DermEquity: A Universal Safety Framework for Medical AI Using Worst-Case Bias Certification and Influence-Based Attribution Validated in Skin Cancer Screening

Lau, Kris (School: Signature School)
Xiu, Angie (School: Carmel High School)
Mi, Alex (School: Carmel High School)

Medical AI systems fail on underrepresented populations in ways conventional fairness metrics cannot detect. Existing evaluations report average disparities at fixed thresholds, providing no safety guarantees across the full clinical operating range. We built DermEquity, a universal pre-deployment bias certification framework with five components: (1) a balanced benchmark of 1,658 skin lesion images stratified by Fitzpatrick skin type and malignancy; (2) Worst-Case Underdiagnosis Gap (WCUG), measuring maximum false negative rate disparity across 101 thresholds; (3) Influence-Based Bias Attribution using leave-one-out analysis to identify training samples driving worst-case bias; (4) external validation on EfficientNetB3, the ISIC 2019 winning model trained on 25,331 images; and (5) cross-domain validation on chest radiography to establish framework universality. Four InceptionV3 models trained on imbalanced Fitzpatrick17k data (0%–10% dark skin) were evaluated with bootstrapped 95% CIs and McNemar's test (a=0.01). A model with 76.4% average accuracy masked a 6.8% melanoma miss rate gap for dark skin at threshold 0.40. Among 200 LOO models, 66% of top-50 bias-amplifying samples were light skin; removing them reduced WCUG by 15.6% (p=0.008) without architectural changes. When applied to EfficientNetB3, DermEquity revealed WCUG of 0.089 despite 73.6% average accuracy. Cross-domain application to NIH ChestXRay-14 (112,120 images, sex-stratified) confirmed WCUG detects worst-case underdiagnosis gaps in female patients invisible to average accuracy. A sub-$30 3D-printed dermoscope integrates with a mobile app running the bias-certified model for community deployment. DermEquity establishes the first universal pre-deployment safety standard for medical AI.

Awards Won: