15. LeCun, Y., Denker, J., Solla, S.: Optimal brain damage. Advances in neural information processing systems **2** (1989)
16. 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)
17. Roh, Y., Lee, K., Whang, S., Suh, C.: Fr-train: A mutual information-based approach to fair and robust training. In: International Conference on Machine Learning. pp. 8147–8157. PMLR (2020)
18. Seyyed-Kalantari, L., Zhang, H., McDermott, M.B., Chen, I.Y., Ghassemi, M.: Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nature medicine **27**(12), 2176–2182 (2021)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
20. 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)
21. Tzeng, E., Hoffman, J., Darrell, T., Saenko, K.: Simultaneous deep transfer across domains and tasks. In: Proceedings of the IEEE international conference on computer vision. pp. 4068–4076 (2015)
22. Wan, M., Zha, D., Liu, N., Zou, N.: In-processing modeling techniques for machine learning fairness: A survey **17**(3) (mar 2023). https://doi.org/10.1145/3551390
23. Wang, A., Russakovsky, O.: Directional bias amplification. In: International Conference on Machine Learning. pp. 10882–10893. PMLR (2021)
24. 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)
25. Wang, Z., Dong, X., Xue, H., Zhang, Z., Chiu, W., Wei, T., Ren, K.: Fairness-aware adversarial perturbation towards bias mitigation for deployed deep models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10379–10388 (2022)
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, Z., Zhao, S., Quan, Q., Yao, Q., Zhou, S.K.: Fairadabn: Mitigating unfairness with adaptive batch normalization and its application to dermatological disease classification. arXiv preprint arXiv:2303.08325 (2023)
28. Yang, J., Soltan, A., Eyre, D., Yang, Y., Clifton, D.: An adversarial training framework for mitigating algorithmic biases in clinical machine learning. NPJ digital medicine **6**, 55 (03 2023)
29. Yao, R., Cui, Z., Li, X., Gu, L.: Improving fairness in image classification via sketching (2022)
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)

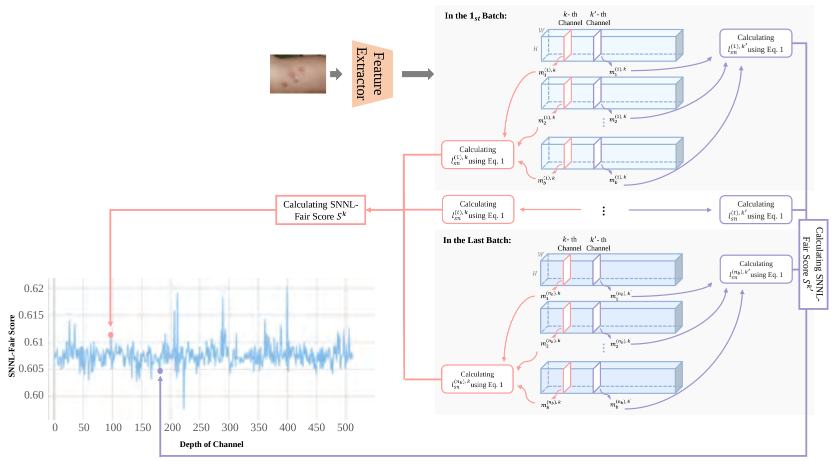


**Fig. 1.** Illustration of SNNL-Fair metric calculation. Here, $t$ represents the batch index, $b$ represents the batch size, $k$ represent the depth of channel, $n_b$ represents total number of batches, and $m_b^{(t),k}$ represent the feature map at the $k$-th channel in the $t$-th batch.

**Table 1.** Additional results of accuracy and fairness on the Fitzpatrick-17k and VGG-11 backbone, using skin tone as the sensitive attribute. The dark skin is the privileged group. *FATE* metrics are evaluated using the vanilla VGG-11 as the baseline. ($pr$ is the pruning ratio, $n$ is the pruning iteration(s), and $pr_c$ is the channel pruning ratio.)

| | | Accuracy | | | Fairness | | |
|---|---|---|---|---|---|---|---|
| Method | Skin Tone | Precision | Recall | F1-score | *Eopp0*↓ / *FATE*↑ | *Eopp1*↓ / *FATE*↑ | *Eodd*↓ / *FATE*↑ |
| AdvConf [4] | Dark | 0.506 | 0.562 | 0.506 | **0.0011** / 0.0676 | 0.339 / -0.0253 | 0.169 / -0.0148 |
| | Light | 0.427 | 0.464 | 0.426 | | | |
| | Avg.↑ | 0.467 | 0.513 | 0.466 | | | |
| | Diff.↓ | 0.079 | 0.098 | 0.080 | | | |
| AdvRef [21] | Dark | 0.514 | 0.545 | 0.503 | **0.0011** / 0.0950 | 0.334 / 0.0160 | 0.166 / 0.0291 |
| | Light | 0.489 | 0.469 | 0.457 | | | |
| | Avg.↑ | 0.502 | 0.507 | 0.480 | | | |
| | Diff.↓ | 0.025 | 0.076 | 0.046 | | | |
| DomainIndep [24] | Dark | 0.547 | 0.567 | 0.532 | 0.0012 / 0.0416 | 0.344 / 0.0118 | 0.172 / 0.0197 |
| | Light | 0.455 | 0.480 | 0.451 | | | |
| | Avg.↑ | 0.501 | 0.523 | 0.492 | | | |
| | Diff.↓ | 0.025 | 0.076 | 0.046 | | | |
| OBD [15] (pr=35%) | Dark | 0.557 | 0.570 | 0.536 | 0.0012 / 0.0691 | 0.360 / -0.0051 | 0.180 / 0.0031 |
| | Light | 0.488 | 0.494 | 0.475 | | | |
| | Avg.↑ | 0.523 | 0.532 | 0.506 | | | |
| | Diff.↓ | 0.069 | 0.076 | 0.061 | | | |
| SCP-FairPrune (Ours) ($pr_c = 2\%, n = 3$) | Dark | 0.568 | 0.576 | 0.547 | 0.0012 / **0.0965** | **0.278** / **0.2495** | **0.139** / **0.2559** |
| | Light | 0.499 | 0.504 | 0.492 | | | |
| | Avg.↑ | 0.533 | 0.540 | 0.520 | | | |
| | Diff.↓ | 0.069 | 0.073 | 0.055 | | | |

**Table 2.** Additional results of accuracy and fairness on the ISIC 2019 dataset and ResNet-18 backbone, using gender as the sensitive attribute. Female is the privileged group. *FATE* metrics are evaluated using the vanilla ResNet-18 as the baseline. ($n$ is the pruning iteration(s), and $pr_c$ is the channel pruning ratio.)

| Method | Gender | Accuracy: Precision | Accuracy: Recall | Accuracy: F1-score | Fairness: *Eopp0*↓ / *FATE*↑ | Fairness: *Eopp1*↓ / *FATE*↑ | Fairness: *Eodd*↓ / *FATE*↑ |
|---|---|---|---|---|---|---|---|
| AdvConf [4] | Female | 0.755 | 0.738 | 0.741 | 0.008 / 0.8684 | 0.070 / -0.5748 | 0.037 / -0.6574 |
| | Male | 0.710 | 0.757 | 0.731 | | | |
| | Avg.↑ | 0.733 | 0.747 | 0.736 | | | |
| | Diff.↓ | 0.045 | 0.020 | 0.010 | | | |
| AdvRef [21] | Female | 0.778 | 0.683 | 0.716 | 0.007 / 0.8674 | 0.059 / -0.5700 | 0.033 / -0.4957 |
| | Male | 0.773 | 0.706 | 0.729 | | | |
| | Avg.↑ | 0.775 | 0.694 | 0.723 | | | |
| | Diff.↓ | 0.006 | 0.023 | 0.014 | | | |
| DomainIndep [24] | Female | 0.729 | 0.747 | 0.734 | 0.010 / 0.8106 | 0.086 / -0.9597 | 0.042 / -0.9061 |
| | Male | 0.725 | 0.694 | 0.702 | | | |
| | Avg.↑ | 0.727 | 0.721 | 0.718 | | | |
| | Diff.↓ | 0.004 | 0.053 | 0.031 | | | |
| SCP-FairPrune (Ours) ($pr_c = 2\%, n = 3$) | Female | 0.787 | 0.701 | 0.736 | **0.006 / 0.9018** | **0.015 / 0.6724** | **0.006 / 0.7411** |
| | Male | 0.765 | 0.712 | 0.735 | | | |
| | Avg.↑ | 0.776 | 0.707 | 0.736 | | | |
| | Diff.↓ | 0.022 | 0.012 | 0.001 | | | |

**Table 3.** Accuracy and fairness of classification results across different baselines with and without the SNNL-based Channel Pruning framework on the Fitzpatrick17k dataset. SCP-"X" refers to applying our framework to the "X" model. "X" model is also the baseline used in *FATE* metric evaluation. Our framework always achieves positive *FATE* suggesting better accuracy-fairness trade-off. ($n$ is the pruning iteration(s), and $pr_c$ is the channel pruning ratio.)

| Method | Skin Tone | Accuracy: Precision | Accuracy: Recall | Accuracy: F1-score | Fairness: Eopp0↓ / FATE↑ | Fairness: Eopp1↓ / FATE↑ | Fairness: Eodd↓ / FATE↑ |
|---|---|---|---|---|---|---|---|
| VGG-11 [19] | Dark | 0.563 | 0.581 | 0.546 | 0.0013 / 0.0000 | 0.361 / 0.0000 | 0.182 / 0.0000 |
| | Light | 0.482 | 0.495 | 0.473 | | | |
| | Avg.↑ | 0.523 | 0.538 | 0.510 | | | |
| | Diff.↓ | 0.081 | 0.086 | 0.073 | | | |
| SCP-VGG-11 ($pr_c = 2\%, n = 3$) | Dark | 0.580 | 0.583 | 0.552 | **0.0013 / 0.0301** | **0.286 / 0.2371** | **0.143 / 0.2433** |
| | Light | 0.511 | 0.506 | 0.498 | | | |
| | Avg.↑ | 0.545 | 0.544 | 0.525 | | | |
| | Diff.↓ | 0.069 | 0.077 | 0.054 | | | |
| HSIC [16] | Dark | 0.548 | 0.522 | 0.513 | 0.0013 / 0.0000 | 0.331 / 0.0000 | 0.166 / 0.0000 |
| | Light | 0.513 | 0.506 | 0.486 | | | |
| | Avg.↑ | 0.530 | 0.515 | 0.500 | | | |
| | Diff.↓ | 0.040 | 0.018 | 0.029 | | | |
| SCP-HSIC ($pr_c = 2\%, n = 3$) | Dark | 0.525 | 0.518 | 0.504 | **0.0012 / 0.0609** | **0.304 / 0.0656** | **0.152 / 0.0683** |
| | Light | 0.477 | 0.510 | 0.479 | | | |
| | Avg.↑ | 0.501 | 0.514 | 0.492 | | | |
| | Diff.↓ | 0.048 | 0.008 | 0.025 | | | |