

Methods for maintaining patient privacy when creating distributable 3D total body photography image sets for machine learning

Dr Liam Caffery¹, Dr Brigid Betz-Stablein², Professor Monika Janda¹, Professor Peter Soyer¹

¹*The University of Queensland, Brisbane, Australia*, ²*QIMR Berghofer Medical Research Institute, Brisbane, Australia*

Parallel Session 1D, Grand Ballroom 6, November 19, 2019, 13:30 - 15:00

Methods for maintaining patient privacy when creating distributable 3D total body photography image sets for machine learning

Caffery LJ¹, Betz-Stablein B², Janda M¹, Soyer HP¹

¹ *The University of Queensland, Brisbane, Australia*

² *QIMR Berghofer Medical Research Institute, Brisbane, Australia*

Three-dimensional total body photography (3D-TBP) is an emerging non-invasive imaging technique used for melanoma screening. A 3D-avatar of the patient's entire skin surface is created during imaging. In Australia, the use of 3D-TBP is being accelerated by the Australian Centre for Excellence in Melanoma Imaging and Diagnosis (ACEMID). The ACEMID project funded by the Australian Cancer Research Foundation will install 15 3D-TBP acquisition devices across Queensland, New South Wales and Victoria. These installations will produce approximately 100,000 3D-avatars within three years which will be labelled and amalgamated into a research database making them an ideal dataset for machine learning applications in the context of melanoma diagnosis.

A principle of ACEMID is to foster machine learning research, however challenges to protecting patient privacy when distributing the dataset to external researchers will need to be addressed before this can occur. Unlike other image-based machine learning datasets (e.g. radiology, dermoscopy), the patient's face and identifying features (e.g. tattoos) are clearly visible. There are additional sensitivities as the dataset includes semi-nude images of the patient. These problems are not unique to 3D-TBP and are likely to affect other photographic image sets.

This presentation will discuss best practice principles according to the privacy by design approach and present options on how to both protect the patient privacy while still allowing interdisciplinary research providing external researchers an accessible datasets for ML. This presentation will be applicable to other areas of research that require sharing of potentially identifiable information in the future.