Incorporating Rather Than Eliminating: Achieving Fairness for Skin Disease Diagnosis Through Group-Specific Experts

Gelei Xu[1], Yuying Duan[1], Zheyuan Liu[1], Xueyang Li[1], Meng Jiang[1], Michael Lemmon[1], Wei Jin[2], and Yiyu Shi[1]

[1] University of Notre Dame, Notre Dame, IN 46556, USA
{gxu4, yduan2, zliu29, xli34, mjiang2, lemmon, yshi4}@nd.edu
[2] Emory University, Atlanta, GA 30322, USA
wei.jin@emory.edu

Abstract. AI-based systems have achieved high accuracy in skin disease diagnostics but often exhibit biases across demographic groups, leading to inequitable healthcare outcomes and diminished patient trust. Most existing bias mitigation methods attempt to eliminate the correlation between sensitive attributes and diagnostic prediction, but those methods often degrade performance due to the lost of clinically relevant diagnostic cues. In this work, we propose an alternative approach that incorporates sensitive attributes to achieve fairness. We introduce FairMoE, a framework that employs layer-wise mixture-of-experts modules to serve as group-specific learners. Unlike traditional methods that rigidly assign data based on group labels, FairMoE dynamically routes data to the most suitable expert, making it particularly effective for handling cases near group boundaries. Experimental results show that, unlike previous fairness approaches that reduce performance, FairMoE achieves substantial accuracy improvements while preserving comparable fairness metrics.

Keywords: Skin Disease · Fairness · Mixture-of-Experts.

1 Introduction

Recently, AI-based systems have been widely adopted for skin disease diagnostics, achieving high accuracy in predicting various conditions [1,15,27,32]. However, these systems often exhibit biases across population groups. For example, a model may perform well overall but yield inconsistent diagnostic accuracy between patients with dark and light skin types [11]. In clinical settings, such biases can lead to inequitable healthcare outcomes and erode patient trust in AI-driven diagnostics, highlighting the urgent need for fairer diagnostic methods.

The most common bias mitigation strategies focus on **removing** the influence of sensitive attributes (e.g. age, gender). Adversarial training, for example, achieves fairness by training an adversary network to predict sensitive attributes from learned representations while optimizing the main model to minimize this

adversary's success, ensuring that sensitive information is not encoded [2, 30]. Regularization-based methods learn fair model by penalizing correlations between sensitive attributes and model's output [14, 19]. More recently, pruning and quantization have been used to improve fairness by removing model components that contribute most to disparities across sensitive attributes [5, 12, 26].

While these approaches promote fairness by suppressing sensitive attributes, they may not be ideal for skin disease diagnosis, where such attributes hold essential clinical value. For example, skin type encodes diagnostic signals such as UV susceptibility, which is crucial for clinical decision-making [4]. Suppressing these attributes can inadvertently discard vital information, leading to significant declines in model performance [23,29]. More concerningly, reducing performance gaps between groups is often achieved not by improving disadvantaged groups but by uniformly degrading performance across all groups [7, 8]. Thus, instead of removing sensitive attributes, this paper advocates for **incorporating** group information to improve fairness. This reflects a shift from suppressing differences to embracing them through *group customization*—using sensitive attributes to deliver predictions tailored to each group's clinical profile. Given the differences in disease distribution across groups, assigning each group its own module allows models to better capture specialized patterns and improve diagnostic accuracy.

To achieve this, we draw inspiration from recent advances in Mixture of Experts (MoE), a framework that enables the training of specialized experts with dynamic input routing [20, 31]. MoE provides a promising solution for group-aware learning by allowing models to leverage different experts based on input characteristics. However, directly applying MoE poses several challenges. **First**, MoE routers automatically learn expert selection, making it difficult to explicitly assign experts to specific groups. **Second**, balancing specialization and generalization is non-trivial. Training an expert solely on its group's data strengthens specialization but weakens generalization due to data scarcity, while incorporating other groups' data improves generalization but introduces distribution shifts, reducing specialization. **Third**, handling continuous attributes (e.g., age) is challenging, especially for intermediate samples that lack clear group assignment. If each expert is optimized for a specific group, borderline cases may experience degraded performance due to ineffective expert coverage. Addressing these challenges is crucial for using MoE in fairness-aware medical AI systems.

