Fairness of AI for Skin Lesion Diagnosis

Supervisors: Prof Andy King (andrew.king@kcl.ac.uk), Dr Sweta Rai

Background: Artificial intelligence (AI) techniques based upon deep learning have shown impressive performance in automatically classifying malignant/benign skin lesions from photographic dermatology images [1] (see Figure). However, subsequent work has raised concerns about possibly *biased* or *unfair* behaviour of these techniques. For example, an AI model for classifying skin lesions that was trained using data from mostly lighter-skinned patients will tend to perform better on other lighter-skinned patients compared to darker-skinned patients, i.e. its performance will be *biased* in favour of lighter-skinned patients [2]. This is concerning because most available databases of dermatology images are either imbalanced by skin tone (with much higher representation of lighter skin tones) or do not report skin tone information [3].

Figure: A dermatology image of a skin lesion

Aims: This project aims to investigate and address bias by skin tone in AI models for skin lesion classification. First, an AI model will be trained to classify lesions as either benign or malignant using a large-scale database of dermatology images, which is imbalanced by skin tone. The performance and bias of this model will be assessed using hold-out test data from the same database (internal validation) as well as using other databases (external validation). Finally, experiments will be performed to investigate methods to *mitigate* the bias found, i.e. to make the performance of the model fairer and more equitable by skin tone. Specifically, planned activities for the project are:
   - Literature review on fairness/bias in AI and AI for skin lesion classification.
   - Train and quantitatively evaluate the performance and bias of AI models for skin lesion classification.
   - Investigate bias mitigation algorithms such as data augmentation and modified batch sample ordering.

Skills training: This project can accommodate two students, one from an engineering programme (BEng/MEng) and one from a medical programme (MBBS). The engineering student will be involved in data curation and using Python to train/evaluate the deep learning models for skin lesion classification. They will also perform quantitative analysis of the models' performance. Therefore, the project is suited for a student who would like to further develop their programming skills, in particular in the area of deep learning, as well as to gain experience of involvement in a real cutting-edge research project. Prior programming experience is required and experience with Python would be preferred but is not essential.

The medical student will perform a literature review on fairness of AI in medical image analysis and skin lesion diagnosis in particular. There would also be the opportunity to become involved in the practical aspects of the project if they wish.

Outputs: As well as writing the final project dissertation, depending on progress there could be the possibility to submit a paper to an international academic conference based on the work. The student would be an author on this paper.

References:

[1] Esteva et al, "Dermatologist-level Classification of Skin Cancer with Deep Neural Networks", Nature, 2017 (https://doi.org/10.1038/nature21056)

[2] Daneshjou et al, "Disparities in Dermatology AI Performance on a Diverse, Curated Clinical Image Set", Science Advances, 2022 (https://doi.org/10.1126/sciadv.abq6147)

[3] Daneshjou et al, "Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review", JAMA Dermatology, 2021 (https://doi.org/10.1001/jamadermatol.2021.3129)