

Here is a concise 3-paragraph summary of the paper:

<paragraph_1> This paper explores potential racial biases in facial expression recognition models using synthetically generated faces that only differ in skin tone. The authors created a dataset of 960 facial expression images of a European male and female face, systematically varying the skin tone across 10 shades from lightest to darkest based on the Monk Skin Tone Scale. They evaluated an action unit estimation network trained on this synthetic dataset to detect disparities in performance across different skin tones. Using paired t-tests, the authors found statistically significant differences in the model's output scores for certain action units and overall pain intensity estimation between the lightest and darkest skin tones, indicating the presence of skin color biases in the model. </paragraph_1>

<paragraph_2> To further investigate these biases, the authors conducted experiments systematically activating specific action units (e.g. brow lowerer, lip corner puller) across the skin tone range while holding other facial attributes constant. They found that the model's ability to track activation of certain action units diminished for darker skin tones, with contrasting biases across different action units. Training the model exclusively on lighter or darker faces exacerbated these skin tone biases, highlighting how the distribution of skin tones in the training data impacts model performance and biases. </paragraph_2>

<paragraph_3> The findings demonstrate the presence of problematic skin color biases in this facial expression model that interact with the biases caused by differences in the training data distribution. The paper highlights the importance of thoughtfully constructing diverse and representative training datasets to mitigate harmful biases in facial recognition models. This work serves as an important step towards ultimately understanding and removing such biases to develop fairer AI systems in the future. </paragraph_3>