Building on these insights, we introduce FairMoE, a framework that employs layer-wise mixture-of-experts modules to serve as group-specific learners for skin disease diagnosis. FairMoE leverages mutual information loss to establish strong correlation between groups and experts, ensuring that each expert specializes in a specific group. To maximize performance across all groups, particularly for data samples that fall around group boundaries, each expert is trained not only on its designated group's subset but also selectively incorporates data from other groups through soft probabilistic routing. Unlike prior methods that often achieve fairness by sacrificing performance in advantaged groups, FairMoE enhances performance across all groups while preserving comparable fairness

metrics, demonstrating its potential as a practical and effective approach for fair skin disease diagnosis. The code is publicly available at link.

2 Motivation

In applications such as recidivism risk assessment [9] or income prediction [3], fairness requires excluding sensitive attributes such as race and gender from training to prevent bias. However, in skin disease diagnosis, these sensitive attributes provide essential clinical insights and cannot simply be removed. For example, skin type offers crucial information for clinicians assessing UV susceptibility [4], while race and gender also influence disease prevalence—malignant melanoma is 20 times more common in African Americans compared to other groups [17]. Similarly, men are twice as likely to develop basal cell carcinoma (BCC) and three times more likely to develop squamous cell carcinoma (SCC) than women [10]. Therefore, removing sensitive attributes as done in most fairness-aware algorithms risks compromising performance, which is not an option in medical diagnostic models where delivering precise diagnoses is absolutely crucial. While existing approaches [18,24] attempt to improve fairness by explicitly incorporating sensitive attributes and training separate models for different demographic groups by only using in-group data, such models often suffer from limited generalization due to reduced training data. Motivated by the needs, this paper proposes FairMoE, which integrates sensitive attributes to enhance fairness while training group-specific experts. Unlike methods mentioned above, FairMoE selectively incorporates out-of-group data to improve generalization, achieving fairness while maintaining model's performance.

3 Methodology

3.1 Problem Definition and Introduction of MoE Model

Problem Definition. Consider a dataset $D = \{(x_i, y_i, c_i)\}_{i=1}^{N}$, where $x_i \in \mathcal{X}$ represents the input image, $y_i \in \mathcal{Y}$ is the ground-truth class label, and $c_i \in \mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ denotes the sensitive attribute (e.g., age, skin type), which we also refer to as "group label" in this paper. We define our MoE model as $\hat{y}_i = F(\theta, x_i)$, where F maps the input x_i to the predicted label $\hat{y}_i$ and is parameterized by weights θ. Our objective is to enhance model performance across different groups by leveraging the sensitive attributes, we incorporate c_i during training to guide the routing mechanism. Accordingly, we denote the MoE model as $\hat{y}_i = F(\theta, x_i, c_i)$, explicitly indicating that c_i is included in the training.

MoE Model. Our MoE model extends a convolutional neural network (CNN) backbone by replacing each convolutional (Conv) layer with a MoE Conv Layer. A MoE Conv Layer comprises a set of expert networks $E_1, E_2, \ldots, E_n$ and a routing network G [20]. Each expert is a Conv layer with the same dimensions as the corresponding Conv layer in the original backbone CNN. The routing network is a significantly smaller Conv layer that directs data flow. Each expert

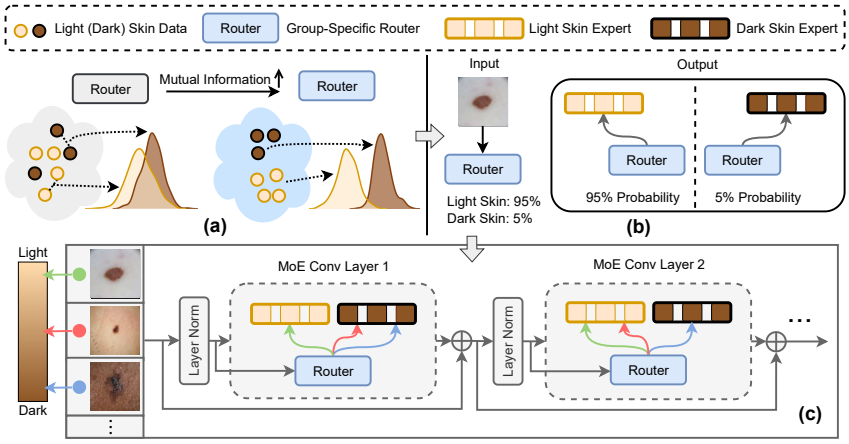


Fig. 1: Overview of FairMoE Framework: (a) Establishing correlations between groups and experts, (b) implementing soft probabilistic expert routing, and (c) illustrating the expert selection process within a ResNet-18 module.

is specialized for a specific group, so we set the number of experts n equal to the number of groups m. During training and inference, the routing network dynamically selects an expert for a given sample and forwards it to the next layer. The core of FairMoE lies in the routing network, which governs how experts receive training data and determines expert selection during inference.

3.2 FairMoE Framework Overview

Figure 1 provides an overview of the FairMoE framework, where the sensitive attribute is binary skin tone, and each layer has two corresponding experts. The design of FairMoE consists of two key components: First, as shown in Figure 1a, the router maps each sample to the expert corresponding to its group. To ensure that samples from a given group are forwarded to their group-specific expert with high probability, the router maximizes the mutual information between the group label C_i and the assigned expert E_i. Second, as shown in Figure 1b, instead of rigid assignments, the router uses a probabilistic selection strategy to mitigate data scarcity in group-specific training, particularly benefiting boundary samples. With both components successfully deployed, Figure 1c illustrates the expert selection process within a ResNet-18 with two MoE Conv Layers. Samples for which the router has high confidence in their group membership are assigned to the corresponding expert with high probability. While for samples near the decision boundary—such as those with intermediate skin tones—expert selection becomes more stochastic, leading to dynamic assignments across different layers.

3.3 Encouraging Specialization Through Group-Aware Assignments

In MoE training, when the router's assignment is not constrained by group information, samples are assigned to experts with nearly uniform probability, preventing the formation of group-specific expertise. However, our approach seeks to specialize experts for distinct groups, ensuring each expert is optimized for classifying data within its respective group.

To quantitatively enforce this expert-group specialization, we define $P(E_i|C_j)$ as the probability of a sample from group C_j being assigned to expert E_i. Using conditional probability, the joint probability is $P(E, C) = P(E|C)P(C)$, where $P(C)$ is a constant depending on the relative sizes of different groups in the dataset. In order to explicitly build dependence between experts and groups, we measure their correlation using mutual information, formulated as:

$$I(C; E) = \sum_{i=1}^{m} \sum_{j=1}^{m} P(C_i, E_j) \log \frac{P(C_i, E_j)}{P(C_i)P(E_j)}. \tag{1}$$

In this context, m represents both the number of experts and the number of groups. If experts are assigned to all groups with equal probability, meaning there is no correlation between groups and experts, then the joint probability satisfies $P(C_i, E_j) = P(C_i)P(E_j)$, resulting in a mutual information value of zero. Conversely, if each expert is exclusively assigned to a single group, the mutual information is maximized. In practice, we incorporate $-I(C; E_Y)$ into the total loss for each MoE layer Y, weighted by a parameter w_{MI}, where E_Y represents all experts within layer Y.

3.4 Soft Probabilistic Expert Routing

If a strong dependency forms between experts and groups, the router effectively functions as a group classifier, consistently assigning each sample to a fixed expert. While this structure enables expert specialization, it introduces two major limitations that ultimately compromise model performance. First, confining an expert to a single group reduces its available training data, weakening generalization, especially in imbalanced datasets. For example, if a group contains only 10% of the dataset, its expert trains on a limited subset, restricting its ability to learn robust patterns and resulting in suboptimal performance. Second, for continuous attributes like skin tone or age, hard group assignments fail to capture boundary features effectively. For instance, when groups are defined as age < 55 and age ≥ 55, experts trained exclusively on these subsets struggle to generalize in transition regions (e.g., ages 50–60), leading to reduced predictive accuracy.

