Case Studies

Scenario 6: AI-Assisted Diagnosis of Skin Cancer Using Dermoscopy Images

Background: AI tool is being developed to assist dermatologists in diagnosing skin cancer from dermoscopy images. Skin cancer is a serious condition that can be effectively treated if caught early. The tool aims to classify images as either malignant (cancerous) or benign (non-cancerous).

Development and Training: The dataset used for training the AI tool includes images from patients of various ethnic backgrounds. However, there are significant demographic disparities in the dataset. Caucasian patients, who have a higher prevalence of certain types of skin cancer, are overrepresented compared to patients who are more likely to have darker skin tones, such as those of African, Asian, or Hispanic descent. To help mitigate bias, skin tone information will be recorded in the training data using the Fitzpatrick skin tone scale, which categorizes skin types from Type I (light skin) to Type VI (dark skin) based on their response to UV light.

Questions to consider:

- What dangers/risks of the use of AI for this problem can you identify at this stage?
- How would you go about addressing these?
- What fairness metric(s) do you think might be appropriate when assessing the AI tool for potential bias?

REMINDER - DEFINITIONS OF FAIRNESS

- *False Negative Rate (FNR): the rate at which positive cases are missed by the classifier*
- *Demographic parity - equal chance of being classified positive for each protected group*
- *Equalised odds - equal true positive rate (TPR) & false positive rate (FPR) for each protected group*
- *Equal opportunity - only equalise either FPR or FNR, not both*

Suggested answers/discussion points:

- What dangers/risks of the use of AI for this problem can you identify at this stage?
    - **Bias in Training Data:** The overrepresentation of Caucasian patients in the training data might lead to an AI model that is less accurate for diagnosing skin cancer in patients with darker skin tones. This could result in higher rates of false negatives or false positives for these groups.
    - **Biased Predictions:** The AI model might have higher false negative rates for underrepresented groups, leading to missed diagnoses of skin cancer in these populations.
    - **Generalization Issues:** The tool may not generalize well to different demographic groups if the characteristics of skin lesions in these groups are not adequately represented in the training data.
    - **Ethical and Equity Concerns:** If the AI tool is less effective for certain demographic groups, it could exacerbate existing health disparities, as individuals from underrepresented groups might receive less accurate diagnoses and subsequent care.
    - **Lack of Diverse Image Characteristics:** Skin cancer can present differently on different skin tones, and a lack of diverse images may mean the model fails to learn important characteristics necessary for accurate diagnosis across all skin types.
- How would you go about addressing these?
    - **Expand and Balance the Dataset:** Actively seek to include more dermoscopy images from patients with darker skin tones. Partner with dermatology clinics and hospitals that serve more diverse populations to obtain a more representative dataset.
    - **Demographic Parity:** Implement demographic parity to ensure that the tool's predictions are equitable across different demographic groups. This means adjusting the model to ensure that the probability of predicting a malignant lesion is similar for all demographic groups, regardless of their representation in the dataset.
    - **Resampling Techniques:** Use oversampling or undersampling techniques to balance the number of images from different demographic groups.
    - **Synthetic Data Generation:** Generate synthetic images to augment the dataset, especially for underrepresented groups, using techniques like GANs (Generative Adversarial Networks).
    - **Bias Mitigation Algorithms:** Apply bias mitigation algorithms during training to adjust the model's learning process and reduce disparities in predictions across demographic groups.
    - **Regular Retraining and Validation:** Continuously retrain and validate the AI model with new, diverse data to ensure it remains effective and fair across all demographic groups.

    - **Thorough Subgroup Analysis:** Perform detailed analysis on the model's performance across different demographic groups, ensuring consistent accuracy and reliability across all skin tones.
    - **Record Skin Tone Information:** Use the Fitzpatrick skin tone scale to record and categorize skin tone information in the training data. This will help in analyzing the model's performance across different skin tones and ensuring fairness.
- What fairness metric(s) do you think might be appropriate when assessing the AI tool for potential bias?
    - **Demographic Parity:** Ensure that the model's positive prediction rate (i.e., the rate at which the model predicts malignant lesions) is equal across different demographic groups.
    - **Equal Opportunity:** Ensure that the true positive rate (recall) is similar across different demographic groups, indicating that the model is equally good at identifying actual cancer cases in all groups.
    - **Calibration:** Verify that the predicted probabilities are reliable and consistent across different demographic groups, meaning that for any given probability score, the actual incidence of malignancy should be the same across all groups.
    - **False Negative Rate (FNR):** Minimize the FNR across all demographic groups to ensure that the AI tool does not miss cancer cases, particularly in underrepresented groups.
    - **Subgroup Analysis:** Conduct thorough subgroup analysis to ensure the AI model performs consistently across various subpopulations, such as different skin tones, age groups, and genders.

Conclusion: This case study emphasizes the importance of addressing data imbalance and ensuring fairness in AI-assisted medical diagnostics, particularly for skin cancer diagnosis using dermoscopy images. By expanding and balancing the dataset, applying bias mitigation techniques, recording skin tone information using the Fitzpatrick skin tone scale, and using appropriate fairness metrics such as demographic parity, equal opportunity, and calibration, the AI tool can be developed to provide accurate and equitable skin cancer diagnoses across all skin tones. This approach will help improve health outcomes, promote equitable healthcare, and reduce disparities in medical care, especially for underrepresented populations.