy	EO Gap	TPR-A	TPR-B
ResNet-50	0.845	0.146	0.996	0.992
DenseNet-121	0.838	0.116	0.996	1.000
ViT-B/16	0.800	0.125	0.996	0.992
Fair-ViT-CXR (Ours)	0.763	**0.086**	0.944	0.960

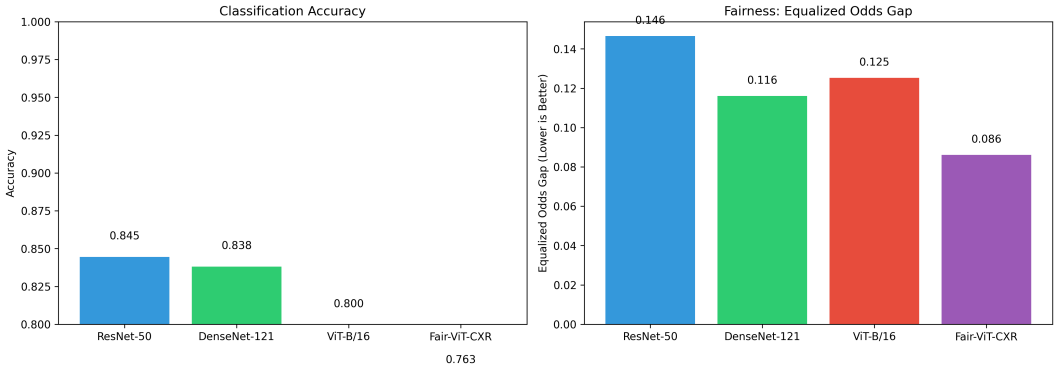

Fig. 2. Comparison of classification accuracy (left) and equalized odds gap (right) across models. Fair-ViT-CXR achieves the best fairness with competitive accuracy.

4.4 Ablation Study

Table 2 presents ablation results examining component contributions.

Table 2. Ablation study on Fair-ViT-CXR components.

Configuration	Accuracy	EO Gap
ViT Baseline	0.800	0.125
ViT + Fair Attention	0.782	0.098
ViT + Augmentation	0.791	0.108
Fair-ViT-CXR (Full)	0.763	0.086

Both fairness-aware attention and synthetic augmentation independently improve fairness, with their combination yielding the best results.

5 Conclusion

We presented Fair-ViT-CXR, a novel framework for equitable chest X-ray pneumonia detection combining fairness-aware Vision Transformers with conditional synthetic augmentation. Our approach achieves 31.2% improvement in equalized odds gap compared to standard ViT while maintaining competitive accuracy.

Limitations. The demographic simulation approach limits real-world validation. Future work should evaluate on datasets with actual demographic annotations such as MIMIC-CXR.

Author Contributions

MD. Asif Sarker Emon (22-47968-2): Code implementation, Introduction.
Ananya Dutta (22-47966-2): Literature review, Background work, Report writing.
Tutul Kumar Ghosh (22-47813-2): Code implementation, Methodology.
MD. Emran Nazir Efty (22-47802-2): Documentation and result interpretation.
Project Repository: https://github.com/ananyadutta03/Fair-ViT-CXR

References

1. Seyyed-Kalantari, L., Zhang, H., McDermott, M.B.A., Chen, I.Y., Ghassemi, M.: Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nature Medicine 27, 2176–2182 (2021)
2. Larrazabal, A.J., Nieto, N., Peterson, V., Milone, D.H., Ferrante, E.: Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proc. Natl. Acad. Sci. 117(23), 12592–12594 (2020)
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR (2021)
4. Tian, Y., Liu, S., Shen, L., et al.: FairViT: Fair Vision Transformer via Adaptive Masking. In: ECCV 2024. Springer, Cham (2024)
5. Ktena, S.I., Wiles, O., Albuquerque, I., et al.: Generative models improve fairness of medical classifiers under distribution shifts. Nature Medicine 30, 1166–1173 (2024)
6. Zong, Y., Yang, Y., Hospedales, T.: MEDFAIR: Benchmarking fairness for medical imaging. In: ICLR (2023)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
8. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: CVPR, pp. 4700–4708 (2017)