To balance specialization and generalization while enhancing performance on group boundary samples, it is crucial to selectively incorporate out-of-group data when training group-specific experts. To minimize distribution shift, priority should be given to samples closer to the target group's distribution. We achieve this by leveraging the router's confidence score as a proxy for a sample's proximity to each group's distribution, using it as the probability for expert selection. For example, a sample with a high probability of belonging to the light-skin

group should primarily train the light-skin expert, whereas assigning it to the dark-skin expert would introduce distribution shift. Conversely, a sample with equal probabilities for both groups should contribute equally to both experts. This approach replaces the rigid, hard-assignment strategy—where each sample is strictly assigned to the expert based on its predicted group—with a soft, probabilistic assignment, where selection probability reflects a sample's contribution to different groups. The probability for selecting expert E_k is given by:

$$P(E_k \mid x) = \frac{\alpha_k \cdot s_k(x)}{\sum_j \alpha_j \cdot s_j(x)}, \tag{2}$$

where $s_k(x)$ denotes the router's confidence score for assigning sample x to group k, and $\alpha_k = \frac{1}{N_k}$ balances expert selection based on group sizes N_k. Without α_k, the router would disproportionately assign samples to majority-group experts, leading to uneven training.

4 Experiments and Results

Dataset and Model. FairMoE is evaluated on two skin disease classification datasets: Fitzpatrick-17k [11] and ISIC 2019 [6, 22]. Fitzpatrick-17k contains 16,577 images representing 114 different skin conditions, with skin types 1-3 grouped as light skin and 4-6 as dark skin, following [26]. ISIC 2019 contains 25331 images across 8 diagnostic categories, where age as the sensitive attribute, grouping individual ≤ 55 as young and > 55 as old. We use ResNet-18 [13] as the backbone for Fitzpatrick-17k and VGG-11 [21] for ISIC 2019, extending their original convolutional layers to MoE Conv layers. The training settings were the same with those described in [26]. A standard preprocessing step for both datasets involves resizing all the images to a uniform size of 128×128 pixels. The mutual information loss constant w_{MI} is set to 0.01.

Fairness Metrics. We employ the multi-class equalized opportunity (Eopp) and equalized odds (Eodd) metrics to evaluate the fairness of our model. We follow [26] for calculating these metrics. We use the FATE metric from [28] to assess the overall performance of the model balancing accuracy and fairness. A higher FATE score indicates a better trade-off between fairness and accuracy.

Baselines. We present a comprehensive comparison of our framework against various bias mitigation baselines. Vanilla is the model trained directly on ResNet-18 or VGG-11 backbone. DomainIndep learns group-specific classifiers with shared parameters [25]. FairAdaBN achieves fairness by modifying the batch normalization layer [28]. SCP-FairPrune [16] and FairQuantize [12] is the latest work achieves fairness on skin disease through pruning and quantization, respectively.

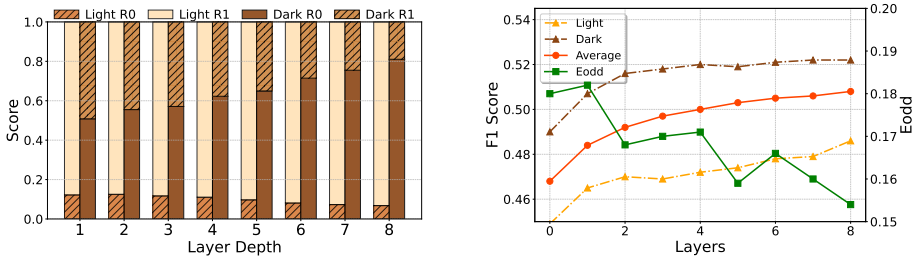
Results on Fitzpatrick-17k Dataset. The upper section of Table 1 presents a comparative analysis of accuracy and fairness across methods on the Fitzpatrick-17k dataset, highlighting FairMoE's superior performance. FairMoE achieves the highest F1-score (0.502), outperforming the second-best baseline SCP-FairPrune (0.476). Additionally, it attains the lowest Eopp1 and Eodd scores, indicating improved fairness. Furthermore, FairMoE substantially increases FATE scores,

Table 1: Results of accuracy and fairness on Fitzpatrick-17k and ISIC2019 datasets. The best performance is highlighted in bold.

