The paper "Analyzing Biases in AU Activation Estimation Toward Fairer Facial Expression Recognition" by Miguel Monares, Yuan Tang, Ritik Raina, and Virginia R. de Sa examines racial biases in facial expression recognition (FER) models, focusing on disparities across different skin tones. The study specifically investigates biases in an action unit (AU) estimation network using synthetic facial data. By employing synthetically generated faces with varied skin tones but identical facial expressions and morphologies, the research highlights the existence of biases and underscores the importance of dataset diversity to achieve robust and fair models.

The authors created a dataset of 940 facial images using Character Creator 4 software, manipulating facial expressions based on the Facial Action Coding System (FACS). They conducted experiments using a pre-trained pain-estimation model, assessing its performance on synthetic faces with different skin tones. The study involved paired t-tests to identify statistically significant differences in model outputs between light and dark-skinned faces, revealing biases in AU activation and PSPI scores. Further analyses showed that biases varied across different action units and were influenced by the training data's skin tone distribution.

The results indicated that models trained on lighter-skinned faces performed poorly on darker-skinned faces and vice versa. This research underscores the need for more inclusive and diverse datasets in training FER models to ensure equitable performance across all demographics. The study's findings contribute to the broader effort to understand and mitigate biases in machine learning models, ultimately aiming for fairer and more accurate FER systems.