

The paper "Analyzing Biases in AU Activation Estimation Toward Fairer Facial Expression Recognition" by Miguel Monares et al. investigates the presence of racial biases in facial expression recognition (FER) models, particularly focusing on action unit (AU) estimation across different skin tones. Using synthetic facial datasets generated with Character Creator 4, the study systematically examines how an AU estimation network's performance varies with skin color. The research highlights significant disparities in the network's accuracy for different skin tones, underscoring the importance of dataset diversity for robust FER model performance.

The methodology involves creating 940 synthetic facial expression images, manipulating 10 specific AUs based on the Facial Action Coding System (FACS). These faces vary only in skin color and gender (male and female), ensuring controlled conditions for bias analysis. The AU estimation network, adapted from a state-of-the-art pain estimation model, was evaluated using paired t-tests to identify statistical differences in AU activation predictions between the lightest and darkest skin tones. Results indicate significant biases in several AUs and the Prkachin and Solomon Pain Intensity (PSPI) scores, suggesting the model's differential performance based on skin color.

Further experiments using models trained exclusively on lighter or darker skin tones demonstrate the impact of training data distribution on model biases. Models trained on lighter skin tones performed poorly on darker faces and vice versa, revealing complex, non-uniform biases across different AUs. These findings stress the need for comprehensive, diverse training datasets to mitigate biases in FER models, contributing to the development of fairer, more accurate AI systems for facial expression recognition.