		Accuracy			Fairness		
Method	Group	Precision	Recall	F1-score	Eopp0 ↓ / FATE ↑	Eopp1 ↓ / FATE ↑	Eodd ↓ / FATE ↑
Fitzpatrick-17k Dataset							
ResNet-18	Dark	0.512	0.511	0.490	0.0031 / 0.000	0.332 / 0.000	0.180 / 0.000
	Light	0.467	0.468	0.449			
	Avg. ↑	0.489	0.490	0.468			
	Diff. ↓	0.045	0.043	0.041			
DomainIndep	Dark	0.501	0.522	0.492	0.0030 / 0.041	0.320 / 0.045	0.163 / 0.103
	Light	0.465	0.479	0.459			
	Avg. ↑	0.484	0.497	0.472			
	Diff. ↓	0.036	0.043	0.033			
FairAdaBN	Dark	0.493	0.495	0.469	0.0030 / 0.007	0.302 / 0.065	0.171 / 0.024
	Light	0.450	0.443	0.435			
	Avg. ↑	0.471	0.473	0.456			
	Diff. ↓	0.043	0.052	0.034			
SCP-FairPrune	Dark	0.512	0.512	0.490	0.0030 / 0.049	0.289 / 0.147	0.164 / 0.106
	Light	0.490	0.472	0.469			
	Avg. ↑	0.495	0.497	0.476			
	Diff. ↓	**0.022**	**0.040**	**0.021**			
FairQuantize	Dark	0.498	0.513	0.480	**0.0028** / 0.082	0.291 / 0.109	0.156 / 0.118
	Light	0.459	0.470	0.448			
	Avg. ↑	0.479	0.489	0.461			
	Diff. ↓	0.043	0.053	0.032			
FairMoE	Dark	**0.540**	**0.554**	**0.522**	0.0030 / **0.105**	**0.285** / **0.214**	**0.154** / **0.217**
	Light	**0.505**	**0.497**	**0.486**			
	Avg. ↑	**0.518**	**0.525**	**0.502**			
	Diff. ↓	0.035	0.057	0.036			
ISIC2019 Dataset							
VGG-11	Young	0.669	0.724	0.687	0.023 / 0.000	0.150 / 0.000	0.078 / 0.000
	Old	0.758	0.798	0.766			
	Avg. ↑	0.723	0.773	0.741			
	Diff. ↓	0.089	0.074	0.080			
DomainIndep	Young	0.677	0.724	0.689	0.021 / 0.090	0.103 / 0.316	0.051 / 0.345
	Old	0.758	0.796	0.766			
	Avg. ↑	0.724	0.772	0.743			
	Diff. ↓	0.081	0.072	0.078			
FairAdaBN	Young	0.643	0.717	0.663	0.018 / 0.196	**0.084** / 0.418	0.060 / 0.209
	Old	0.727	0.770	0.748			
	Avg. ↑	0.699	0.762	0.725			
	Diff. ↓	0.085	0.053	0.086			
SCP-FairPrune	Young	0.662	0.724	0.683	0.020 / 0.125	0.098 / 0.341	0.068 / 0.123
	Old	0.745	0.792	0.761			
	Avg. ↑	0.717	0.769	0.737			
	Diff. ↓	0.083	0.068	0.078			
FairQuantize	Young	0.657	0.718	0.676	0.019 / 0.163	0.088 / 0.403	0.060 / 0.220
	Old	0.743	0.785	0.755			
	Avg. ↑	0.707	0.765	0.733			
	Diff. ↓	0.086	0.067	0.080			
FairMoE	Young	**0.711**	**0.768**	**0.725**	**0.016** / **0.340**	0.089 / **0.441**	**0.046** / **0.445**
	Old	**0.782**	**0.819**	**0.788**			
	Avg. ↑	**0.760**	**0.795**	**0.767**			
	Diff. ↓	**0.072**	**0.051**	**0.063**			

with a 46% increase in FATE (Eopp0) and 104% in FATE (Eodd) compared to SCP-FairPrune, demonstrating the effectiveness of FairMoE.

Results on ISIC 2019 Dataset. Similar to Fitzpatrick-17k, FairMoE outperforms all baselines in accuracy and two fairness metrics. Unlike most methods that sacrifice accuracy for fairness, FairMoE effectively avoids this trade-off by employing group-specific experts, allowing each group to optimize its own performance rather than relying on a shared model structure. Compared to Do-



(a) (b)

Fig. 2: (a) Router scores for dark and light skin experts across model layers. (b) Impact of the number of MoE Conv layers on model performance.

mainIndep, which also utilizes group-specific modules for fairnessis, FairMoE achieves significantly better accuracy and fairness scores. Additionally, FairMoE exhibits the smallest gap across all accuracy metrics, indicating greater benefits for unprivileged groups and further enhancing fairness.

Case Study. In the Fitzpatrick-17k dataset, although light-skin samples outnumber dark-skin samples by nearly a 2:1 ratio, its classification performance is significantly lower for light skin. We hypothesize that dark-skin samples shift the model's focus, negatively impacting light-skin predictions, and that training group-specific experts may better improve light-skin performance. Figure 2a shows MoE routing scores for each skin type as network depth increases. R0 corresponds to the dark-skin expert, and R1 to the light-skin expert. The result shows that the tendency to select specific experts strengthens with network depth, aligning with findings in [5] that bias increases in deeper layers. Moreover, the privileged group (dark skin) shows no preference for its expert in shallow layers, with selection gradually increasing in deeper layers. In contrast, the unprivileged group (light skin) strongly favors its expert from the outset, with this preference intensifying as depth increases. These results suggest that group-specific experts may provide greater benefits to unprivileged groups.

Ablation Study. To evaluate the impact of each MoE layer, we incrementally added them, starting from the final layer and gradually extending to earlier layers, as shown in Figure 2b. The figure shows that increasing the number of MoE layers gradually improves model performance while generally improves fairness, with this effect being more pronounced in deeper layers and less significant in shallower ones. These findings underscore a trade-off between performance, fairness, and computational resources, highlighting the flexibility of MoE layer selection based on resource availability. When computational resources are abundant, the full MoE structure can be employed for optimal performance. Conversely, in resource-constrained scenarios, MoE layers can be selectively applied to only the deeper layers, striking a balance between efficiency and effectiveness.

5 Conclusion

In this paper, we contend that eliminating sensitive attributes to achieve fairness is impractical in skin disease diagnosis, as these attributes provide essential diagnostic cues. Instead, we propose FairMoE, a model with layer-wise mixture-of-experts modules that incorporate group information for specialized learning. FairMoE leverages mutual information loss to establish expert-group correlations and enhances generalization through soft probabilistic routing, selectively incorporating out-of-group data. Experimental results demonstrate that FairMoE significantly improves accuracy while maintaining comparable fairness metrics.

Acknowledgments. This project is supported in part by NIH Grant R01EB033387. Yuying Duan and Michael Lemmon acknowledge the support from NSF under Grant No. CNS-2228092.

Disclosure of Interests. The authors declare no competing interests.

References

1. Ahammed, M., Al Mamun, M., Uddin, M.S.: A machine learning approach for skin disease detection and classification using image segmentation. Healthcare Analytics **2**, 100122 (2022)
2. Alvi, M., Zisserman, A., Nellåker, C.: Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings. In: Proceedings of the European conference on computer vision (ECCV) workshops. pp. 0–0 (2018)
3. Asuncion, A., Newman, D.: Uci machine learning repository (November 2007), `https://archive.ics.uci.edu/ml/`
4. Caini, S., Gandini, S., Sera, F., Raimondi, S., Fargnoli, M.C., Boniol, M., Armstrong, B.K.: Meta-analysis of risk factors for cutaneous melanoma according to anatomical site and clinico-pathological variant. European journal of cancer **45**(17), 3054–3063 (2009)
5. Chiu, C.H., Chung, H.W., Chen, Y.J., Shi, Y., Ho, T.Y.: Toward fairness through fair multi-exit framework for dermatological disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 97–107. Springer (2023)
6. Combalia, M., Codella, N.C., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., Carrera, C., Barreiro, A., Halpern, A.C., Puig, S., et al.: Bcn20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288 (2019)
7. Duan, Y., Tian, Y., Chawla, N., Lemmon, M.: Post-fair federated learning: Achieving group and community fairness in federated learning via post-processing. arXiv preprint arXiv:2405.17782 (2024)
8. Duan, Y., Xu, G., Shi, Y., Lemmon, M.: The cost of local and global fairness in federated learning. arXiv preprint arXiv:2503.22762 (2025)
9. Flores, A.W., Bechtel, K., Lowenkamp, C.T.: False positives, false negatives, and false analyses: A rejoinder to machine bias: There's software used across the country to predict future criminals, and it's biased against blacks. Federal Probation **80**, 38 (2016)

10. Gordon, R.: Skin cancer: an overview of epidemiology and risk factors. In: Seminars in oncology nursing. vol. 29, pp. 160–169. Elsevier (2013)
11. Groh, M., Harris, C., Soenksen, L., Lau, F., Han, R., Kim, A., Koochek, A., Badri, O.: Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1820–1828 (2021)
12. Guo, Y., Jia, Z., Hu, J., Shi, Y.: Fairquantize: Achieving fairness through weight quantization for dermatological disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 329–338. Springer (2024)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
14. Jung, S., Lee, D., Park, T., Moon, T.: Fair feature distillation for visual recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12115–12124 (2021)
15. Kampen, P.J.T., Christensen, A.N., Hannemose, M.R.: Is this hard for you? Personalized human difficulty estimation for skin lesion diagnosis . In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15012, pp. 175 – 184. Springer Nature Switzerland (October 2024)
16. Kong, Q., Chiu, C.H., Zeng, D., Chen, Y.J., Ho, T.Y., Hu, J., Shi, Y.: Achieving fairness through channel pruning for dermatological disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 24–34. Springer (2024)
17. Narayanan, D.L., Saladi, R.N., Fox, J.L.: Ultraviolet radiation and skin cancer. International journal of dermatology **49**(9), 978–986 (2010)
18. Puyol-Antón, E., Ruijsink, B., Piechnik, S.K., Neubauer, S., Petersen, S.E., Razavi, R., King, A.P.: Fairness in cardiac mr image analysis: an investigation of bias due to data imbalance in deep learning based segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 413–423. Springer (2021)
19. Quadrianto, N., Sharmanska, V., Thomas, O.: Discovering fair representations in the data domain. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8227–8236 (2019)
20. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538 (2017)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
22. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**(1), 1–9 (2018)
23. Wang, J., Wang, X.E., Liu, Y.: Understanding instance-level impact of fairness constraints. In: International Conference on Machine Learning. pp. 23114–23130. PMLR (2022)
24. Wang, M., Deng, W.: Mitigating bias in face recognition using skewness-aware reinforcement learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9322–9331. IEEE (2020)
25. Wang, Z., Qinami, K., Karakozis, I.C., Genova, K., Nair, P., Hata, K., Russakovsky, O.: Towards fairness in visual recognition: Effective strategies for bias mitigation.

In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8919–8928 (2020)
26. Wu, Y., Zeng, D., Xu, X., Shi, Y., Hu, J.: Fairprune: Achieving fairness through pruning for dermatological disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 743–753. Springer (2022)
27. Xu, G., Wu, Y., Hu, J., Shi, Y.: Achieving fairness in dermatological disease diagnosis through automatic weight adjusting federated learning and personalization. arXiv preprint arXiv:2208.11187 (2022)
28. Xu, Z., Zhao, S., Quan, Q., Yao, Q., Zhou, S.K.: Fairadabn: Mitigating unfairness with adaptive batch normalization and its application to dermatological disease classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 307–317. Springer (2023)
29. Zafar, M.B., Valera, I., Gomez-Rodriguez, M., Gummadi, K.P.: Fairness constraints: A flexible approach for fair classification. Journal of Machine Learning Research **20**(75), 1–42 (2019)
30. Zhang, B.H., Lemoine, B., Mitchell, M.: Mitigating unwanted biases with adversarial learning. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. pp. 335–340 (2018)
31. Zhang, Y., Cai, R., Chen, T., Zhang, G., Zhang, H., Chen, P.Y., Chang, S., Wang, Z., Liu, S.: Robust mixture-of-expert training for convolutional neural networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 90–101 (2023)
32. Zhou, J., Wu, Z., Jiang, Z., Huang, K., Guo, K., Zhao, S.: Background selection schema on deep learning-based classification of dermatological disease. Computers in Biology and Medicine **149**, 105966 (2022)